Estimating stochastic volatility and jumps using high-frequency data and Bayesian methods[1]

MILAN FIČURA[2], JIŘÍ WITZANY[3]

Abstract: We are comparing two approaches for stochastic volatility and jumps estimation in the EUR/USD time series - the non-parametric power-variation approach using high-frequency returns, and the parametric Bayesian approach (MCMC estimation of SVJD models) using daily returns. We find that both of the methods do identify continuous stochastic volatility similarly, but they do not identify similarly the jump component. Firstly - the jumps estimated using the non-parametric high-frequency estimators are much more numerous than in the case of the Bayesian method using daily data. More importantly - we find that the probabilities of jump occurrences assigned to every day by both of the methods are virtually no rank-correlated (Spearman rank correlation is 0.0148) meaning that the two methods do not identify jumps at the same days. Actually the jump probabilities inferred using the non-parametric approach are not much correlated even with the daily realized variance and the daily squared returns, indicating that the discontinuous price changes (jumps) observed on high-frequencies may not be distinguishable (from the continuous volatility) on the daily frequency. As an additional result we find strong evidence for jump size dependence and jump clustering (based on the self-exciting Hawkes process) of the jumps identified using the non-parametric method (the shrinkage estimator).

AMS/JEL classification: C11, C14, C15, C22, G1

Keywords: Stochastic volatility, Bayesian inference, MCMC, Realized variance, Bipower variation, Shrinkage estimator, Jump clustering, Self-Exciting jumps, Hawkes process

Introduction

Numerous studies have shown that the asset price processes exhibit properties like the fat tails of the return distribution, long-term shifts in the levels of volatility, abrupt changes in price and volatility as well as multifractal scaling of return moments. In order to tackle these empirical properties of the price dynamics it has proved useful to decompose the variability of the price process into two components - diffusive stochastic volatility and discontinuous price jumps (Craine et. al. 2000, Eraker 2004, Witzany 2013).

Stochastic volatility, which governs the variability of the asset price diffusion, follows for most assets a very persistent mean-reverting process, generating long-terms changes in the magnitude of asset returns - co called volatility clustering (for further discussion see Stádník 2014). This increases the kurtosis of the long-horizon conditional return distribution and the fatness of its tails, with important implications for long-maturity derivatives pricing (see Eraker et al. 2003 or Fulop, Li and Yu 2014).

[1] The research has been supported by the Czech Science Foundation Grant P402/12/G097 Dynamical Models in Economics

[2]University of Economics, Faculty of Finance and Accounting, Department of Banking and Insurance, Winston Churchill Square 4, 130 67, Prague 3, Czech Republic, E-mail: xficm03@vse.cz.

[3]University of Economics, Faculty of Finance and Accounting, Department of Banking and Insurance, Winston Churchill Square 4, 130 67, Prague 3, Czech Republic, E-mail: jiri.witzany@vse.cz.

The jump component, on the other hand, adds large instantaneous changes to the price process (jumps), which are often correlated with macroeconomic news announcements (but may not be, see Andersen et al. 2007). Statistical characteristics of this abrupt price movements are usually much less persistent than the changes in stochastic volatility, although some dependences in the intensity of the jump occurrence (jump clustering) and their absolute magnitudes do exists (Andersen et al. 2007, Fulop, Li and Yu 2014). Compared to the stochastic volatility of the diffusion process, jumps increase the fatness of the tails of the conditional return distribution mainly in short horizons, where the stochastic volatility is not able to explain the full magnitude of the extreme instantaneous price movements (for analysis of the effects of jumps in price as well as volatility see Eraker et al. 2003).

Models using the decomposition of the price variability into stochastic volatility and jumps - so called Stochastic-Volatility-Jump-Difusion models (SVJD) - have achieved very good performance in a wide area of financial applications where modeling of the conditional return distribution is of interest (for more about these models see Shephard 2005).

One of the most fruitful areas of parametric SVJD models applications is the area of option pricing. As can be seen empirically from the existence of volatility smiles, the models with constant volatility - like the Black-Scholes model with geometric Brownian motion (GBM) as its underlying price process - greatly undervalue options that are far from the money, because the GBM process underestimates the probability of extreme price movements. This can be partially solved by adding stochastic volatility to the price process as in the case the Heston model (Heston 1993). But as has been shown in empirical studies, both of the variability components - stochastic volatility as well as jumps - are necessary to explain the empirically observed option prices (see Eraker et al. 2003, Eraker 2004 or Fulop Li and Yu 2014). Stochastic volatility of the diffusion component plays role mainly in long-maturity far-out-of-the-money option pricing, while the jump component is crucial in particular when short-maturity far-out-of-the-money options are priced.

Another possible application of the SVJD models is in the area of risk management where measures of risk such as expected volatility, Value at Risk (VaR) or Expected Shortfall (ES) need to be estimated.

Value at Risk is defined as some low percentage quantile (typically 5%, 1% or 0,1%) of the conditional return distribution. The size of VaR is thus heavily dependent on the expected fatness of the tails of the return distribution. As it has been shown in Witzany (2013), the incorporation of stochastic volatility into the underlying price process plays crucial role mainly for long horizon VaR estimation, while the incorporation of jumps is important mainly for short-horizon, high probability (i.e. very low quantile) VaR estimation.

Recently, separate modeling of continuous volatility and jumps has become an area of interest also in the area of volatility forecasting (Andersen et al. 2007, Lanne 2006, Corsi et al. 2010). The resulting models utilize the fact that both of the components exhibit entirely different temporal dynamics with different levels of persistence. So it has been shown that the diffusive stochastic volatility can best be modeled using long-memory autoregressive models (HAR as in Corsi et al. 2010 or ARFIMA as in Ishida and Watanabe 2009), while the jump component is better tackled using a far less persistent processes (in the continuous time setting a recent

advancement was to use the Hawkess process, see Chen and Poon 2013 or Fulop, Li and Yu 2014).

Finally the separate modeling of stochastic volatility and jumps can be used for volatility risk premium estimation - i.e., the premium of the option implied volatility compared to the expected future realized volatility. As has been shown in numerous studies the value of the volatility risk premium heavily influences the expected returns of option strategies as well as volatility derivatives (Bakshi and Kapadia 2003, Eraker 2009 or Carr and Wu 2009). Interestingly the value of the volatility premium depends differently on the expected diffusion volatility and the expected jump volatility (Todorov 2009, Chen and Poon 2013).

The main challenge in stochastic volatility and jumps modeling is to disentangle their appropriate contributions to the overall return variability in the past. Unfortunately the only variable that we can directly observe on the market is the asset price sampled at some discrete points in time. The process of the stochastic volatility, as well as the processes governing the occurrence, size and intensity of the jumps are unobservable (i.e., latent) and they have to be estimated (which is usually possible only with a limited level of precision).

