BIROn - Birkbeck Institutional Research Online

Cartea, Alvaro and Karyampas, Dimitrios (2009) The relationship between the volatility of returns and the number of jumps in financial markets. Working Paper. Birkbeck College, University of London, London, UK.

Birkbeck
UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

The Relationship between the Volatility of Returns and the Number of Jumps in Financial Markets

Álvaro Cartea
Universidad Carlos III de Madrid

Dimitrios Karyampas
Birkbeck, University of London

November 2009

Birkbeck Working Papers in Economics & Finance

The Relationship Between the Volatility of Returns and the Number of Jumps in Financial Markets
Working Version

ÁLVARO CARTEA
Universidad Carlos III de Madrid
and
DIMITRIOS KARYAMPAS[*]
Birkbeck, University of London

First draft: June 10, 2009; This version: November 17, 2009

ABSTRACT

The contribution of this paper is two-fold. First we show how to estimate the volatility of high frequency log-returns where the estimates are not affected by microstructure noise and the presence of Lévy-type jumps in prices. The second contribution focuses on the relationship between the number of jumps and the volatility of log-returns of the SPY, which is the fund that tracks the S&P 500. We employ SPY high frequency data (minute-by-minute) to obtain estimates of the volatility of the SPY log-returns to show that: (i) The number of jumps in the SPY is an important variable in explaining the daily volatility of the SPY log-returns; (ii) The number of jumps in the SPY prices has more explanatory power with respect to daily volatility than other variables based on: volume, number of trades, open and close, and other jump activity measures based on Bipower Variation; (iii) The number of jumps in the SPY prices has a similar explanatory power to that of the VIX, and slightly less explanatory power than measures based on high and low prices, when it comes to explaining volatility; (iv) Forecasts of the average number of jumps are important variables when producing monthly volatility forecasts and, furthermore, they contain information that is not impounded in the VIX.

---

1. Introduction

Modeling and forecasting volatility of asset prices is a crucial task in finance. The recent financial crisis has highlighted the importance that investors place on the returns and volatilities of assets. During the crisis, the volatility of most financial assets almost doubled and, at the same time, changes in volatility (known as the volatility of volatility) also increased, reflecting the "puzzled" expectations and reactions of investors in the risky and uncertain environment. One example is the hike in the VIX, a measure of the implied volatility of the S&P 500 index options, that rose from an average value of 25% during 2007 to 70% towards the end of 2008.

Over the past years the literature on volatility estimation and forecasting has been very extensive. The common feature of most of these new studies is that high frequency, instead of daily, stock returns are employed.

Many methods have been proposed to estimate daily volatility using data at higher frequencies. One of the best known approaches is known as 'realized volatility' where volatility is calculated at a 5–minute sampling frequency, see Andersen and Bollerslev (1998). There are other more recent developments that estimate volatility at even higher frequencies (improving the consistency of the estimators relative to those based on the sparse sampling approach) some of which are also designed to address the problems stemming from the microstructure noise when sampling at high frequencies, see Zhang et al. (2005), Aït-Sahalia et al. (2005), Barndorff-Nielsen et al. (2008).

Another generation of papers also focuses on how to make the best use of ultra high frequency data to measure volatility of returns, but recognizes that discontinuities or jumps in the log-returns process must be accounted for, otherwise the volatility estimators will be considerably upward biased, see for example Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2006), Andersen et al. (2007), Mancini (2007), and Corsi et al. (2008). The key point about these estimators is that although they can handle rare big jumps they are not designed to deal with microstructure noise. Therefore, one way to deal with the noise is to use the sparse sampling approach.

More recent papers concentrate on the high frequency dynamics of prices and volatility of stock prices. Todorov (2009) investigates the temporal variation in the variance risk premium paying particular attention to jumps in stock prices as well as jumps in the volatility. Jacod and Todorov (2009) derive tests to decide whether jumps in volatility and jumps in prices occur simultaneously. The work of Todorov and Tauchen (2009) examines the path properties of the volatility where

one of their empirical findings is that the S&P 500 and the VIX jump at the same time. And Maheu and McCurdy (2009) examine the value that high frequency measures of volatility provide in characterizing the forecast density returns.

The contribution of this paper is two-fold. First we show how to estimate the volatility of log-returns where the estimates are not affected by the problems arising from microstructure noise and the presence of jumps. Here, jumps refer to price revisions that are *not* produced by Brownian motion or Gaussian shocks, but produced by either: large and rare Poisson-type events; or small infinite activity jumps, both of which we consider to be Lévy-type jumps.

The second contribution of our paper focuses on the link between volatility and the jumps in log-returns. We show that the number of jumps within a trading day helps to explain and forecast future volatility. To the best of our knowledge this is the first time that the number of jumps has been used as a measure of jump activity to explain and forecast the volatility of price innovations.

We show the empirical performance of our volatility estimator, and the link between the number of jumps and volatility, by employing minute-by-minute observations of the SPY, the fund tracker of the S&P 500, from January 2000 to December 2006. We highlight two of our empirical findings.

First, in addition to other well-documented variables such as: high, low, open, closing prices, volume and the number of trades, we show that the number of jumps in the SPY is a crucial variable in explaining the SPY volatility. We show that: a) the explanatory power of our proposed jump activity measure, given by the number of jumps, is higher than the explanatory power of previous jump activity measures when explaining the volatility component of log-price innovations. b) We show that the number of jumps in the SPY prices has more explanatory power with respect to daily volatility than other variables based on: volume, number of trades, and open and close. And c) We show that the number of jumps in the SPY prices has a similar explanatory power to that of the VIX, and slightly less explanatory power than measures based on high and low prices, when it comes to explaining volatility.

Second, using the number of jumps as an explanatory variable increases the forecasting ability of autoregressive volatility models. Results show that the incorporation of forecasts of the monthly average number of jumps in our volatility models leads to better monthly volatility forecasts and contain relevant information which is not impounded in the VIX. Hence, these models can be used to produce better volatility forecasts (model based forecasts) in addition to the well known and widely used market based forecasts such as the VIX.

The rest of the paper is organized as follows. Section 2 reviews the recent literature on volatility estimation using high frequency data. It also describes the estimators that have been proposed to account for the existence of jumps. Section 3 reviews the methodology of Lee and Mykland (2008) and Lee and Hannig (2009) to detect Lévy-type jumps in the SPY and discusses how we use these tests to produce daily volatility estimates not affected by jumps or microstructure noise. Section 4 describes the data used in our empirical study. Section 5 looks at different models proposed in the literature to explain and forecast volatility and presents the empirical results. Finally, Section 6 concludes.

2. Literature Review

In this section we review volatility estimators that use high frequency financial data. We focus on the different methods that have been proposed to estimate the true volatility of financial assets and deal with the problems arising in the presence of microstructure noise and jumps in the prices.

2.1. Volatility estimators when log-returns are described by Brownian motion

The initial approaches to volatility with high frequency data incorporate the concept of realized variance. The idea is to use intra-day returns to get a better estimate of daily volatility, an estimate that also captures the intraday variation of the financial asset.

We assume that the log-price of a security follows

$$X_t = \sigma W_t, \tag{1}$$

where W_t is a standard Brownian motion. In equation (1) the drift is not included because at high frequencies it is negligible relative to the diffusion.

The variance of (1) is defined as

$$RV_{X,T}^{(all)} = [X, X]_T^{(all)} := \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})^2, \tag{2}$$

where $RV_{X,T}^{(all)}$ is known as the realized variance of the log-returns process and is equal to the sum of the squared differences of X_t. The notation (all) means that we use all observations in the sample.

We also assume that the observations are equally spaced, so the time interval between them is constant and equal to Δ. The observations are recorded at times $t_i = i\Delta$ with $t_N = N\Delta = T$ for $i = 0, \ldots, N$, thus N denotes the number of observations between time 0 and T where, for practical purposes, it represents one trading day.

One problem arising from high frequency financial data is the presence of market microstructure noise. As a consequence, the true or efficient log-price, denoted by X_t, is contaminated by the microstructure noise ε_t and what we observe is a noisy log-price:

$$Y_t = X_t + \varepsilon_t. \tag{3}$$

Zhang et al. (2005) show that, at high frequencies, using (2) to calculate the realized variance of the log-price X_t is dominated by the variance of the noise term, hence we would obtain a biased estimate of the volatility. To overcome this problem, a typical approach is to sparse sample the data at frequencies that lessen the impact of microstructure noise on the volatility estimator. A common approach is to use 5-minute intervals and compute the realized variance with 78 observations within the day.

Zhang et al. (2005) show that arbitrary sparse sampling, such as always sampling at 5-minute intervals, regardless of the individual characteristics of the asset under study, is not the optimal way to proceed when plenty of data are available. They propose alternative non-parametric ways to estimate volatility without arbitrarily excluding large amounts of data. The best estimator they propose is the Two-Scale Realized Variance estimator $TSRV$. It is given by

$$TSRV_{Y,T} = RV_{Y,T}^{(avg)} - \frac{\bar{N}}{N} RV_{Y,T}^{(all)}, \tag{4}$$

where $\bar{N} = \frac{N-G+1}{G}$, $RV_Y^{(avg)} = \frac{1}{G} \sum_{g=1}^{G} RV_Y^{\mathcal{G}(g)}$ and $RV_Y^{\mathcal{G}(g)}$ is the realized variance for each grid and G is the total number of grids.

An alternative way to estimate volatility at higher frequencies is the parametric method of Aït-Sahalia et al. (2005). This method is based on the idea that the noisy returns r_i follow an $MA(1)$ process because r_i is defined as $r_i = \sigma(W_{t_i} - W_{t_{i-1}}) + \varepsilon_{t_i} - \varepsilon_{t_{i-1}} = \zeta_i + \eta\zeta_{i-1}$. Thus, they propose a maximum likelihood estimation method which produces fully efficient volatility estimates (MLE, hereafter) as well as estimates for the variance of the microstructure noise, σ_ε^2 (where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$). The crucial point in this approach is that we should specify correctly the

distribution of high frequency returns. If the price process is given by (3) then the MLE estimate proposed by Aït-Sahalia et al. (2005) is the most efficient estimate of volatility that we can get in the presence of microstructure noise. However, if the process is not described by (3) the efficiency of the estimator will be affected. In fact, stock price dynamics are poorly described by Brownian motion or a Gaussian process because price revisions also exhibit large jumps and small movements that cannot be attributed to a Gaussian process.