The lack of efficient estimation methods for SVJD models has hindered the past development of associated applications that could fully utilize their analytical power. Fortunately, in the last two decades two highly efficient groups of estimation techniques emerged - Bayesian methods used for estimation of parametric SVJD models and high-frequency power-variation estimators used to estimate stochastic volatility and jumps non-parametrically (for a review of the historical developments see Andersen et al. 2005).

The first group of techniques is based on Bayesian inference methods - using extensive simulations - in order to estimate the parameters as well as the latent state variables (i.e., the latent processes) of parametrically specified SVJD models. Among the most popular techniques in this category is the MCMC algorithm (Eraker et al. 2003, Johannes and Polson 2009, Witzany 2013, etc.), the EMM method (Gonzales et al. 2011), as well as a group of methods called particle filters (Fulop, Li and Yu 2014). All of the techniques have traditionally been applied to daily returns but they can be possibly used with intraday returns as well. A great advantage of this group of methods is that they are able to estimate the latent state variables as well as the parameters of the specified stochastic processes, which can then be used to analytically express the conditional future return distribution or to approximate it arbitrarily well using Monte-Carlo simulations of the future price evolution. This has proved to be useful in the area of option pricing, especially for exotic options, where the simulation of future price paths may often pose the only practical way for how to price them.

The greatest disadvantage of the Bayesian methods and the parametric SVJD models is their high computational intensity which is probably the main reason why they are less popular in some empirical applications - like volatility forecasting - than other families of models (GARCH models, realized volatility models, option volatility models, etc.), for discussion see Andersen et al. 2005).

The other set of techniques used for volatility and jump estimation is based on the recently developed asymptotic distribution theory for power-variation measures (Barndorff-Nielsen

and Shephard 2004) and it estimates the latent stochastic volatilities and jump components non-parametrically using estimators calculated from high-frequency data. A great advantage of these methods (in certain applications) is that there is no need to specify any parametric processes for the stochastic volatility and jumps and the methods should be theoretically be valid for a wide range of possible processes (i.e. the methods are model-free).

The primary idea of these methods is to utilize the properties of power-variation measures which when sampled at sufficiently fine frequencies (limiting to infinity) should converge towards either the quadratic variation - which is a measure of the overall price variability over a given period of time, or to the integrated variance - which is a measure of the diffusion variability component. The jump component can then easily be estimated as the difference between the estimates of the quadratic variation and the integrated variance.

The first of these estimators was the realized variance introduced by Andersen and Bollerslev (1998) which converges - as the sampling frequency goes to zero - to the quadratic variance of the price process. In the subsequent years many similar estimators emerged, converging either to the quadratic variance - realized kernels (see Barndorff-Nielsen et al. 2008) - or to the integrated variance - bipower variation (introduced by Barndorff-Nielsen and Shephard 2004), multi-power variation measures (see Ysusi 2006 and Shi 2009) or nearest-neighbor truncation measures (introduced by Andersen et al. 2010).

So the researchers and practitioners are now possessing two efficient methods for stochastic volatility and jump estimation, and the natural question is, if these methods do provide similar results - in general - or if there are some systematic differences between them. The previous studies indicate that the Bayesian parametric methods (using daily data) usually identify significantly less jumps than the high-frequency power-variation methods (for discussion see Andersen et al. 2007). But is this the only difference?

In order to examine this question in more detail we applied a representative method from both of the estimation frameworks to estimate the stochastic volatility and jumps in the past history of EUR/USD exchange rate (during the period between 3.2.2009 and 15.4.2014) and we compared the results.

As a representative method for the parametric Bayesian approach we utilized the log-variance SVJD model developed in Witzany (2013), with the slight difference that instead of Poisson jumps we worked with Self-Exciting jumps governed by a Hawkes process in order to tackle the possible jump clustering. The parameters of the model were estimated using a complex MCMC algorithm combining the Gibbs Sampler and Metropolis-Hastings algorithm (as in Witzany 2013) with the addition of additional Random-Walk Metropolis-Hastings step for the estimation of the parameters of the Hawkes process.

As a representative method for the non-parametric power-variation approach we used the realized variance as an estimator of the quadratic variation (introduced by Andersen and Bollerslev 1998), bipower variation as an estimator of the integrated variance (introduced by Barndorff-Nielsen and Shephard 2004) and the so called shrinkage estimator (developed in Andersen et al. 2007) for the estimation of the statistically significant jumps. In order to

analyze jump clustering we further used the Hawkes process (as in the previous case) applied on the jumps identified beforehand using the shrinkage estimator.

Surprisingly the results of our research indicate that the two methods do not identify jumps at the same times at all, respectively that the inferred probabilities of jump occurrences are not correlated between the two methods (using the Spearman rank correlation coefficient). We attribute these results to the fact that both of the methods use entirely different information source to identify the jumps (the Bayesian method uses daily returns, while the non-parametric method uses 15 minute returns).

In accord with other empirical studies we found that the non-parametric power-variation approach identifies significantly more jumps than the Bayesian approach. This is in line with the intuitive notion that it should be easier to identify discontinuous price changes from 15 minute returns than from daily returns where they mix together with the diffusion.

Finally, we found that when estimated using the shrinkage estimator the jumps do exhibit certain clustering (based on the Hawkess process), no-matter what significance levels we use for the calculation of the shrinkage estimator. We also found evidence of clustering of the sizes of the jumps. Using the Bayesian approach however, these effects were not present.

The rest of the paper is organized as follows. In the first chapter we present the general price process and the non-parametric estimators used to estimate its quadratic variation and integrated variance. In the second chapter we present the parametric SVJD model with self exciting jumps and the Bayesian methods used for the estimation of its parameters and latent state variables. Finally, in the third chapter there is the empirical research on the EUR/USD exchange rate where we compare the results of the two estimation approaches. In the conclusion we sum up the most important results of the research.

1 Non-parametric estimation of volatility and jumps

Let us assume that the logarithmic price of an asset follows a general Stochastic-Volatility Jump-Diffusion process defined by the following stochastic differential equation:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + j(t)dq(t) \tag{1}$$

where $\mathrm{p(t)}$ is the logarithm of the asset price, $\mu(\mathrm{t})$ is the instantenous drift rate, $\sigma(\mathrm{t})$ is the instentenous volatility, $\mathrm{W(t)}$ is the Wiener process, $\mathrm{j(t)}$ is a process determining the size of the jumps and $\mathrm{q(t)}$ is a counting process whose differential determines the times of jump occurrence and is often called the jump indicator.

The logarithmic return over a given time period between $\mathrm{t}-1$ and t (in our case it will be one day) can be expressed as:

$$r(t) = p(t) - p(t-1) = \int_{t-1}^{t} \mu(\tau)d\tau + \int_{t-1}^{t} \sigma(\tau)dW(\tau) + \sum_{t-1\leq\tau<t} \kappa(\tau) \tag{2}$$

where $\kappa(t) = j(t)I[q(t) = 1]$, $I(.)$ is the indicator function and the sum of $\kappa(t)$ measures the impact of jumps during the given period of time (from $t-1$ to t).

The impact of the potentially stochastic drift rate $\mu(t)$ is often omitted (especially when working with exchange rates). The daily return then reduces to:

$$r(t) = \int_{t-1}^{t} \sigma(\tau) dW(\tau) + \sum_{t-1 \le \tau < t} \kappa(\tau) \tag{3}$$

The total variability of the stochastic process over a given period of time can then be measured using its quadratic variation in the following form:

$$QV(t) = \int_{t-1}^{t} \sigma^2(s) ds + \sum_{t-1 \le s < t} \kappa^2(s) \tag{4}$$

where the first term - representing the continuous component of price variability - is called *integrated variance*, and the second term - representing the discontinuous component of price variability - is called *jump volatility*. So we can write:

$$QV(t) = IV(t) + JV(t) \tag{5}$$

where $IV(t)$ is the integrated variance and $JV(t)$ the jump volatility.

Both of the measures, $IV(t)$ and $JV(t)$, are unobservable on the market and they have to be estimated. One possible approach is to parametrically define the processes $\sigma(t)$, $j(t)$ and $q(t)$ and estimate their parameters as well as the latent state variables (i.e. the latent processes) using Bayesian inference methods. This approach will be further described in chapter 2.

Another approach is to use the non-parametric high-frequency power-variation measures which - when sampled at a sufficiently high frequency - should converge either to the quadratic variation or to the integrated variance - no-matter what is the exact form of the underlying processes $\sigma(t)$, $j(t)$ and $q(t)$.

The most well-known non-parametric high-frequency estimator of quadratic variation is the *realized variance* which is defined - for a given frequency - as the sum of squared returns on some higher frequency (Andersen and Bollerslev 1998). So we can - for example - get a consistent and unbiased estimate of the daily quadratic variation simply by computing the squares of all of the 15 minute returns during that day and adding them together.

Generally - if we denote $\mathrm{r}(\mathrm{t}, \Delta)$ as the logarithmic return between $\mathrm{t} - \Delta$ and t, the realized variance can be defined as:

$$\mathrm{RV}(\mathrm{t}, \Delta) = \sum_{\mathrm{j}=1}^{1/\Delta} \mathrm{r}^2(\mathrm{t} - 1 + \mathrm{j}\Delta, \Delta), \tag{6}$$

and it holds that $\mathrm{RV}(\mathrm{t}, \Delta) \to \mathrm{QV}(\mathrm{t})$ as $\Delta \to 0$.

Although the realized variance should theoretically provide an unbiased and consistent estimate of the underlying quadratic variation, in practical applications it may be positively biased, because the returns on very high frequencies (tick, second, minute, etc.) are heavily

plagued by microstructure noise (discreteness of the price grid, bid ask bounce, etc.), inducing negative autocorrelation in these returns which violates the semi-martingale property that is necessary for the above convergence to hold (see Andersen et al. 2005).

To cope with the microstructure noise problem it is possible to either use the returns on lower frequencies (5-30 minutes) where the autocorrelation is not an issue (but the resulting estimator becomes more noisy), or to use more advanced estimators, such as different sub-sampling schemes (Zhang, Mykland and Ait-Sahalia 2005) or realized kernels (Barndorff-Nielsen et al. 2008), which should provide robust estimates of the underlying quadratic variation even in the presence of microstructure noise.

In order to further decompose the estimated quadratic variation into the appropriate contributions of integrated variance and jump volatility, it is possible to use specific multi-power-variation measures which should converge - with increasing sampling frequency - directly to the integrated variance.

The first one and most well-known of these estimators is the realized bipower variation (Barndorff-Nielsen and Shephard 2004) which is defined as:

$$\mathrm{BV}(t,\Delta) = \frac{\pi}{2}\sum_{j=2}^{1/\Delta} |r(t-1+j\Delta,\Delta)||r(t-1+(j-1)\Delta,\Delta)|, \tag{7}$$

and it holds that $\mathrm{BV}(t,\Delta) \to \mathrm{IV}(t)$ as $\Delta \to 0$.

By utilizing the products of absolute returns for two consecutive periods instead of squared returns from a single period the jump volatility component becomes eliminated as the Δ goes to zero and the realized bipower variation converges to the integrated variance.

The contribution of the jump component can then be estimated as:

$$\mathrm{RJV}(t,\Delta) = \mathrm{RV}(t,\Delta) - \mathrm{BV}(t,\Delta) \tag{8}$$

where $\mathrm{RJV}(t,\Delta)$ is called realized jump volatility and $\mathrm{RJV}(t,\Delta) \to \sum_{t-1\le s<t} \kappa^2(s)$ as $\Delta \to 0$.

As long as we are not able to sample the absolute returns at infinitely fine frequency - which is never possible in practical applications - the estimates of jump volatility based on bipower variation are always plagued by some noise. This noise may cause the values of the estimator to be negative from time to time (actually if there were no jumps in the underlying price process the estimator would be negative in approximately 50% of the days), which is in contradiction with the idea that the estimator represents integrated variance which cannot be negative. Because of that it is reasonable to discard the negative values and estimate the jump contribution as:

$$\mathrm{RJV}(t,\Delta) = \max\{\mathrm{RV}(t,\Delta) - \mathrm{BV}(t,\Delta), 0\} \tag{9}$$

Consequently, in order to ensure that the estimates of the integrated variance and jump volatility sum into the estimate of the quadratic variation (realized variance) it is good to re-write the estimate of the integrated variance as follows:

$$\mathrm{EIV}(t,\Delta) = \mathrm{RV}(t,\Delta) - \max\{\mathrm{RV}(t,\Delta) - \mathrm{BV}(t,\Delta), 0\} \tag{10}$$

where $\mathrm{EIV}(t,\Delta)$ is the adjusted estimate of integrated variance.

Besides bipower variation, there were introduced - in the last few years - several more advanced estimators that should arguably provide faster convergence to the underlying integrated variance (resp. jump volatility). For alternative multi-power-variation measures see Ysusi (2006), Shi (2009) and Corsi et al. (2010). For an entirely new group of nearest-neighbor-truncation estimators see also Andersen et al. (2010).