2.2. Volatility estimators when log-returns are described by Brownian motion and Poisson jumps

In this section we extend the dynamics of the efficient price (1) to incorporate shocks to price increments, in form of Poisson jumps, that better capture the price dynamics observed in the markets. Empirical studies argue that price dynamics contain such discontinuities, see for example Andersen et al. (2003), Barndorff-Nielsen and Shephard (2006), Lee and Mykland (2008) and many others.

So far the literature has included Poisson-type jumps, in the sense that they are relatively large and occur very seldom. Hence, we extend (1) in the following way:

$$X_t = \int_0^t \sigma dW_s + \int_0^t \kappa dN_s, \tag{5}$$

where κ is the random jump size and N_t a Poisson counting process with an adapted stochastic intensity parameter λ_t.

Using the definition of the RV in equation (2) it can be easily shown that in the presence of jumps the RV is a biased estimate of the true volatility.

$$RV_{X,T}^{(all)} = [X,X]_T + [J,J]_T = [X,X]_T + \sum_{i=1}^{N_T} \kappa_{\tau_i}^2, \tag{6}$$

where the quantity $\sum_{i=1}^{N_t} \kappa_{\tau_i}^2$ is the contribution of the jumps process to the $RV_{X,T}^{(all)}$.

To our knowledge, the first attempt to derive consistent estimates of the volatility σ of the Brownian part of the process X, in the presence of Poisson-type jumps, was that of *Power Variation*, introduced by Barndorff-Nielsen and Shephard (2004). The most widely used estimator that focuses

on the continuous part of (6) is the well-known *Bipower Variation*, defined as

$$BPV_t = \mu^{-2} \sum_{i=2}^{N} |r_{i-1}| \, |r_i| \,, \tag{7}$$

where r_i indicates the log-return, N is the total number of observations and $\mu \simeq 0.7979$. Its more general specification, given by the *Multipower Variation* is given by

$$MPV_t = \Delta^{1-\frac{1}{2}(\gamma_1 + \ldots + \gamma_M)} \sum_{j=M}^{N} \prod_{k=1}^{M} |r_{j-k+1}|^{\gamma_k} \,, \tag{8}$$

with γ_k, $k = 1, \ldots, M$, positive constants.

We denote the difference between the $RV_{X,t}^{(all)}$ and BPV_t by $J_t = RV_{X,t}^{(all)} - BPV_t$, first introduced by Barndorff-Nielsen and Shephard (2004). This quantity may be considered as an estimate of the jump activity during day t. The intuition behind this activity measure is that since the BPV estimator is a consistent estimator of the quadratic variation of X ($[X, X]_t$ as defined in (6)), and $RV_{X,t}^{(all)}$ is an estimate for both the continuous and the discontinuous part of X, the difference between $RV_{X,t}^{(all)}$ and BPV_t can be considered as an estimator of the component $\sum_{i=1}^{N_t} \kappa_{\tau_i}^2$ in (6). Even though one expects the difference $RV_{X,t}^{(all)} - BPV_t$ to be non-negative, one finds, in empirical studies, that this is not the case, and the solution has been to truncate J_t at 0 (see Andersen et al. (2007)); in other words

$$J_t = \max(RV_{X,t}^{(all)} - BPV_t, 0). \tag{9}$$

Finally, the J_t has been used in several studies in the literature to build jump detection tests and examine the informational content of jumps in volatility forecasts, see for instance Corsi et al. (2008) and Becker et al. (2009).

An alternative volatility estimator can be found in Mancini (2007). This estimator is based on a threshold approach labeled the *Threshold Realized Variance* (TRV) and defined by

$$TRV_{X,T} = \sum_{i=1}^{N} r_i^2 \mathbb{1}_{\left[r_i^2 \leq \Theta(\Delta)\right]} \,, \tag{10}$$

where $\Theta(\Delta)$ is the threshold function, N the number of observations, $\mathbb{1}_{[\cdot]}$ the indicator function and r_i the log return.

Finally, the last estimator we review here is an extension of Multipower Variation which in-

corporates the concept of the threshold approach. The estimator called $TBPV$, which stands for *Threshold Bipower Variation*, proposed by Corsi et al. (2008) is given by

$$TBPV_{X,T} = \mu^{-2}TMPV_{X,T} = \mu^{-2}\sum_{i=2}^{N}|r_{i-1}|\,|r_i|\,\mathbb{1}_{\left[|r_{i-1}|^2\leq\Theta_{i-1}\right]}\mathbb{1}_{\left[|r_i|^2\leq\Theta_i\right]}, \tag{11}$$

where

$$TMPV_{X,T} = \Delta^{1-\frac{1}{2}(\gamma_1+...+\gamma_M)}\sum_{j=M}^{N}\prod_{k=1}^{M}|r_{j-k+1}|^{\gamma_k}\,\mathbb{1}_{\left[|r_{j-k+1}|^2\leq\Theta_{j-k+1}\right]},$$

r_i is the log return, Θ_j the threshold function, $\mathbb{1}_{[\cdot]}$ the indicator function, γ_k, $k = 1,\ldots,M$, are positive constants and $\mu = 0.7979$ as above.

The $TBPV$'s advantage is that it gives unbiased estimates of volatility when consecutive jumps appear in our price process. The simpler Multipower Variation is highly affected by the presence of consecutive jumps and the bias of the volatility estimator could be extremely large.

Note that all estimators described in subsection 2.2 focus on the importance of discontinuities in the log-price dynamics, but ignore the effects of market microstructure noise even though they have been designed to use high frequency observations. Therefore, to mitigate the effects of microstructure noise on the volatility estimates, when employing these estimators capable of dealing with jumps, the sparse sampling approach has been employed.

3. Jump detection tests: the MLE-F as an alternative volatility estimator

From the review of volatility estimators presented in Section 2, it is clear that we can find ways of estimating the volatility of the diffusion part of the price process when microstructure noise or Poisson-type jumps are present. However, how can we estimate the volatility of the Brownian component in log-returns when more general processes are assumed to drive the price dynamics? How can we deal with the biases introduced into the volatility estimates by: (i) microstructure noise and (ii) jumps in the log-prices? In this section we provide an answer to both these questions.

In the literature we can find abundant evidence to demonstrate that the discontinuities present in the price innovations are better captured by a more general Lévy process where the arrival of jumps is not exclusively of Poisson-type, see Carr and Madan (1999), Carr and Wu (2003), Carr and Wu (2004), Carr and Wu (2007), Bakshi et al. (2008), and in portfolio management theory,

see Aït-Sahalia et al. (2009).

Considering only Poisson jumps ignores other Lévy-type jumps that are frequent and small. Thus, if these small price deviations, which are not Gaussian, are confounded with the Gaussian movements of the price, the estimator will produce incorrect volatility estimates.

Therefore, our aim is to propose a volatility estimator that is neither affected by Lévy-type jumps (infinite activity and Poisson) nor microstructure noise. First we assume that we observe the noisy log-prices

$$Y_t = X_t + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is the microstructure noise and the true price is given by

$$X_t = \int_0^t \sigma_s dW_s + \int_0^t dL_s, \tag{12}$$

where σ_t is the volatility, dW_t the increments of Brownian motion, and dL_t are the increments of a pure jump Lévy process.

Our goal is to produce consistent and unbiased estimates of the volatility parameter σ_t. We use the high frequency data within every trading day to estimate the intraday volatility and we assume that volatility can vary from day-to-day but that it is a constant within one trading day; an assumption that is supported by the findings in Oomen (2006). It is also possible to relax this assumption and allow for volatility to change within the day, see Christensen et al. (2009).

We deal with the two problems, jumps in returns and microstructure noise, in sequence. We start with the high frequency observations Y_t and:

1. Remove price revisions that come from Lévy shocks by:

   - Employing the non-parametric tests proposed by Lee and Mykland (2008) and Lee and Hannig (2009) to determine which price innovations come from a Gaussian process and which come from Lévy jumps. (The jump detection tests are discussed below in subsection 3.1).

   - Removing the log-returns that are not Gaussian, i.e. removing the jump component $\int_0^t dL_s$ from X_t in equation (12).

2. Once we have removed the jumps, our new series, which we denote $\widetilde{Y}_t$, is given by

$$\widetilde{Y}_t = \int_0^t \sigma_s dW_s + \varepsilon_t \tag{13}$$

9

which is the Gaussian component of the true log-price plus the microstructure noise. This allows us to employ the MLE proposed by Aït-Sahalia et al. (2005) on the series $\tilde{Y}_t$, which produces the most efficient estimate of daily volatility σ_t that we can obtain in the presence of microstructure noise.

3.1. Detecting jumps

Detecting Poisson-type jumps

Lee and Mykland (2008) propose a non-parametric test based on high frequency data to detect jumps that are generated by a non-homogeneous Poisson-type jump process. The test-statistic is based on the idea that if a jump occurred at time t_i, the return would be much larger than with usual innovations, while the *instantaneous volatility*, which in this case is an estimator not affected by jumps, would remain at the usual level. The statistic is

$$\mathcal{L}(i) := \frac{\log\left(\frac{S(t_i)}{S(t_{i-1})}\right)}{\widehat{\sigma}(t_i)}, \tag{14}$$

where the instantaneous volatility $\widehat{\sigma}(t_i)$ is given by

$$\widehat{\sigma}^2(t_i) := \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} \left|\log\left(\frac{S(t_j)}{S(t_{j-1})}\right)\right| \left|\log\left(\frac{S(t_{j-1})}{S(t_{j-2})}\right)\right|$$

where $S(t_i)$ denotes the stock price at time t_i.

Then, the ith observation is considered a jump if

$$\frac{\max|\mathcal{L}(i)| - C_n}{S_n} > 4.6001$$

where

$$C_n = \frac{(2\log n)^{1/2}}{c} - \frac{\log\pi + \log(\log n)}{2c(2\log n)^{1/2}}, \qquad S_n = \frac{1}{c(2\log n)^{1/2}},$$

$c = (2/\pi)^{1/2}$, n is the total number of observations and K is the time "window" used to calculate the instantaneous volatility.

Detecting infinite activity jumps from a Lévy jump process

In a recent working paper, Lee and Hannig (2009) propose a similar non-parametric test to detect Lévy-type jumps: jumps that are difficult to locate due to their infinite activity and their small size which makes it difficult to differentiate them from price changes that are a result of a Gaussian shock.

Being able to detect all Lévy-type jumps in log prices allows us to separate the contribution to the noisy log-price Y_t that comes from a Gaussian shock or a jump. This decomposition is crucial because we can reduce the empirical problem of estimating the daily volatility σ to one that we can solve by applying the MLE estimator to the noisy series $\widetilde{Y}_t$, as defined by (13).

3.2. The MLE-F as an alternative volatility estimator

Once we have filtered the jumps, by detecting and removing them from the price series, we are left with the noisy log-price series $\widetilde{Y}_t$ given by equation (13). It is clear that the log-price process $\widetilde{Y}_t$ only contains a combination of two Gaussian shocks: price innovations in the continuous part of the true log-price process X_t and the microstructure noise ε_t.

Therefore, using $\widetilde{Y}_t$ as the starting point significantly simplifies the problem of estimating daily volatility because we can apply the MLE approach on the new vector of log-prices $\widetilde{Y}_t$. In discrete time, the vector of log-prices is $\widetilde{Y}_{t_i}$ where $t_i = i\Delta$ with $t_N = N\Delta = T$ for $i = 0, \ldots, N$, and Δ is the time-step between observations. And since we measure daily volatility, $T = 1$, i.e. one trading day, and N is the number of (Gaussian) observations in the trading day.

The efficient estimator for σ^2, which we label MLE-F to underscore the fact that we employ an MLE on a log-price series where jumps have been filtered, is calculated in the following way.

First, we note that the log-returns of $\widetilde{Y}_t$ follows an $MA(1)$:

$$\widetilde{Y}_{t_i} - \widetilde{Y}_{t_{i-1}} = \widetilde{r}_i = \zeta_i + \eta\zeta_{i-1} \tag{15}$$

where the ζ_i's are i.i.d. $N(0, \gamma^2)$ and

$$\gamma^2(1 + \eta^2) = \sigma^2\Delta + 2\sigma_\varepsilon^2,$$

$$\gamma^2\eta = -\sigma_\varepsilon^2.$$

Then we estimate σ^2 and the variance of the microstructure error σ_ε^2 by estimating γ^2 and η which result from maximizing the log-likelihood function

$$l(\eta, \gamma^2) = -\ln \det(V)/2 - N/2 \ln(2\pi\gamma^2) - \frac{1}{2\gamma^2}\tilde{r}'V^{-1}\tilde{r}, \tag{16}$$

where $\gamma^2 V$ is the covariance matrix of the returns:

$$V = \begin{pmatrix} 1+\eta^2 & \eta & 0 & \cdots & 0 \\ \eta & 1+\eta^2 & \eta & \ddots & \vdots \\ 0 & \eta & 1+\eta^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \eta \\ 0 & \cdots & 0 & \eta & 1+\eta^2 \end{pmatrix},$$

and $\tilde{r} = (\tilde{r}_1, \ldots, \tilde{r}_N)'$ is the returns vector that does not contain jumps.

Another important feature of this approach is that the MLE is theoretically robust to misspecification of the marginal distribution of the noise process. Furthermore, instead of assuming that microstructure noise is i.i.d. we could introduce dependence in the noise process, for example that the microstructure noise follows and $MA(1)$, and adopt the extended MLE approach of Gatheral and Oomen (2009).

4. Data

The data used in this paper are taken from the Trade and Quote (TAQ) and CRSP databases and our analysis is based on S&P 500 index. Instead of using S&P 500 TAQ data, we proceed as in Verardo and Patton (2009) and employ the exchange traded fund tracking the S&P 500 index; the S&P 500 SPDR traded on Amex with ticker SPY and available on the TAQ database. As Verardo and Patton point out "this fund is very actively traded and...the fund's price does not deviate from the fundamental value of the underlying index."

The SPY transactions included here are trades from 9.30am to 4.00pm on a sample period that runs from January 2000 to December 2006. The total period consists of $1,759$ trading days. The first 12 days are used as an initial window (K, as defined above in (14)) to employ the test statistic that detects the jumps in the SPY log-returns. Therefore the sample used to compute realized variances and jump activity consists of $1,747$ days, i.e. $681,330$ high frequency (minute-by-minute)

observations.

A practical problem we face when using the high frequency data is that data are irregularly spaced and it is not possible to find a sequence of trades with exactly $\Delta = 60$ seconds between them. To overcome this, we prepare the data by designating the last trade within the preceding 60 seconds as the observation for the minute in question. For example, if over the period 9:35:00am to 9:36:00am the last trade took place at 9:35:40 (40 seconds after 9:35am) we take that observation and record it as the 9:36am observation. Moreover, although very seldom, in the event that there was no trade the minute in question we use the last observation available. For example, if over the period 9:35:00am to 9:36:00am there was no trade, we assume that there was a trade and take the price of the trade used for the slot 9:33:00am to 9:34:00am.

Below, in the empirical section, we produce monthly volatility forecasts based on different econometric models and we test their forecasting ability. In addition to model-based volatility forecasts, we also use the VIX index over the same period, January 2000 to December 2006. VIX is the implied volatility of the S&P 500 provided by the Chicago Board of Options Exchange (CBOE). VIX is derived from call and put options on the S&P 500 index and can be used as a market implied volatility forecast for the next 22 trading days. Several studies show that the VIX conveys information about the future realized volatility, so it can be used to enhance econometric forecasting volatility models, see, for instance, Fleming (1998), Jiang and Tian (2005), Szakmary et al. (2003) and Blair et al. (2001). Other variables such as the volume, high and low prices, open and close are taken from CRSP (The Center for Research in Security Prices).

Once we have prepared the data in a minute-by-minute format we apply both jump detection tests, described above, to obtain both the position and total number of jumps within each trading day. Here we denote the total number of jumps during day t by NJ_t. The total number of jumps for the entire period examined here is $1,899$, thus implying a mean value of 1.087 jumps per day t in the SPY. The maximum number of jumps in one day is 32, which represents 8.2% of the minute-by-minute trades taking place on 28 August 2001. We also calculate the jump activity measure defined above, $J_t = \max(RV_t^{(all)} - BPV_t, 0)$. Figure 1 presents the SPY return for the above period, where red circles and black dots indicate big and small jumps respectively.

Figure 1 About Here

Finally, in Figure 2 we show the QQ Plots of the SPY returns before and after removing the Lévy-type jumps. The blue circles depict the QQ Plot of the SPY returns and it is evident that

they are not Normal. The light blue dots represent the QQ Plot of the filtered SPY returns where it is clear that these filtered returns are very 'close' to being Normally distributed.

Figure 2 About Here

5. Empirical Results: Explaining and Forecasting Volatility

In this section we investigate the relation between alternative volatility estimators and two jump activity measures, the one we propose (the actual number of jumps NJ_t) and the one proposed previously by the literature J_t. We present the models used to explain the SPY volatility; the models that are employed to obtain good volatility forecasts; and discuss the empirical results.

The volatility estimators used here are: (i) MLE, (ii) BPV, (iii) $TBPV$ and (iv) MLE-F. As discussed above the MLE approach assumes that there is i.i.d. microstructure noise in the log-prices and, consequently, the returns follow an $MA(1)$ process. We show that this is the case in Figure 4 where it is clear that there is significant autocorrelation of the SPY returns at lag 1.

Figure 4 About Here

As a preliminary result we present the correlations between the daily volatility estimates that are obtained from the four estimators in Table 1. Moreover, although below we only focus on these four estimators, Table 1 also shows the correlation of the volatility estimates when $|r_t|$ and $GARCH$ are used as alternative volatility estimators, using daily data. The table also shows the correlation of the VIX with the other volatility estimates. For clarity, in all tables results that are in bold and italic denote parameters that are not significant at the 5% level and results that are in bold are those that we highlight and discuss in the text.

From Table 1 we can observe that the two estimators that show the lowest correlations with the other measurements of volatility are the ones based on daily data (for instance, $Cor(|r_t|$, $MLE - F) = 0.77$, i.e. the smallest correlation value in the first MLE-F column), while the correlation between them ($|r_t|$, $GARCH$) is equal to 96% (in bold). We can also see that the highest correlation between the daily volatility estimates obtained with the MLE-F and the other estimators is of 0.97, which results from the volatility estimates of the $TBPV$ proposed by Corsi et al. (2008).

Table 1 About Here

14

Figure 3 shows the time series of the daily volatility estimates using the MLE, BPV, $TBPV$ and MLE-F. It is interesting to observe that the MLE-F does not exhibit too many 'spikes' whereas the BPV volatility estimates exhibit large fluctuations which could result from the fact that, in theory, the BPV volatility estimates are not affected by big and rare jumps, but are affected by small jumps and by microstructure noise.

Figure 3 About Here

5.1. The informational content of the realized number of jumps

In this section we investigate whether the jump activity of the SPY has any explanatory power on the volatility of its log-returns. To do so, we apply several time series models that have jump activity measures as explanatory variables. We proceed in the following way: (i) We apply an autoregressive model without any jump activity measure as an independent variable. (ii) We then extend the autoregressive model by including the jump activity measure J_t , defined in (9), as a further explanatory variable. (iii) Instead of including J_t as explanatory variable, we include the number of jumps NJ_t as an alternative jump activity measure. (iv) And finally, we extend the autoregressive model to include both jump activity measures: J_t and NJ_t. These four steps allow us to compare the performance of the two jump activity measures, in terms of their explanatory power, in the different volatility models where we use an autoregressive model as the benchmark.

All models in this section examine the relations between the logarithmic value of volatility and the independent variables. The estimation is done using the four procedures: MLE, BPV, $TBPV$ and MLE-F. We use logarithmic values of volatility because they seem to be normally distributed relative to the actual volatility values σ_t. The autoregressive model $AR(1)$ for log-volatility is

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + u_t, \tag{17}$$

where σ_t are the daily estimates provided by the four alternative volatility estimators and u_t is the noise term of the regression.

Table 2 shows that all $AR(1)$ coefficients are significant and the parameter responsible for the speed of mean reversion shows that volatility is a highly mean-reverting process. This fact has been extensively documented in the volatility literature (see for instance Merville and Pieptea (1989), Bali and Demirtas (2008)).

Table 2 About Here

The adjusted R^2 value for all regressions is very high, with values close to 75%. For all regressions we use the Heteroscedasticity and Autocorrelation Consistent covariance estimates proposed by Newey and West (1987).

At this point we incorporate into our analysis the effects of the jump activity as another explanatory variable in the regressions of the SPY return volatility. We do this in the following set of regressions where we extend the $AR(1)$ process, see equation (17), by incorporating a jump activity measure and its past value. The first jump activity measure we include is the J_t given above in equation (9). The other jump activity measure is the one we propose here, and is examined for the first time in this paper, which is given by the actual number of jumps occurring within each trading day.