A well known problem of the bipower variation, as well as of most of the other high-frequency estimators is, that they indicate a presence of jumps on almost every day. This is in contradiction with the believed notion that jumps appear only infrequently. Besides that, it is also probable that many of the alleged jumps are in fact erroneously identified due to the same reason as why some of the alleged jumps acquire negative values (i.e., due to the noise caused by the non-infinitely precise sampling frequency).

So it seems reasonable to work only with the statistically significant jumps. Using the asymptotic theory of power-variations (introduced in Barndorff-Nielsen and Shephard 2004), Andersen, Bollerslev and Diebold (2007) developed the so called shrinkage estimator for the jump contribution. It is based on the idea that appropriately normalized differences between realized variance and bipower variation should - in the absence of jumps - asymptotically converge to the standard normal distribution. For the normalization it is necessary to use the so called integrated quarticity:

$$\mathrm{IQ}(t) = \int_{t-1}^{t} \sigma^4(s)ds \tag{11}$$

which can be consistently estimated (even in the presence of jumps) using the realized tri-power quarticity:

$$\mathrm{TQ}(t,\Delta) = \frac{\pi^{3/2}}{4\Delta}\Gamma\left(\frac{7}{6}\right)^{-3}\sum_{j=3}^{1/\Delta}|r(t-1+j\Delta,\Delta)|^{4/3}|r(t-1+(j-1)\Delta,\Delta)|^{4/3}|r(t-1+(j-2)\Delta,\Delta)|^{4/3}, \tag{12}$$

i.e., it hold that $\mathrm{TQ}(t,\Delta) \to \mathrm{IQ}(t)$ as $\Delta \to 0$.

Using $\mathrm{RV}(t,\Delta)$, $\mathrm{BV}(t,\Delta)$ and $\mathrm{TQ}(t,\Delta)$ the authors define the variable $Z(t,\Delta)$ which asymptotically follows a standardized normal distribution as long as the the underlying process does not contain jumps:

$$Z(t,\Delta) = \frac{[\mathrm{RV}(t,\Delta) - B\mathrm{V}(t,\Delta)]\mathrm{RV}(t,\Delta)^{-1}}{\sqrt{[(\pi/2)^2 + \pi - 5]\max\{1, TV(t,\Delta)B\mathrm{V}(t,\Delta)^{-2}\}\Delta}} \tag{13}$$

Large values of $Z(t,\Delta)$ are thus indicating that a jump occurred during the given period. Therefore, the statistically significant jumps on a significance level α can be identified as:

$$\mathrm{RJV}(t,\Delta) = I\{Z(t,\Delta) > \Phi(\alpha)^{-1}\}[\mathrm{RV}(t,\Delta) - \mathrm{BV}(t,\Delta)] \tag{14}$$

where $I\{.\}$ is the indicator function and $\Phi(\alpha)^{-1}$ is the quantile function of the standard normal distribution. It is also worth noting that for $\alpha = 0.5$ the equation (14) becomes equal to the equation (9) and only the negative jumps become eliminated by the estimator.

Since we again want the jump estimator and integrated variance estimator to sum into realized variance, it is possible to re-estimate the integrated variance as:

$$\mathrm{EIV}(\mathrm{t},\Delta) = \mathrm{RV}(\mathrm{t},\Delta) - I\{Z(\mathrm{t},\Delta) > \Phi(\alpha)^{-1}\}[\mathrm{RV}(\mathrm{t},\Delta) - \mathrm{BV}(\mathrm{t},\Delta)] \quad (15)$$

In addition to that the shrinkage estimator further allows us to get the probabilities of jump occurrence on any given day by simply applying the standard normal cumulative distribution function to the variable $Z(\mathrm{t},\Delta)$. These probabilities can then be compared with the Bayesian probabilities of jump occurrences calculated using a parametric model.

2 Bayesian estimation of volatility and jumps

In order to estimate the continuous stochastic volatility and jumps in parametric model framework using Bayesian inference methods, we need to firstly specify the underlying latent processes for $\mu(\mathrm{t})$, $\sigma(\mathrm{t})$, $\mathrm{j}(\mathrm{t})$ and $\mathrm{q}(\mathrm{t})$ in equation (1) as well as the latent process $\lambda(\mathrm{t})$ which determines the intensity of jumps and thus governs the possible jump clustering. We will further present a parametric SVJD model inspired by Witzany 2013, with the addition of Self-Exciting Hawkes jumps instead of the independent Poisson jumps.

In the model we assume that the drift of the logarithmic return process is constant: $\mu(\mathrm{t}) = \mu$. The equation (1) then changes to:

$$dp(t) = \mu dt + \sigma(t)dW(t) + j(t)dq(t) \quad (16)$$

In order to model the instantaneous stochastic volatility $\sigma(t)$ we will utilize the log-variance model which uses the mean-reverting Ornstein-Uhlenbeck process for the logarithm of the return variance. The process can be written as follows:

$$dh(t) = \kappa[\theta - h(t)]dt + \xi dW_V(t) \quad (17)$$

where $h(t)$ is the logarithm of the return variance: $h(t) = \ln[\sigma^2(t)]$ and $W_V(t)$ is a separate Wiener process governing the evolution of the stochastic volatility which may or may not be correlated with the Wiener process for the logarithmic price $W(t)$ (we will further assume no correlation). Parameter θ represents the long-term level of volatility, κ determines the strength of the mean-reversion relationship and ξ is the volatility of volatility which we assume to be constant (although it may not be).

Considering the process of jump sizes $j(t)$, we will model it as a series of normally distributed independent random variables: $j(t) \sim N(\mu_J, \sigma_J)$. This means that we will omit the possible temporal dependence in jumps sizes, the correlation between the jump volatility and diffusion volatility as well as the fat tails of the jump distribution.

Finally, the process of jump occurrences $q(t)$ will be modeled using a self-exciting Hawkes process with the exponential decay function. In this process the intensity of jumps $\lambda(\mathrm{t})$, defined by the relationship $\Pr[\mathrm{dq}(\mathrm{t}) = 1] = \lambda(\mathrm{t})\mathrm{dt}$, increases (by a fixed amount) every time a jump occurs and then decays exponentially back to its long-term level. The jump intensity process $\lambda(\mathrm{t})$ can be expressed using the following differential equation which is deterministic in its relation to the process $q(t)$:

$$d\lambda(t) = \kappa_J\big[\theta_J - \lambda(t)\big]dt + \xi_J dq(t) \tag{18}$$

where θ_J is the long-term jump intensity, κ_J determines the rate of the exponential decay of the jump intensity towards its long-term level and ξ_J measures the immediate increase in jump intensity after a jump occurs (i.e., when $\mathrm{dq(t)} = 1$).

By solving the differential equation (16) it is possible to express the value of the jump intensity $\lambda(t)$ at any given point in time using the following relationship:

$$\lambda(t) = \theta_J + \int_{-\infty}^{t} \xi_J e^{-\kappa_J(t-s)} dq(s) = \theta_J + \sum_{dq(s)=1, s\le t} \xi_J e^{-\kappa_J(t-s)} \tag{19}$$

So we can see that every single jump that occurred in the past increases the jump intensity initially by amount ξ_J but its contribution decays over time by the exponential rate κ_J.

It is also worth noting that many recent studies (Eraker et al. 2003, Eraker 2004 or Fulop, Li and Yu 2014) add also the possibility of jumps in the volatility process which may be correlated with the price jumps, however, we do not include this feature.

The final SVJD model used in our study is specified by the equations (14), (15) and (16). In order to estimate the relevant parameters as well as the latent state variables it is necessary to convert the continuous-time model into the discrete time setting. For this purpose we will use the Euler discretization. Since the processes will be estimated using daily returns, the discretization will be done so that the units of time ($\mathrm{t} = 1,2,....$) represent days. It is further important to note that for the discretization of the jump intensity process it was necessary to apply a simplifying assumption that no more than one jump can occur during any given day. This assumption was of course not applied for the non-parametric methods of jump detection where unlimited amount of jumps could occur on any given day.

The discrete version of equation (14) governing the evolution of daily log-returns is:

$$r(t) = \mu + \sigma(t)\varepsilon(t) + J(t)Q(t) \tag{20}$$

where $r(t)$ is the daily logarithmic return defined as $r(t) = p(t) - p(t-1)$ and $p(t)$ is the logarithm of the closing price at day t. Parameter μ represents the conditional mean of the daily returns, $\sigma(t)$ is the daily stochastic volatility, $\varepsilon(t) \sim N(0,1)$ is a standard normal random variable, $J(t) \sim N(\mu_J, \sigma_J)$ is a normally distributed random variable determining the size of the jumps and $Q(t) \sim \mathrm{Bern}[\lambda(t)]$ is a variable following a Bernoulli process with intensity $\lambda(t)$.

The discrete version of equation (15) governing the daily stochastic volatility is:

$$h(t) = \alpha + \beta h(t-1) + \gamma \varepsilon_V(t) \tag{21}$$

where $h(t) = \ln[\sigma^2(t)]$ is the logarithm of the daily return variance, $\alpha = (1-\beta)\theta$ determines the long-term volatility, β is the autoregressive coefficient between today and yesterday log-variance, γ is the volatility of the volatility and $\varepsilon_V(t) \sim N(0,1)$ is a series of standard normal random variables which theoretically may be correlated with $\varepsilon(t)$ but in our model we will not include this possibility. As can be noticed from the equation (19), it does in fact represent the AR(1) process of the log-variance.

Finally, the discrete version of equation (16) governing the jump intensity $\lambda(t)$ is:

$$\lambda(t) = \alpha_J + \beta_J \lambda(t-1) + \gamma_J Q(t-1) \tag{22}$$

where $\lambda(t)$ is the jump intensity at day t, $\alpha_J = (1 - \beta_J)\theta_J$ determines the long-term jump intensity, β_J is the rate of the exponential decay of the jump intensity and γ_J gives the increase of intensity in the following day after a jump occurrence.

For notational simplicity we will further denote $V(t) = \sigma^2(t)$ as the daily stochastic variance which theoretically corresponds to the integrated variance from Section 2.

The final model in the discrete time setting has 3 equations (20, 21 and 22), with 9 parameters to be estimated ($\mu, \alpha, \beta, \gamma, \theta_J, \beta_J, \gamma_J, \mu_J, \sigma_J$) and 3 vectors of latent state variables: **V**, **J**, and **Q**.

To estimate the parameters of the model and the past values of the latent state variables we will use a Markov Chain Monte Carlo algorithm (MCMC) constructed according to Witzany (2013), based on the results in Jacquier et al. (2007) and Johannes and Polson (2009).

MCMC is a Bayesian estimation method that enables us to sample from high-dimensional multivariate densities by constructing a Markov Chain that converges to the target join density (it is its equilibrium distribution) but uses only the information about the conditional low-dimensional densities that are far easier to analytically express and sample from.

Let us assume that our goal is to estimate a vector of parameters (and possibly latent state variables) denoted as $\Theta = (\theta_1, \dots, \theta_k)$ and that we are able to analyticaly express (and sample from) all of the univariate conditional densities $p(\theta_j | \theta_i, i \neq j, \text{data})$ but not the multivariate joint density $p(\Theta|\text{data})$. Then we can use the *Gibbs sampler* and construct a Markov Chain whose iterations would converge to the multivariate join posterior density $p(\Theta|\text{data})$ and will therefore generate asymptotically unbiased samples from it.

The procedure for Gibbs Sampler is as follows:

0. Assign a vector of initial values to $\Theta^0 = (\theta_1^0, \dots, \theta_k^0)$ and set $j = 0$

1. Set $j = j + 1$

2. Sample $\theta_1^j \sim p(\theta_1 | \theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})$

3. Sample $\theta_2^j \sim p(\theta_2 | \theta_1^j, \theta_3^{j-1}, \dots, \theta_k^{j-1}, \text{data})$

...

k+1. Sample $\theta_k^j \sim p(\theta_k | \theta_1^j, \theta_2^j, \dots, \theta_{k-1}^j, \text{data})$ and return to step 1.

According to the Clifford-Hammersley theorem (for explanation see Johannes and Polson 2009) the univariate conditional distributions $p(\theta_j | \theta_i, i \neq j, \text{data})$ fully characterize the join distribution $p(\Theta|\text{data})$ and it can also be proved that the Markov Chain constructed according to the Gibbs Sampler converges to the multivariate joint distribution $p(\Theta|\text{data})$. So, we only have to calculate enough iterations of the Gibbs Sampler, discard the first ones (where the chain did not converge yet) and use the remaining ones to estimate all of the required statistical properties of the joint posterior distribution $p(\Theta|\text{data})$.

The necessary conditional densities $p(\theta_j|\theta_i, i \neq j, \text{data})$ are usually obtained by applying the Bayes theorem to the likelihood function and the prior density:

$$p\left(\theta_1\middle|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data}\right) \propto L\left(\text{data}\middle|\theta_1, \theta_2^{j-1}, \dots, \theta_k^{j-1}\right) * \text{prior}\left(\theta_1\middle|\theta_2^{j-1}, \dots, \theta_k^{j-1}\right) \quad (23)$$

where $L(.)$ denotes the likelihood function, prior(.) the Bayesian prior density for the given parameter and $\propto$ represents a proportionate relationship. In our applications we will further use an uninformative prior (i.e. $\text{prior}(\theta_i) \propto 1$) and assume that all of the parameters are independent (which is not true in general).