Therefore, in the case where we test the explanatory power of the jump activity measure J_t we employ

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 J_t + \beta_3 J_{t-1} + u_t, \tag{18}$$

where σ_t are the daily estimates provided by the four alternative volatility estimators and u_t is the noise term of the regression. Similarly, by replacing J_t with NJ_t we test the explanatory power of the actual number of jumps as a new measure to explain volatility by running the regression

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 NJ_t + \beta_3 NJ_{t-1} + u_t, \tag{19}$$

and finally, we run the regression that extends the $AR(1)$ to include both jump activity measures

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 J_t + \beta_3 J_{t-1} + \beta_4 NJ_t + \beta_5 NJ_{t-1} + u_t. \tag{20}$$

The estimates of the coefficients are presented in Table 3.

Table 3 About Here

The J_t variable is significant for the MLE, BPV, and $TBPV$ volatility estimates while its lagged value J_{t-1} is only significant for the MLE and BPV, all at a 5% significance level. J_t does not provide any informational content in the new MLE-F volatility estimator introduced in this paper. By looking at the R^2s of the AR(1) in Table 2 and the R^2s in Table 3, we see

16

that the incorporation of J_t and its lagged value increase the models' explanatory power for the MLE, BPV, and $TBPV$ but not the MLE-F. Moreover, by looking at the information criteria we see that they all have smaller values than in the previous pure $AR(1)$ regressions, while the log-likelihood maximum values ($\log L$) are greater in this case.

Table 3 shows interesting results about the jump activity measure NJ_t. The coefficients for the number of jumps NJ_t and NJ_{t-1} are positive and negative respectively and highly significant in all cases, even for the MLE-F volatility estimate. We see that for BPV, for instance, the adjusted R^2 value increases from 69.99%, in the $AR(1)$ model (Table 2), to 80.9% with the inclusion of the NJ_t and NJ_{t-1} regressors. This R^2 value is even higher than in the case where we only had the J_t jump activity measure, where the adjusted R^2 was 72.4%, as shown in the first column of $\log BPV$ results in Table 3.

In the pure $AR(1)$ the $\log L$ value for the BPV is -298.44, but when the NJ's are included, see equation (19), the $\log L$ increases to 98.39 which implies that the extended $AR(1)$ is a much better model to explain the market's variation. Moreover, in this extended model, the information criteria values decrease when NJ_t and NJ_{t-1} are included. This implies that the NJ_t, and its lagged value, increase the explanatory power of the volatility models more than the J_t and J_{t-1} variables do.

Similarly, for the other models, we observe that the adjusted R^2s and information criteria show that the extended model (19) performs better. For example, in the case of the MLE, the adjusted R^2 value increases from 72.4% in the $AR(1)$ model to 77.7% in the model with NJ_t and NJ_{t-1}, while the $\log L$ value for the pure $AR(1)$ is -42.09, see Table 2, and for the extended $AR(1)$ is 145.13 as shown in Table 3. And, furthermore, the adjusted R^2 when daily volatility is calculated using the $TBPV$, is 86.8% in the extended model whereas it is 84.7% in the pure $AR(1)$ model. Finally, when we use the MLE-F to calculate daily volatility, the $\log L$ and R^2 in the pure $AR(1)$ case are 235.80 and 79.2% respectively, and in the extended case the $\log L$ and R^2 are 304.38 and 80.8% respectively.

In general, including the number of jumps as an explanatory variable increases the R^2 of the four estimators. This increase is higher for the MLE and the BPV, while not as large as that for the $TBPV$ and the MLE-F. But for all estimators the increase of the explanatory power due to the number of jumps NJ_t and its lagged value is much higher than that of the J_t and J_{t-1} as jump activity measures.

The last test that we require to show that NJ_t conveys more information than J_t to explain the volatility of the SPY log-returns, is to run the regression (20) where both J_t and NJ_t, and their lagged values, are used as independent variables at the same time. The result of this last set of regressions is the last column of each of the four estimators in Table 3.

In most cases NJ_t and NJ_{t-1} overlap the effect that J_t and J_{t-1} have on volatility. J_t and its lagged value become statistically insignificant for BPV, $TBPV$ and MLE, while NJ_t and NJ_{t-1} still have significant coefficients for all estimators. The explanatory power of the model, where both J_t and NJ_t, and their lagged values, are taken into account, is close to the model where only NJ_t was used to explain volatility.

5.2. Using alternative explanatory variables: VIX, high and lows and number of trades

So far we have focused on two jump activity measures when extending the pure $AR(1)$ log-volatility process to identify which of these two competing measures has more power to explain the daily volatility of the SPY returns. There are, however, many other well studied factors that have been successfully employed to explain the behavior of the volatility of assets' log-returns. Therefore, now that we have established that the jump activity measure NJ_t is an important variable to consider when studying the volatility of the SPY, we need to judge whether, in the presence of other significant explanatory variables, the number of jumps in prices is still a variable with explanatory power.

The other variables we consider here are: the VIX, the difference between (log) close and (log) open, and the difference between (log) high and (log) lows (for previous studies that use such variables as sources of volatility see, for instance, Alizadeh et al. (2002), Garman and Klass (1980), Rogers and Satchell (1991) Yang and Zhang (2000)). Variables such as the volume and the number of transactions within each day have been also used in the literature to explain assets' volatility. We shall also consider these two other variables in our analysis, but do it in subsection 5.3 below.

The regression we use is the following

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 NJ_t + \beta_3 NJ_{t-1} + \beta_4 \log HL_t + \beta_5 OC_t + \beta_6 \log VIX_t + u_t, \quad (21)$$

where $HL_t := \log(\text{high}_t/\text{low}_t)$ and $OC_t := |\text{close}_t/\text{open}_t|$. Table 4 below presents the results of

18

equation (21).

When comparing the results of running the regression (21) with those of the regression (19) we find that all independent variables are statistically significant for the MLE estimator and the adjusted R^2 resulting from running (21) is 87.4% which is higher than the 77.7% obtained from running regression (19).

Regarding the BPV estimator, the adjusted R^2 also increases from 80.9%, in the case where we had only the NJs as explanatory variables, to 86.3% with the incorporation of the extra variables in (21). In addition, the variable OC_t is insignificant at a 5% significance level.

The explanatory power of the model for the $TBPV$ case becomes also higher with the extra explanatory variables. The new R^2 is 90.8%, which is higher than the 86.8% obtained in model (19). The new $\log L$ value has nearly doubled, increasing from 487.47 to 802.09.

For the MLE-F volatility estimator the incorporation of the extra variables increases the adjusted R^2 from 80.8% to 88.8%. Moreover, the coefficient of NJ_{t-1} is insignificant.

Finally, information criteria for all four estimators become lower when the extra explanatory variables are incorporated in the model described by (19), implying that this model better explains the SPY volatility.

Therefore, it is interesting to see that our proposed jump activity measure, given by the number of jumps within each day, is still statistically significant after the incorporation of the extra explanatory variables in the model. Thus, in addition to the well documented explanatory variables of volatility, the number of jumps in the underlying asset is also a driving source of volatility. The coefficient of the NJ_t is positive while the coefficient of NJ_{t-1} is negative.

5.3. Volume vs. number of trades

In this section we examine the relationship that the volume of trades and the number of trades have with volatility. The seminal work of Clark (1973) argues that markets operate at different trading rates over different periods of time. Therefore, in Clark's work it is argued that the trading volume is positively related to the number of intraday transactions and hence that the trading volume is related to the variability of price change. This positive relationship between price changes and

trading volume, without controlling for the number of transactions, is further investigated in Epps and Epps (1976), Gallant et al. (1992), Tauchen and Pitts (1983) and Karpoff (1987).

However, in Jones et al. (1994) it is shown that the positive relationship between volume and volatility, measured as absolute or squared price changes, actually reflects the positive relationship between volatility and the number of transactions. In addition, Ross (1989) also documented that the number of trades, rather than the volume, is the driving source of asset price volatility.

Here we run three different regressions to examine which jump activity measure, NJ_t or J_t, is more important in explaining volatility when volume and/or number of trades are also included as regressors. First, we run a regression that uses, in addition to the explanatory variables of equation (21), volume as an independent variable to explain volatility. Second, we run the same model but instead of volume we use the number of trades as an explanatory variable. Finally, we incorporate both volume and number of trades in (21).

The three regressions are:

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 NJ_t + \beta_3 NJ_{t-1} + \beta_4 \log HL_t + \beta_5 OC_t + \beta_6 \log VIX_t + \beta_7 VOL_t + u_t, \quad (22)$$

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 NJ_t + \beta_3 NJ_{t-1} + \beta_4 \log HL_t + \beta_5 OC_t + \beta_6 \log VIX_t + \beta_7 NT_t + u_t, \quad (23)$$

and

$$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} + \beta_2 NJ_t + \beta_3 NJ_{t-1} + \beta_4 \log HL_t + \beta_5 OC_t + \beta_6 \log VIX_t + \beta_7 VOL_t + \beta_8 NT_t + u_t,$$
$$(24)$$

where VOL_t is the volume traded throughout day t, NT_t number of trades throughout day t (both measures are between 9.30am and 4:00pm), and the other regressors are as described above in equation (21).

Table 5 presents the results for all of the above models for the different volatility estimators: MLE, BPV, $TBPV$ and $MLF - F$

Table 5 About Here

The BPV estimator. For the model given by (22) the adjusted R^2 value remains the same as it was for (21). The coefficient of the volume regressor is significant but very small, thus not too relevant to explain the SPY volatility. The explanatory power for the model (23) is again at the

level of 86.3% with the coefficients of NT_t and OC_t being statistically insignificant. When both volume and number of trades are included in the model as explanatory variables of volatility, they are statistically significant and the adjusted R^2 increases from 80.9% in the simple case without the extra variables (see Table 3, 2nd column for BPV) to 86.4%. VIX_t exhibits a high coefficient, close to the coefficient of the autoregressive term. The value of the HL_t coefficient is also relatively high, implying that a great deal of variation in the SPY log-returns is explained by the difference between the maximum and the minimum SPY prices within each day.

The $TBPV$ **estimator**. In the case of $TBPV$ the volume is insignificant for models given by (22) and (24), while the number of trades is significant in the latter model, something that also increases the $\log L$ value to 803.97 from the 802.09 obtained in equation (21).

The MLE **estimator**. For the MLE estimator the volume is significant when running model (22), and NT_t is significant when (23) is implemented. But both are statistically insignificant when both are included in the regression (24) as explanatory variables for volatility. The coefficient of the autoregressive term becomes very small, with VIX_t and HL_ts' coefficients taking values close to or even higher than the β_1 estimate.