In order to utilize the Gibbs sampler it is necessary to normalize the right hand side of the equation (23) (i.e. to replace the proportionate relationship with equality). This can be done by integrating the RHS of the equation over θ_1 in order to get the density $p\left(\text{data}\middle|\theta_2^{j-1}, \dots, \theta_k^{j-1}\right)$ and then dividing the RHS by this density.

If the integration on the right hand side of the equation (23) is not possible then the Gibbs Sampler cannot be used. Instead, we can use the so called *Metropolis-Hastings algorithm*, which is a rejection sampling algorithm where a proposal value of each parameter is drawn at first from a proposal density and the new parameter value is then either accepted or rejected based on a given probability. For example, in step 2 we may sample at first a proposal value of the parameter θ_1^j from a propoisal distribution $q(\theta_1|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})$ and then either accept it or reject it (i.e. reset the parameter $\theta_1^j = \theta_1^{j-1}$) based on a given probability.

Specifically, Step 2 in the Gibbs Sampler is replaced by the following two step procedure:

A. Draw θ_1^j from the proposal density $q(\theta_1|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})$

B. Accept θ_1^j with the probability $\alpha = \min(R, 1)$, where R denotes the acceptance ratio defined as:

$$R = \frac{p(\theta_1^j|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data}) q(\theta_1^{j-1}|\theta_1^j, \theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})}{p(\theta_1^{j-1}|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data}) q(\theta_1^j|\theta_1^{j-1}, \theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})} \quad (24)$$

which is in practice evaluated by sampling $u \sim U(0,1)$ from a uniform distribution and accepting θ_1^j if and only if $u < R$.

It can again be shown (see Johaness and Polson 2009) that the resulting Markov-chain converges to the joint posterior distribution $p(\Theta|\text{data})$.

They are many different versions of the Metropolis-Hasting algorithm which often differ in the proposal distribution assumed.

A very popular and easy to use version of the method is the *Random-Walk Metropolis-Hastings* algorithm where the proposal distribution is:

$$\theta_1^j \sim \theta_1^{j-1} + N(0, c) \quad (25)$$

where c is a meta-parameter which may influence the computational efficiency of the algorithm and the practice is to set it so that approximately 50% of the proposals get accepted.

A great advantage of the Random-Walk Metropolis-Hastings algorithm is that its proposal distribution is symmetric which means that the probability of going from θ_1^{j-1} to θ_1^{j} is the same as of going from θ_1^{j} to θ_1^{j-1}. Because of that, the proposal densities in equation (24) cancel out. Consequently, by utilizing the relationship (23) and assuming non-informative priors, the acceptance ratio reduces to the likelihood ratio:

$$R = \frac{L(\text{data}|\theta_1^{j}, \theta_2^{j-1}, \dots, \theta_k^{j-1})}{L(\text{data}|\theta_1^{j-1}, \theta_2^{j-1}, \dots, \theta_k^{j-1})} \tag{26}$$

So, in theory, as long as we are able to analytically express the likelihood function, we can use this algorithm to estimate the joint posterior density of the parameters.

Another popular version of the method is the *Independence Sampling Metropolis-Hastings algorithm* where the proposal density $q(\theta_1|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})$ does not depend on the current value of the relevant parameter θ_1^{j-1} (given the values of the other parameters). In this case the proposal densities in the acceptance ratio in equation (22) do not cancel out. Also to use the algorithm efficiently it is necessary to choose the proposal density $q(\theta_1|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data})$ so that its shape closely resembles the shape of the target density $p\left(\theta_1\middle|\theta_2^{j-1}, \dots, \theta_k^{j-1}, \text{data}\right)$ which is known only up to a normalizing constant.

In our case we want to estimate the vector of a few model parameters Θ and a large number of latent state variables X. Since we know from the Bayes theorem that:

$$p(\Theta, \text{X}|\text{data}) \propto p(\text{data}|\Theta, \text{X}) * p(\text{X}, \Theta) \tag{27}$$

We can estimate iteratively the parameters and the latent state variables:

$$\begin{aligned} p(\Theta|\text{X}, \text{data}) &\propto p(\text{data}|\Theta, \text{X}) * p(\text{X}|\Theta) * p(\Theta) \\ p(\text{X}|\Theta, \text{data}) &\propto p(\text{data}|\Theta, \text{X}) * p(\Theta|\text{X}) * p(\text{X}) \end{aligned} \tag{28}$$

In order to achieve better computational efficiency we will combine in our application different versions of the MCMC algorithm for different variables. Specifically we use Gibbs Sampler to estimate the parameters $\mu, \sigma, \mu_j, \sigma_j, \alpha, \beta, \gamma$ and latent state variables **Q** and **J**. We use the Independence-Sampling Metropolis-Hastings to estimate the latent stochastic variances **V** (the proposal density is based on Jacquier et al. 1994) and we use the Random-Walk Metropolis Hastings to estimate the parameters of the Hawkes process $\theta_J, \beta_J, \gamma_J$.

The exact formulation of the estimation algorithm and the derivation of the conditional distributions is described in detail in Witzany (2013). We add just an additional step to the algorithm to estimate the parameters of the Hawkes process using the Random-Walk Metropolis Hastings algorithm with the likelihood function in the following form:

$$L = \prod_{i=1}^{T} \lambda_i^{J_i} (1 - \lambda_i)^{1-J_i} \tag{29}$$

3 Empirical results

The aim of the research was to compare the results of both of the techniques for stochastic volatility estimation and jump identification - i.e., the nonparametric estimators using high-frequency data and the parametric methods using Bayesian inference on daily data. The research was performed on the time series of EUR/USD exchange rate over the period from 3.2.2009 to 15.4.2014. All of the non-parametric high-frequency estimators were calculated using 15 minute returns because at lower frequencies the autocorrelation of the series was too high which may potentially negatively affect the convergence of the estimators to the underlying quantities. We further removed all the days with less than twenty 15-minute returns from the sample. All of the calculations were performed in Matlab with the exception of the fractional parameter *d* estimates which were calculated in the program Cronos.

In the non-parametric approach, we firstly used the appropriate power-variation estimators in order to estimate the past history of the quadratic variation and its components. Quadratic variation, representing the overall price variability, was estimated using the realized variance. Integrated variance, representing the variability of the diffusion component, was estimated using the bipower variation, and the jump volatility, representing the variability of the jump component, was estimated using the shrinkage estimator with alternative probability levels of 0%, 50%, 90%, 95%, 99%, 99.9% and 99.99%. In some of the analyses we also recalculated the estimates of integrated variance based on the differences between the realized variance and the jump volatility estimated using the shrinkage estimator (i.e., based on equation 15).