The $MLE - F$ **estimator**. The crucial result regarding which of the variables, volume or number of trades, is the driving source of volatility is given by the set of regressions applied to our volatility estimator; the estimator that is neither affected by microstructure noise nor jumps. When (22) is implemented, volume is statistically significant but its coefficient value is very small. In the case where only number of trades are used in the model, the NT_t's coefficient is positive, equal to 0.056, and statistically significant at 5% significance level. The interesting point occurs when both volume and number of trades are used as independent variables in the model (24). In this case, the coefficient of volume becomes statistically insignificant, while that of the number of trades is still significant and its value increases from 0.056 to 0.06.

In general, NT_t is statistically significant for most volatility estimators (except MLE) in model (24). On the other hand, volume is insignificant for the MLE, $TBPV$ and MLE-F. This implies that the number of trades, instead of volume, is a better explanatory variable of volatility. Our results agree with those of Jones et al. (1994) and Ross (1989), where it is documented that the number of trades, rather than the volume, is the driving source of asset price volatility.

Regarding the number of jumps, we see that for all models (22), (23) and (24), NJ_t and NJ_{t-1} are significant for the volatility estimators MLE, BPV and $TBPV$ and only NJ_t is significant for

the $MLF - F$. Hence, we see that the number of jumps is still relevant in the presence of other variables that incorporate information such as high, low, volume, number of trades, open, close and the VIX.

Finally, for comparative purposes, Table 7 shows adjusted R^2s resulting from further tests on the explanatory power of the different variables used to explain daily volatility. The first row of the Table shows the regressions we run where daily log-volatility is explained by one of the following variables: $\log \sigma_{t-1}$, $\log HL_t$, OC_t, $\log VIX_t$, NJ_t, NT_t or VOL_t . The results show that the number of jumps NJ_t in the SPY prices has more explanatory power with respect to daily volatility than other variables based on: volume, number of trades, open and close. Moreover, NJ_t has a similar explanatory power to that of the VIX, and slightly less explanatory power than measures based on high and low prices, when it comes to explaining volatility.

Table 7 About Here

5.4. Forecasting volatility

At this point we examine whether the number of jumps in the SPY improves the ability to forecast the S&P 500 volatility. Previous studies show that jumps play an important role in both forecasting and explaining volatility. For instance, Corsi et al. (2008) provide evidence that volatility forecasts exhibit lower Root Mean Square Error (RMSE) when jump activity measures are included in the autoregressive models of volatility. And Andersen et al. (2003) show that the jump measures based on BPV help to explain realized volatility.

Here we follow a similar approach to those of Corsi et al. (2008) and Andersen et al. (2003), but propose the use of the number of jumps as a key variable to explain and forecast volatility. Furthermore, from this point onwards, we only use the MLE-F to produce volatility figures because it is the only estimator that deals with both jumps and microstructure noise.

The goal is to obtain, at time t, a volatility forecast for the next 22 trading days. That is, obtain a model-based forecast (MBF) of the volatility of log-returns of the S&P 500 for the next month. Our notation for the monthly volatility forecast at time t is $\hat{\sigma}_{t \to t+22}$. This means that at time t we produce a forecast which is the square root of the average variance for the next 22 trading days. In our study, the first day for which we produce a forecast is October 5 2006 and the last day is November 9 2006. The forecast we obtain on October 5, $\hat{\sigma}_{\text{Oct } 5 \to \text{Nov } 9}$, is a forecast of the square root of the average variance between October 5 and November 9 of 2006. And our last forecast for

November 9 2006, $\hat{\sigma}_{\text{Nov 9}\rightarrow\text{Dec 11}}$, is a forecast of the square root of the average variance between November 9 and December 11 of 2006.

In the literature we can find several MBFs together with market-based forecasts such as the VIX (for more details see Becker et al. (2009)). Becker et al. present MBFs that incorporate the BPV jump activity measure J_t as one of the variables in the models. The jump activity measure we use here to forecast volatility is based on $\overline{NJ}_{t\rightarrow t+22}$, which denotes the average of the number of jumps for a period of 22 trading days starting at time t.

When forecasting volatility, the VIX is an important variable to consider because it is a measure of the implied volatility of options on the S&P 500 index and therefore it conveys (risk-neutral) forward looking information about what market participants expect the volatility of the S&P 500 to be over the coming month. Therefore, although the VIX contains information about the volatility of the risk-neutral log-returns of the S&P 500, and our objective is to obtain forecasts of the volatility of the S&P 500 log-returns under the data generating measure, we expect the VIX to increase the forecasting ability of MBFs of volatility.

To test whether the VIX can help to forecast realized volatility, we run the following regression:

$$\log \sigma_{t\rightarrow t+22} = c + \beta \log VIX_t + u_t, \tag{25}$$

where $\sigma_{t\rightarrow t+22}$ are obtained using the MLE-F over the period from t to $t+22$. Note that when we calculated daily volatility we used the minute-by-minute returns from 9.30am until 4.00pm for one day and applied the MLE-F to that data set. Now, since we are interested in the square root of the average variance from t to $t+22$, we proceed as with the daily estimates, but our minute-by-minute data stretch from 9.30am of day t until 4.00pm of day $t+22$. The estimate of β in equation (25) is close to 1.07 and the adjusted R^2 is 75.26%. We consider this to be a reasonable indication that the VIX is as a good market forecast of the square root of the average variance over a month and may be used to enhance our model-based volatility forecasts in the models below.

Another model that can be used to produce volatility forecasts is

$$\log \sigma_{t\rightarrow t+22} = c + \beta_1 \log \sigma_{t-22\rightarrow t} + \beta_2 \log VIX_{t-1} + u_t, \tag{26}$$

which is an extension of (25) where $\sigma_{t-22\rightarrow t}$ denotes the square root of the average variance between time $t-22$ and t.

For our purposes, we are interested in a further extension to model (26), where the average number of jumps between time t and $t + 22$ is used as an explanatory variable. However, at time t we do not know the number of jumps that will occur between t and $t+22$. Instead, we need to use a forecast of the average number of jumps. Before doing this, we first show that the realized average number of jumps between t and $t + 22$, $\overline{NJ}_{t \to t+22}$, is a relevant variable in explaining monthly volatility. And, we run the following regression

$$\log \sigma_{t \to t+22} = c + \beta_1 \log \sigma_{t-22 \to t} + \beta_2 \log VIX_{t-1} + \beta_3 \overline{NJ}_{t \to t+22} + u_t \qquad (27)$$

to show that $\overline{NJ}_{t \to t+22}$ can explain $\sigma_{t \to t+22}$.

<center>Table 8 About Here</center>

Table 8 presents the results from estimating models (26) and (27). Once again, the average number of jumps plays an important role in explaining monthly volatility. The explanatory power of the model that includes the VIX and the average number of jumps is 83.50%, whilst in the model described by (26), where only the VIX is employed, the adjusted R^2 is 78.30%. The high adjusted R^2 values indicate that model (26) can be used to forecast monthly SPY volatility. In the case of model (27) it shows that the average number of jumps is an important variable to explain monthly SPY volatility and, therefore, it provides an indication that forecasts of the expected number of jumps may be an important variable to generate monthly volatility forecasts.

In the second column of the table, we present the model where instead of the VIX variable we only use the average number of jumps to explain volatility. The interesting point in this regression is that the adjusted R^2 is equal to 80.60%, higher than the 78.30% for the model where only the VIX is employed. This highlights the importance of incorporating the average number of jumps in the model to explain monthly volatility.

We emphasize that model (27) employs the number of jumps between t and $t + 22$ to explain monthly volatility at time t. However, as discussed above, if our intention is to forecast rather than explain volatility at time t over the period $t \to t + 22$ we must use a forecast of the $\overline{NJ}_{t \to t+22}$ parameter.

Therefore, we obtain forecasts of the monthly average number of jumps by employing a moving average process of order 5 ($MA(5)$) and denote these forecasts $\widehat{\overline{NJ}}_{t \to t+22}$. Table 6 shows the results for the $MA(5)$ model. The adjusted R^2 of 90% indicates that the moving average specification can

<center>24</center>

explain most of the variability of the average jumps activity measure. The model gives us forecasts for the period 05 October 2006 to 09 November, 2006.

<center>Table 6 About Here</center>

Finally, the model that we use to derive volatility forecasts is:

$$\log \sigma_{t \to t+22} = c + \beta_1 \log \sigma_{t-22 \to t} + \beta_2 \log VIX_{t-1} + \beta_3 \widehat{\overline{NJ}}_{t \to t+22} + u_t. \tag{28}$$

The monthly volatility forecasts are for the period 5 October 2006 to 9 November 2006, i.e. 26 monthly volatility forecasts. Note that this model only differs from (27) in that we use the $MA(5)$ forecasts $\widehat{\overline{NJ}}_{t \to t+22}$ as the regressor .

We present four indicators of how good our forecasts are relative to the true values:

- Root Mean Squared Error (RMSE)

- Mean Absolute Error (MAE)

- Mean Absolute Percentage Error (MAPE)

- Theil Inequality index, defined as

$$\text{TI} = \frac{\sqrt{\sum_{T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}}{\sqrt{\sum_{T+1}^{T+h} \hat{y}_t{}^2 / h} + \sqrt{\sum_{T+1}^{T+h} y_t^2 / h}},$$

where $\hat{y}_t$ is the forecast, y_t the observed value and h the total number of forecasts. The TI lies between zero and one, with zero indicating a perfect fit.

The last column of Table 8 presents the estimates of model (28). We see that when using the forecast of the average number of jumps, the adjusted $R^2 = 83.5\%$ which incidentally is the same adjusted R^2 that we obtained when the actual average number of jumps was used as regressor in model (27).

The last four rows of Table 8 present the four indicators for models (26) and (28) for the last 22 days of our sample. We see that the values of the RMSE, MAE and MAPE are smaller in the case where the forecast of the average number of jumps is incorporated in the model to forecast monthly volatility. Model (28) has a higher adjusted $R^2 = 83.5\%$ than that of model (26) which is 78.30%. Furthermore, the TI coefficient of model (28) is 0.016 which indicates a very good fit and it is also lower that the TI of model (26).

<center>25</center>

Therefore, in this subsection we have shown that employing forecasts of the average number of jumps provides crucial information that increases our ability to forecast the SPY one-month volatility. Jumps are important when considering the forecasting problem and one should take them into account in order to obtain better volatility forecasts. Information such as the expected number of jumps should be used in addition to the market-based forecast VIX which does not seem to include all the information relating to the jump activity.

5.5. Why is the realized number of jumps helpful in explaining volatility?

Our empirical results highlight how important the realized number of jumps in the SPY prices are in explaining the volatility of the SPY log-returns. In all the models that we run we find that daily volatility σ_t depends positively on the realized number of jumps NJ_t and negatively on the realized number of jumps for the previous day NJ_{t-1}.