Table 1 shows the main summary statistics of realized variance and bipower variation as well as of their logarithmic and root transformations. For the Jarque-Bera test and the Ljung-Box test (with lag 20) we report only the test statistic for comparison. We do not report the p-values, which were extremely low for all of the series (i.e., all of the series exhibit very strong non-normality and autocorrelation on any commonly used probability level).

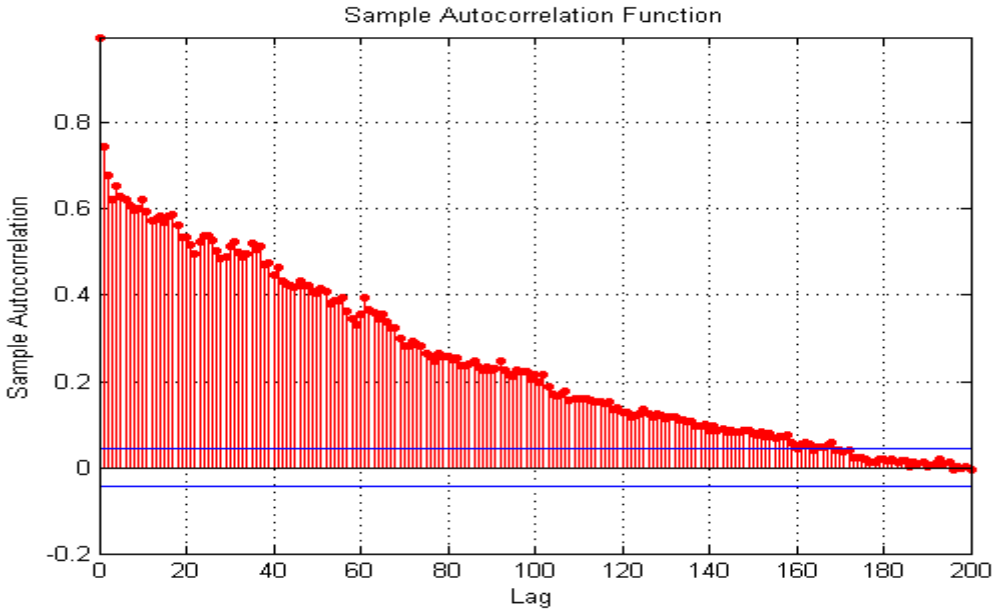
Table 1 - Basic statistics for the realized volatility and bipower variation series

	mean	st.dev	skew	kurt	JB-test	acf_1	LB-test
RV	4.14E-05	4.60E-05	4.58	37.57	111014.48	0.746	15410.05
BV	3.72E-05	4.17E-05	4.23	30.96	74071.31	0.785	17128.07
ln(RV)	-10.4376	0.7922	0.34	3.24	43.98	0.762	19601.80
ln(BV)	-10.5535	0.8002	0.38	3.19	54.00	0.792	20793.67
RV^(1/2)	0.0059	0.0026	1.89	9.06	4426.47	0.786	19555.14
BV^(1/2)	0.0056	0.0025	1.86	8.49	3815.76	0.815	20825.31

As we can see from Table 1, the mean and standard deviation of the bipower variation is approximately 10% lower than in the case of the realized variance. The bipower variation has also slightly lower skewness and kurtosis and is thus a little bit closer to normality. Both of the series are however highly non-normal as can be seen from the extremely high values of their skewness, kurtosis and the JB-test. In comparison, the logarithmic transformations seem to be much closer to the normal distribution. This supports our choice of the log-variance model as the parametric model of volatility. The choice of the AR(1) process for the log-

variance is also somewhat supported by the strong autocorrelation at that lag. From the shape of the autocorrelation function at lags larger than one it is however clear that the AR(1) process may not be an appropriate choice because it fails to capture the long-memory of the volatility process which is present in both of the volatility series (RV and BV) as well as their transformations. Picture 1 shows the autocorrelation function of realized variance.

Picture 1 - The autocorrelation function of the realized variance (RV)



The autocorrelation function of the bipower variation looks very similar to the one of the realized variance. Also considering the logarithmic and root transformations their autocorrelation functions are actually even more persistent than the original series. In order to quantify the persistence of the time series we calculated the values of the fractional parameter *d* by estimating a fractionally integrated process (FI) and an ARFIMA model. In case of the realized variance the value of the parameter *d* is 0,434 for FI and 0,494 for ARFIMA. Considering that 0,5 is the threshold of non-stationarity it is clear that the RV process is very persistent. Actually for the logarithmic transformation of RV the parameter *d* could not even be estimated by our algorithm because the series is non-stationary.

Considering the jump-component - as already noted - we estimated it using the shrinkage estimator with probability levels of 0%, 50%, 90%, 95%, 99%, 99.9% and 99.99%. It is clear that the 0% probability level means that we are accepting all of the identified jumps (i.e. even the negative jumps) while the probability level of 50% corresponds to all of the positive jumps (i.e., we discard only the negative estimated jumps as in equation 9).

We will further proceed as in Andersen et al. (2007) and analyze separately the possible dependencies in the sizes of the jumps and the dependencies in the occurrences of the jumps. To analyze the dependencies in jump sizes we constructed time series only of the realized jumps (i.e., skipping all the days where jumps did not occur).

Table 2 shows some basic statistics about the identified jumps.

Table 2 - Statistics of jumps identified using shrinkage estimator with different alphas

	number	rate	mean	st.dev	skew	kurt	JB-test
JV_0%	2083	100.00%	4.17E-06	9.06E-06	5.697	61.633	309645.39
JV_50%	1620	77.77%	5.96E-06	9.41E-06	6.114	62.765	251194.04
JV_90%	812	38.98%	9.42E-06	1.19E-05	5.121	42.320	55856.44
JV_95%	596	28.61%	1.05E-05	1.23E-05	4.758	37.629	32027.99
JV_99%	330	15.84%	1.27E-05	1.47E-05	4.374	30.062	11122.59
JV_99.9%	132	6.34%	1.51E-05	1.30E-05	2.678	12.737	679.27
JV_99.99%	63	3.02%	1.83E-05	1.47E-05	2.591	11.246	248.97

We can see that the amount of jumps greatly decreases with higher confidence levels used for their estimation. Together with it, the mean size of the jumps and their standard deviation increases with the confidence level, while the skewness and kurtosis are decreasing and the jumps distribution becomes closer to the normal one. Table 3 shows the statistics related to the dependency structure of the jump sizes.