We provide the following intuitive explanation of the relationship between the number of jumps in the price and its volatility. In the literature, the arrival of information is linked to trading activity, number of trades, volume of trades, etc. and therefore information flows are a driving source of volatility, see for example Clark (1973), Tauchen and Pitts (1983) and Ross (1989). Jumps in stock prices can be attributed to the release of important pieces of information, Lee and Mykland (2008). In addition, our results indicate that when there are jumps in stock prices, the volatility also increases. Our interpretation is twofold: (i) that there is information that moves stock prices; and (ii) that the way in which market participants interpret this new information, can also affect the volatility of the price innovations. For instance, when market participants are broadly in agreement, the volatility of price revisions, after the price jump, should remain close to the current volatility levels. On the other hand, if most market participants do not agree on how to interpret the arrival of news, the volatility will increase.

Although the objective of the recent work of Todorov (2009) is to study the temporal variation in the market variance risk premium in the presence of jumps both in the volatility and the stock price process, the author shows that there is a strong relationship between the jumps in the S&P 500 index futures contract and the jumps in its variance. The author also conjectures that situations where jumps in volatility and prices occur at the same time, may be caused by arrival of information that simultaneously impacts the price the volatility jumps due to the fact that market participants cannot agree on the effect that this information will have in the market.

6. Conclusions

The first contribution of the paper is to propose the MLE-F volatility estimator, an estimator that is neither affected by jumps (Lévy-type that include infinite activity and Poisson) nor microstructure noise. We propose a two-step procedure, that takes into account recent developments in jump detection tests and high frequency volatility estimation, to obtain a fully efficient volatility estimator in the presence of noise and price discontinuities.

The second contribution is to examine the relationship between the jump activity of the SPY, the exchange traded fund tracking the S&P 500 index, and its volatility. We employ high frequency data (minute-by-minute between 9:30am and 4:00pm from January 2000 to December 2006) and deconstruct the SPY high frequency returns into its Lévy-type jumps and Gaussian components. This deconstruction of the dynamics of log-prices allows us to propose for the first time the realized number of jumps at time t, denoted by NJ_t, as a new jump activity measure to explain and to forecast the volatility of the SPY.

We summarize our main empirical findings when the number of jumps in the SPY is employed to explain and forecast the volatility of log-returns of the SPY.

The number of jumps in the SPY is an important variable in explaining the daily volatility of the SPY log-returns. In our study, we obtain daily volatility estimates with four different estimators: MLE, BPV, $TBPV$ and our $MLE - F$. We find that the number of jumps in the SPY has more explanatory power with respect to daily volatility than other widely used variables such as: the volume of trades, the number of trades, and the ratio between open and close.

Furthermore, we show that our jump activity measure NJ_t has more explanatory power than the well studied jump activity measure J_t, which is based on the difference between the quadratic variation of the log-prices and the Bipower Variation estimator of Barndorff-Nielsen and Shephard (2004).

Our results also show that the number of jumps NJ_t has a similar explanatory power than that of the VIX, but slightly lower explanatory power than that of the high-low measure $HL_t = \log(\text{high}_t/\text{low}_t)$, when explaining daily volatility.

We further examine the question of whether the volume or the number of trades is the driving source of volatility. We show that the number of trades, and not the volume, is the significant variable in explaining volatility. This result agrees with Jones et al. (1994) and Ross (1989), where

27

it is documented that the number of trades, rather than the volume, is the driving source of asset price volatility.

We emphasize that in our analysis the number of jumps NJ_t is statistically significant in all models, even after the inclusion of combinations of all, or some, of the explanatory variables that we have mentioned above: volume of trades, number of trades, ratio between open and close, log-high minus log-low, VIX and the jump activity measure J_t.

Finally, we examine whether the number of jumps in the SPY log-returns help us to better forecast SPY monthly volatility.

First, we obtain monthly volatility estimates with the MLE-F volatility estimator. Then, we show that: the average number of jumps, for the month where volatility has been estimated explains monthly volatility; and the explanatory power of the realized average number of jumps is on a par with that of the VIX.

Second, since we do not know the average of the number of jumps over the period for which we want to forecast monthly volatility, we predict this average using an $MA(5)$. Next, we use these predictions to forecast monthly volatility and show that a model that includes these $MA(5)$ predictions exhibits lower forecasting errors than the models without these predictions. Finally, our results show that the expected jumps contain relevant information when forecasting monthly volatility which is not impounded in the VIX.

References

Aït-Sahalia, Y., Cacho-Diaz, J. and Hurd, T.: 2009, Portfolio choice with jumps: A closed form solution, *Ann. Appl. Probab* **19**(2), 556–584.

Aït-Sahalia, Y., Mykland, P. and Zhang, L.: 2005, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* **18**(2), 351–416.

Alizadeh, S., Brandt, M. and Diebold, F.: 2002, Range-based estimation of stochastic volatility models, *The Journal of Finance* **57**(3), 1047–1091.

Andersen, T. and Bollerslev, T.: 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* **39**(4), 885–905.

Andersen, T., Bollerslev, T. and Diebold, F.: 2003, Some like it smooth, and some like it rough: untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility, *working paper* .

Andersen, T., Bollerslev, T. and Diebold, F.: 2007, Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *The Review of Economics and Statistics* **89**(4), 701–720.

Bakshi, G., Carr, P. and Wu, L.: 2008, Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies, *Journal of Financial Economics* **87**(1), 132–156.

Bali, T. and Demirtas, O.: 2008, Testing mean reversion in financial market volatility: Evidence from S&P 500 index futures, *Journal of Futures Markets* **28**(1), 1–33.

Barndorff-Nielsen, O., Hansen, P. and Lunde, A.: 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O. and Shephard, N.: 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* **2**(1), 1–37.

Barndorff-Nielsen, O. and Shephard, N.: 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* **4**(1), 1–30.

Becker, R., Clements, A. and McClelland, A.: 2009, The jump component of S&P 500 volatility and the VIX index, *Journal of Banking and Finance* **33**(6), 1033–1038.

Blair, B., Poon, S. and Taylor, S.: 2001, Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns, *Journal of Econometrics* **105**(1), 5–26.

Carr, P. and Madan, D.: 1999, Option valuation using the fast fourier transform, *Journal of Computational Finance* **2**(4), 61–73.

Carr, P. and Wu, L.: 2003, What type of process underlies options? a simple robust test, *The Journal of Finance* **58**(6), 2581–2610.

Carr, P. and Wu, L.: 2004, Time-changed Levy processes and option pricing, *Journal of Financial Economics* **71**(1), 113–141.

Carr, P. and Wu, L.: 2007, Stochastic skew in currency options, *Journal of Financial Economics* **86**(1), 213–247.

Christensen, K., Oomen, R. and Podolskij, M.: 2009, Realised quantile-based estimation of the integrated variance. Working Paper.

Clark, P.: 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**(1), 135–155.

Corsi, F., Pirino, D., Renò, R., Francesco, P., Buffi, V. and Pontecorvo, L.: 2008, Volatility forecasting: the jumps do matter, *working paper* .

Epps, T. and Epps, M.: 1976, The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, *Econometrica* **44**(2), 305–321.

Fleming, J.: 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* **5**(4), 317–345.

Gallant, A., Rossi, P. and Tauchen, G.: 1992, Stock prices and volume, *The Review of Financial Studies* **5**(2), 199–242.

Garman, M. and Klass, M.: 1980, On the estimation of security price volatilities from historical data, *Journal of Business* pp. 67–78.

Gatheral, J. and Oomen, R.: 2009, Zero-intelligence realized variance estimation. Forthcoming in Finance and Stochastics.

Jacod, J. and Todorov, V.: 2009, Do price and volatility jump together? Forthcoming in Annals of Applied Probability.

Jiang, G. and Tian, Y.: 2005, The model-free implied volatility and its information content, *Review of Financial Studies* **18**(4), 1305–1342.

Jones, C., Kaul, G. and Lipson, M.: 1994, Transactions, volume, and volatility, *Review of Financial Studies* pp. 631–651.

Karpoff, J.: 1987, The relation between price changes and trading volume: A survey, *The Journal of Financial and Quantitative Analysis* **22**(1), 109–126.

Lee, S. and Hannig, J.: 2009, Detecting jumps from Lévy jump diffusion processes, *working paper* .

Lee, S. and Mykland, P.: 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial studies* **21**(6), 2535.

Maheu, J. and McCurdy, T.: 2009, Do high-frequency measures of volatility improve forecasts of return distributions?, *Fortcoming Journal of Econometrics* .

Mancini, C.: 2007, Non parametric threshold estimation for models with stochastic diffusion coefficients and jumps, *Forthcoming in Scandinavian Journal of Statistics* .

Merville, L. and Pieptea, D.: 1989, Stock-price volatility, mean-reverting diffusion, and noise, *Journal of Financial Economics* **24**(1), 193–214.

Newey, W. and West, K.: 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* pp. 703–708.

Oomen, R.: 2006, Properties realized variance under alternative sampling schemes, *Journal of Business & Economic Statistics* **24**(2), 219–237.

Rogers, L. and Satchell, S.: 1991, Estimating variance from high, low and closing prices, *The Annals of Applied Probability* **1**(4), 504–512.

Ross, S.: 1989, Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy, *Journal of Finance* pp. 1–17.

Szakmary, A., Ors, E., Kim, J. and Davidson III, W.: 2003, The predictive power of implied volatility: Evidence from 35 futures markets, *Journal of Banking and Finance* **27**, 2151–2175.

Tauchen, G. and Pitts, M.: 1983, The price variability-volume relationship on speculative markets, *Econometrica* pp. 485–505.

Todorov, V.: 2009, Variance risk premium dynamics: The role of jumps. Forthcoming in Review of Financial Studies.

Todorov, V. and Tauchen, G.: 2009, Volatility jumps. Working Paper.

Verardo, M. and Patton, A.: 2009, Does beta move with news? systematic risk and firm-specific information flows, *working paper* .

Yang, D. and Zhang, Q.: 2000, Drift-independent volatility estimation based on high, low, open, and close prices, *The Journal of Business* **73**(3), 477–492.

Zhang, L., Mykland, P. and Aït-Sahalia, Y.: 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**, 1394–1411.

Tables and Figures

Table 1: Correlation matrix of volatility estimates from high and low frequency estimators and VIX

	MLE-F	MLE	BPV	$TBPV$	RV_{5min}	$GARCH$	$\lvert r_t \rvert$	VIX
MLE-F	1	0.96	0.86	0.97	0.80	0.83	0.77	0.83
MLE	0.96	1	0.85	0.94	0.85	0.79	0.73	0.80
BPV	0.86	0.85	1	0.89	0.85	0.73	0.68	0.73
$TBPV$	0.97	0.94	0.89	1	0.84	0.84	0.77	0.83
RV_{5min}	0.80	0.85	0.85	0.84	1	0.69	0.64	0.69
$GARCH$	0.83	0.79	0.73	0.84	0.69	1	0.96	0.89
$\lvert r_t \rvert$	**0.77**	0.73	0.68	0.78	0.64	**0.96**	1	0.86
VIX	0.83	0.80	0.73	0.83	0.69	0.89	0.86	1

Table 2: $AR(1)$ model for volatility estimators

Volatility estimator		Coefficient	Std. Error	Prob	adjusted R^2	$\log L$
$\log MLE$	c	0.376	0.033	0.000	0.724	-42.09
	$\log MLE(-1)$	**0.851**	0.013	0.000		
$\log BPV$	c	0.430	0.041	0.000	0.699	-298.44
	$\log BPV(-1)$	**0.837**	0.015	0.000		
$\log TBPV$	c	0.203	0.019	0.000	0.847	359.59
	$\log TBPV(-1)$	**0.921**	0.007	0.000		
$\log MLE - F$	c	0.275	0.024	0.000	0.792	235.80
	$\log MLE - F(-1)$	**0.890**	0.009	0.000		

Table 3: Incorporating the jump activity measures

	log MLE			log BPV			log TBPV			log MLE − F		
c	0.369	0.309	0.310	0.411	0.281	0.286	0.205	0.173	0.173	0.279	0.254	0.249
	(0.030)	(0.027)	(0.027)	(0.040)	(0.035)	(0.035)	(0.019)	(0.017)	(0.017)	(0.024)	(0.022)	(0.022)
$\log \sigma_{t-1}$	0.853	0.871	0.870	0.843	0.885	0.884	0.919	0.927	0.927	0.889	0.893	0.895
	(0.012)	(0.011)	(0.011)	(0.015)	(0.013)	(0.014)	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
J_t	908.02	-	**146.05**	1394.02	-	**213.21**	432.55	-	**-135.99**	**144.96**	-	-369.74
	(283.84)		(334.51)	(416.25)		(395.24)	(137.03)		(122.45)	(116.69)		(115.62)
J_{t-1}	-595.31	-	*146.05*	-881.70	-	*-1.520*	*-213.47*	-	*149.04*	*32.13*	-	354.50
	(176.01)		(243.18)	(164.51)		(245.08)	(247.56)		(230.58)	(102.67)		(103.74)
NJ_t	-	0.047	**0.045**	-	0.072	**0.069**	-	0.032	**0.034**	-	0.026	**0.031**
		(0.004)	(0.004)		(0.007)	(0.007)		(0.003)	(0.003)		(0.004)	(0.003)
NJ_{t-1}	-	-0.030	-0.029	-	-0.054	-0.054	-	-0.019	-0.020	-	-0.014	-0.018
		(0.004)	(0.005)		(0.007)	(0.008)		(0.003)	(0.004)		(0.003)	(0.004)
adjusted R^2	0.737	0.777	**0.777**	0.724	0.809	**0.809**	0.849	0.868	**0.868**	0.792	0.808	**0.809**
Akaike info	0.002	-0.162	-0.161	0.258	-0.108	-0.108	-0.423	-0.554	-0.554	-0.267	-0.344	-0.355
Schwarz	0.015	-0.149	-0.141	0.271	-0.096	-0.088	-0.410	-0.541	-0.535	-0.254	-0.332	-0.336
Hannan-Quinn	0.007	-0.157	-0.154	0.263	-0.103	-0.100	-0.418	-0.549	-0.547	-0.262	-0.339	-0.348
$\log L$	1.96	**145.13**	**146.15**	-221.29	**98.39**	**99.95**	373.12	**487.47**	**489.66**	236.89	304.38	**315.88**

Table 4: The incremental information of VIX, Open-Close and High-Low

Independent Variables	$\log MLE$	$\log BPV$	$\log TBPV$	$\log MLE - F$
c	2.707	1.781	2.619	3.350
	(0.496)	(0.580)	(0.381)	(0.448)
$\log \sigma_{t-1}$	0.384	0.497	0.609	0.458
	(0.024)	(0.037)	(0.023)	(0.022)
NJ_t	0.031	0.061	0.022	0.012
	(0.003)	(0.007)	(0.002)	(0.004)
NJ_{t-1}	-0.012	-0.029	-0.011	***-0.005***
	(0.004)	(0.006)	(0.003)	(0.003)
$\log HL_t$	0.537	0.354	0.357	0.489
	(0.056)	(0.049)	(0.042)	(0.053)
OC_t	-1.199	***-0.973***	-1.721	-1.968
	(0.439)	(0.546)	(0.354)	(0.394)
$\log VIX_t$	0.413	0.431	0.299	0.537
	(0.043)	(0.051)	(0.035)	(0.039)
adjusted R^2	**0.874**	**0.863**	**0.908**	**0.888**
Akaike info	-0.732	-0.435	-0.911	-0.883
Schwarz	-0.710	-0.413	-0.888	-0.861
Hannan-Quinn	-0.724	-0.426	-0.903	-0.875
$\log L$	646.11	386.50	802.09	785.16

Table 5: Volume vs. Number of Trades

	log MLE			log BPV			log TBPV			log MLE − F		
c	2.578	2.421	2.420	1.947	1.956	1.764	2.606	2.523	2.460	3.222	3.055	3.048
	(0.512)	(0.520)	(0.508)	(0.584)	(0.596)	(0.605)	(0.389)	(0.402)	(0.396)	(0.462)	(0.472)	(0.461)
$\log \sigma_{t-1}$	0.389	0.386	0.386	0.478	0.489	0.476	0.611	0.613	0.608	0.465	0.461	0.460
	(0.024)	(0.025)	(0.003)	(0.038)	(0.037)	(0.038)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
NJ_t	0.031	0.031	0.031	0.061	0.061	0.061	0.022	0.021	0.022	0.012	0.012	0.012
	(0.003)	(0.003)	(0.004)	(0.007)	(0.007)	(0.007)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
NJ_{t-1}	-0.012	-0.012	-0.012	-0.028	-0.029	-0.028	-0.011	-0.011	-0.011	***-0.006***	***-0.006***	***-0.006***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\log HL_t$	0.518	0.518	0.518	0.384	0.368	0.383	0.354	0.349	0.353	0.471	0.470	0.470
	(0.058)	(0.057)	(0.059)	(0.048)	(0.046)	(0.048)	(0.045)	(0.043)	(0.045)	(0.055)	(0.054)	(0.056)
OC_t	-1.203	-1.117	-1.117	-0.941	***-1.011***	***-0.845***	-1.726	-1.703	-1.646	-1.975	-1.884	-1.878
	(0.443)	(0.445)	(0.448)	(0.545)	(0.545)	(0.555)	(0.354)	(0.357)	(0.358)	(0.397)	(0.399)	(0.401)
$\log VIX_t$	0.434	0.459	0.460	0.413	0.407	0.443	0.301	0.313	0.326	0.377	0.405	0.406
	(0.047)	(0.050)	(0.048)	(0.051)	(0.054)	(0.056)	(0.037)	(0.042)	(0.040)	(0.043)	(0.047)	(0.045)
VOL_t	0.000	-	***0.000***	0.000	-	-0.000	***0.000***	-	***0.000***	0.000	-	***0.000***
	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)
NT_t	-	0.054	***0.055***	-	***-0.037***	0.067	-	***0.020***	0.052	-	0.056	0.060
		(0.018)	(0.029)		(0.022)	(0.027)		(0.017)	(0.023)		(0.017)	(0.026)
adjusted R^2	0.874	**0.875**	**0.875**	0.863	0.863	**0.864**	0.907	**0.908**	**0.908**	0.888	0.888	**0.888**
Akaike info	-0.736	-0.737	-0.736	-0.441	-0.436	-0.442	-0.910	-0.911	-0.911	-0.888	-0.890	-0.890
Schwarz	-0.711	-0.712	-0.708	-0.416	-0.411	-0.414	-0.885	-0.886	-0.882	-0.863	-0.865	-0.861
Hannan-Quinn	-0.726	-0.728	-0.726	-0.432	-0.426	-0.432	-0.900	-0.901	-0.900	-0.878	-0.881	-0.878
$\log L$	650.13	**651.80**	**651.79**	392.86	**388.36**	**394.71**	802.17	**803.02**	**803.97**	783.00	**785.30**	**785.31**

Table 6: Forecasting average monthly number of jumps

	$NJ_{t \to t+22}$		
	Coefficient	Std. Error	Prob
c	1.091	0.032	
$MA(1)$	1.056	0.032	0.000
$MA(2)$	0.963	0.040	0.000
$MA(3)$	0.919	0.041	0.000
$MA(4)$	0.832	0.038	0.000
$MA(5)$	0.466	0.030	0.000
Adjusted R^2	**0.900**		

Table 7: Adjusted R^2 values for volatility driving sources (in percentages)

$\log \sigma_t = c + \beta_1 \log \sigma_{t-1} +$	u_t	$\beta_2 \log HL_t + u_t$	$\beta_2 OC_t + u_t$	$\beta_2 \log VIX_t + u_t$	$\beta_2 NJ_t + u_t$	$\beta_2 NT_t + u_t$	$\beta_2 VOL_t + u_t$
MLE	72.44	83.23	72.90	79.34	76.20	72.53	72.62
BPV	69.97	77.92	70.17	76.94	77.24	70.47	69.95
$TBPV$	84.69	88.77	85.05	87.35	86.25	84.69	84.89
$MLE\text{-}F$	79.22	86.55	79.90	83.72	80.43	79.22	79.45

Table 8: Forecasting monthly volatility

| Independent Variables	$\log MLE - F_{t\rightarrow t+22}$	$\log MLE - F_{t\rightarrow t+22}$	$\log MLE - F_{t\rightarrow t+22}$	$\log MLE - F_{t\rightarrow t+22}$
c	*-0.192*	0.142	-0.355	-0.355
	(0.104)	(0.068)	(0.097)	(0.097)
$\log MLE - F_{t-22\rightarrow t}$	0.483	0.855	0.487	0.487
	(0.073)	(0.026)	(0.052)	(0.052)
$\log VIX_{t-1}$	0.517	-	0.497	0.497
	(0.086)		(0.067)	(0.067)
$\overline{NJ}_{t\rightarrow t+22}$	-	0.200	0.196	-
		(0.020)	(0.020)	
$\widehat{\overline{NJ}}_{t\rightarrow t+22}$	-	-	-	0.196
				(0.020)
adjusted R^2	0.783	**0.806**	**0.835**	**0.835**
Akaike info	-0.423	-0.532	**-0.692**	**-0.693**
Schwarz	-0.413	-0.522	**-0.680**	**-0.680**
Hannan-Quinn	-0.419	-0.528	**-0.688**	**-0.688**
$\log L$	363.29	455.58	**594.06**	**594.07**
Forecasting statistics			with $\widehat{\overline{NJ}}_{t\rightarrow t+22}$	
Root Mean Squared Error	0.249			**0.237**
Mean Absolute Error	0.210			**0.194**
Mean Absolute Percent Error	2.793			**2.591**
Theil Inequality Index	0.017			**0.016**
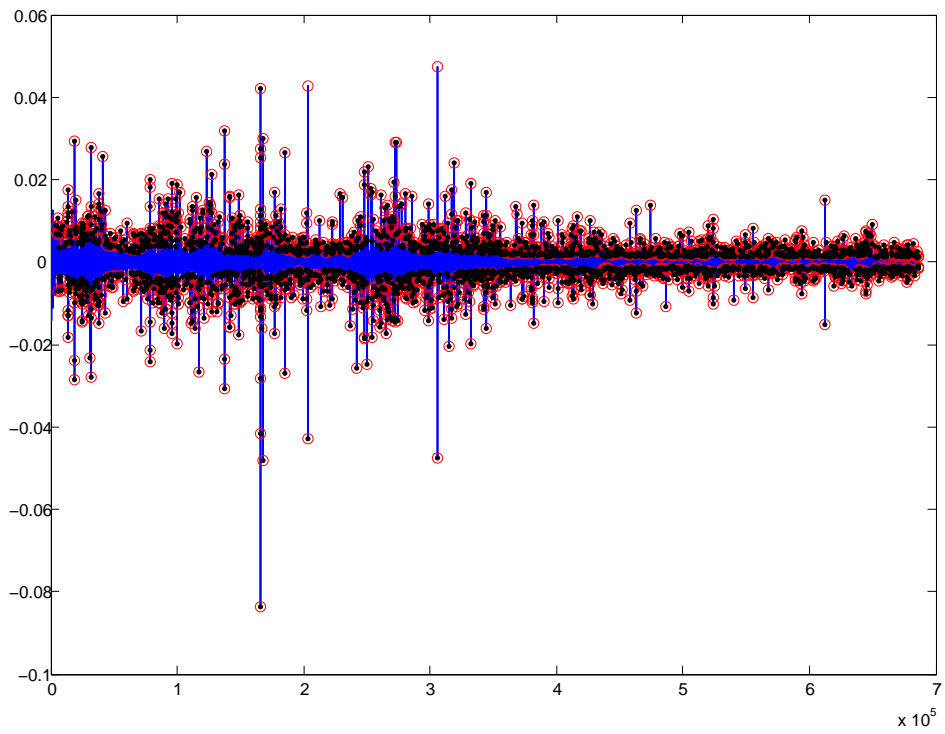
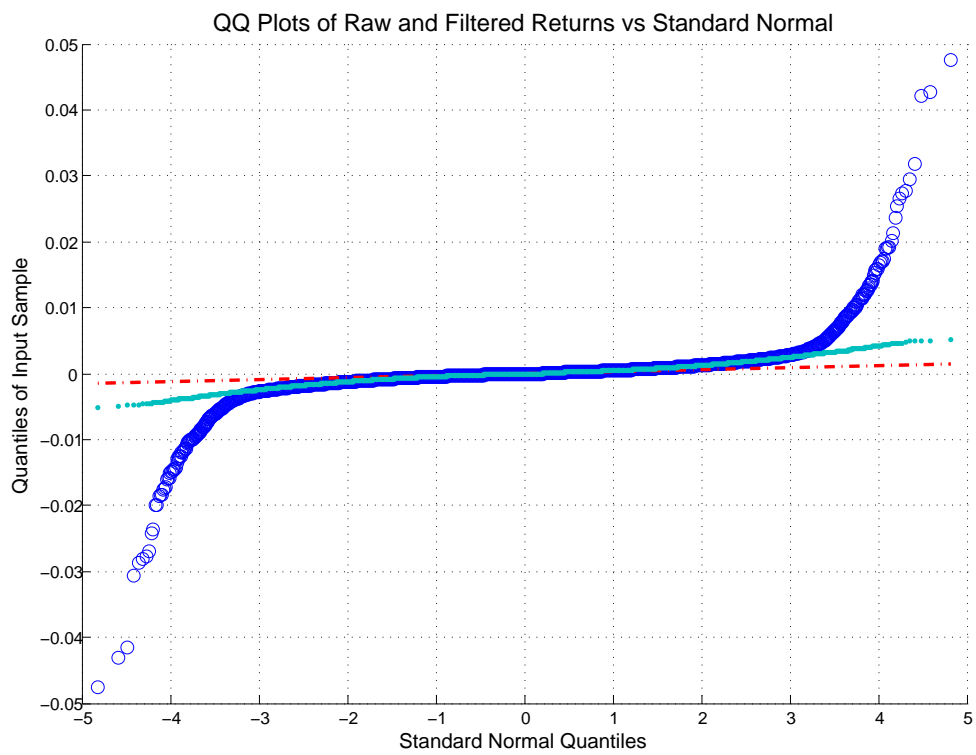Figure 1.   S&P 500 Lévy and Poisson-type jumps.

Figure 2. QQ plot for raw (blue circles) and filtered (light blue dots) SPY returns
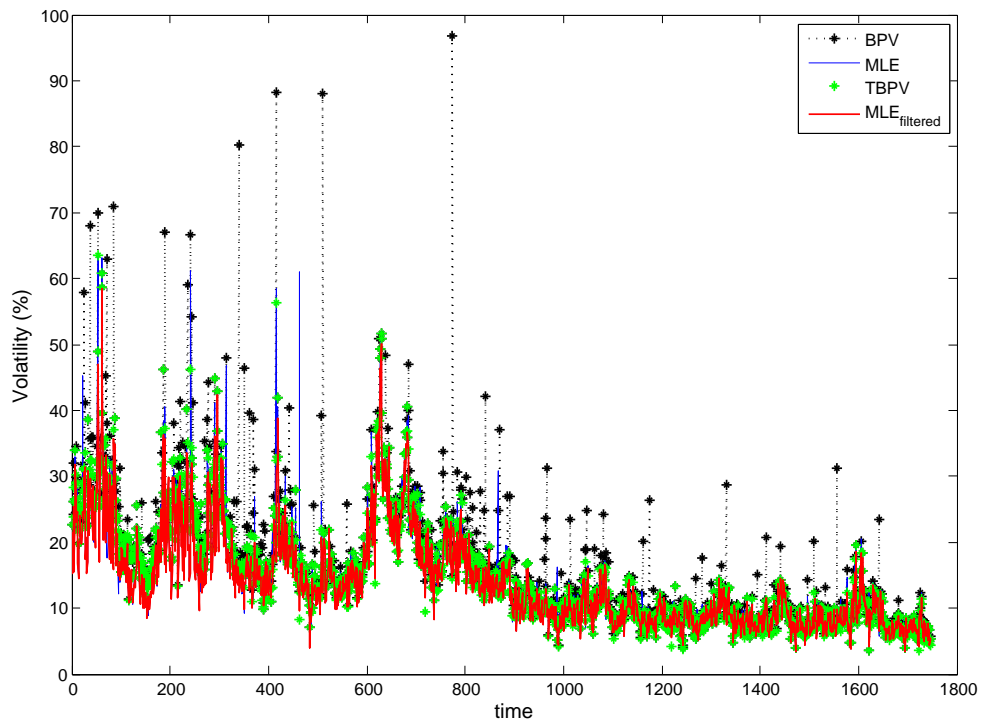
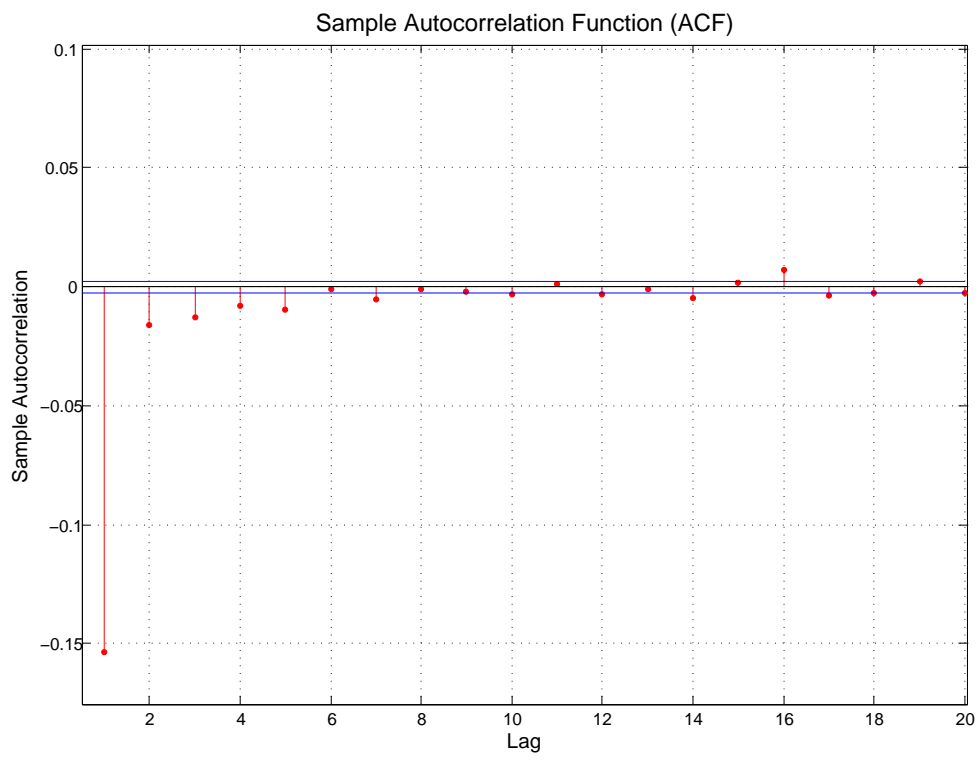**Figure 3.** **SPY daily volatilities using the** MLE, BPV, $TBPV$ **and** MLE-F.

Figure 4. **SPY returns autocorrelogram**