Table 3 - Autocorrelations and the order of integration of the jump sizes

	acf_1	LB-test	LB-pval	d (FI)	d (ARFIMA)
JV_0%	0.088	243.282	0.000	0.119	0.362
JV_50%	0.198	919.942	0.000	0.162	0.422
JV_90%	0.332	985.041	0.000	0.251	0.402
JV_95%	0.264	514.019	0.000	0.226	0.374
JV_99%	0.223	228.486	0.000	0.216	0.376
JV_99.9%	0.039	8.826	0.985	0.038	0.106
JV_99.99%	-0.246	25.331	0.189	-0.136	0.094

As can be seen from Table 3, the sizes of the jumps identified at confidence levels lower or equal to 99% exhibit some autocorrelation on lag 1 as well as jointly on the first 20 lags (as seen from the values of the Ljung-Box test). Then at 99.9% and 99.99% levels the autocorrelation disappears indicating that the jumps probably occur too infrequently at these confidence levels to be somehow serially correlated in terms of their size.

One can observe similar behavior of jump size persistence estimated by the fractional parameter *d*. The most persistent jumps are identified at 50% and 90% probability levels. Nevertheless, their persistence is far lower than in case of the realized variance.

In order to analyze possible dependency in the occurrences of jumps, we first calculated the autocorrelations of the time series of the number of days between when the jumps occur. Surprisingly, the resulting autocorrelations indicated virtually no statistically significant relationship for any of the time series.

As an alternative approach we estimated the parameters of the Hawkes process on the jump time series identified using the shrinkage estimator. The estimation was done by maximizing the likelihood function (according to equation 29) using constrained optimization in Matlab. Table 4 shows the resulting parameters of the Hawkes process:

Table 4 - Parameters of the Hawkes process for the jumps identified using the shrinkage estimator on different probability levels

	ThetaJ	BetaJ	GammaJ
JV_90%	0.3453	0.7380	0.0300
JV_95%	0.2139	0.9650	0.0090
JV_99%	0.1247	0.8841	0.0247
JV_99.9%	0.0561	0.9042	0.0111
JV_99.99%	0.0301	0.8268	0.0009

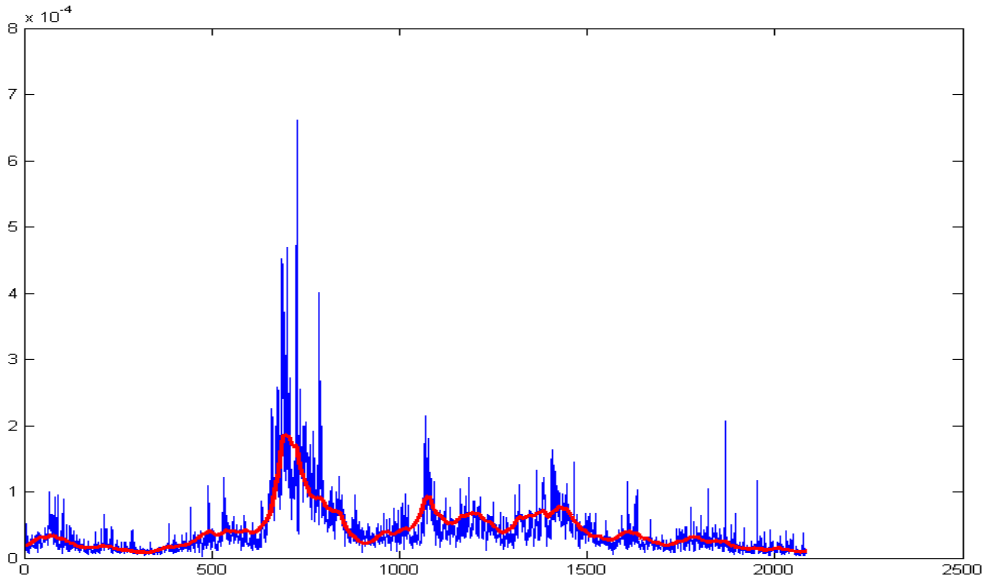
As can be seen from the table, the parameters do indicate some kind of jump clustering. The parameter BetaJ determining the decay rate is relatively close to one for all of the jump series and the parameter GammaJ determining the increase in jump intensity after a jump occurrence is clearly not negligible (with the possible exception of JV_99,99%). Hence - for most of the series - after a jump occurs the probability of subsequent jumps increases and stays increased for several days until the slow exponential decay returns it back to its long-term level ThetaJ.

We further constructed a simple Random-Walk Metropolis-Algorithm (based only on the last step of our full algorithm for the parametric stochastic volatility model) in order to estimate the posterior Bayesian distributions of the parameters of the Hawkes process applied to the jump series identified by the shrinkage estimator with confidence level of 99%. The results in the form of bivariate (i.e., integrated over the third parameter) contour plots of the posterior distributions can be seen in Appendix 1. It is clear that especially for the parameter BetaJ the Bayesian confidence interval would be very wide. There is also apparent some dependence between the parameters (especially involving BetaJ).

The second method of stochastic volatility and jump estimation that we tested is the fully parametric approach. We used a sophisticated MCMC algorithm based on the methodology described in section 2 in order to estimate all of the parameters and latent state variables of the parametric SVJD model described by the equations (20), (21) and (22). As already noted the model has 9 parameters $(\mu, \alpha, \beta, \gamma, \theta_J, \beta_J, \gamma_J, \mu_J, \sigma_J)$ and 3 vectors of latent state variables: **V**, **J**, and **Q**. We estimated them using 5000 MCMC iterations. The convergence of the algorithm was relatively good for the parameters of the stochastic volatility process but not so good for the parameters of the Hawkes process which is however to be expected considering the extremely low number of jumps estimated by the model (as will be seen later).

Picture 2 compares the Bayesian stochastic volatility estimates with the realized variance estimates of the quadratic variation. The Bayesian estimates of the stochastic volatility were constructed using the averages from the last 4500 MCMC iterations. It should be noted that the Bayesian stochastic volatility represents in fact an estimate of the integrated variance and it should be thus compared rather to bipower variation instead of the realized variance. We choose to compare it to the realized variance mainly because of the extreme differences in the properties of the jump component estimated using both of the methods.

Picture 2 - Comparison of the Bayesian stochastic volatility with the realized variance



From Picture 2 we can see that, at least visually, the Bayesian stochastic volatility seems to fit the realized variance quite well. In order to evaluate the fit more quantitatively we calculated the Minzer-Zarnowitz regressions for the stochastic volatility estimates as well as for several other benchmark models. As the dependent variable we used alternatively the realized variance and the bipower variation. Table 5 shows the results of the regression.

Table 5 - In-sample fitting performance of the stochastic volatility model

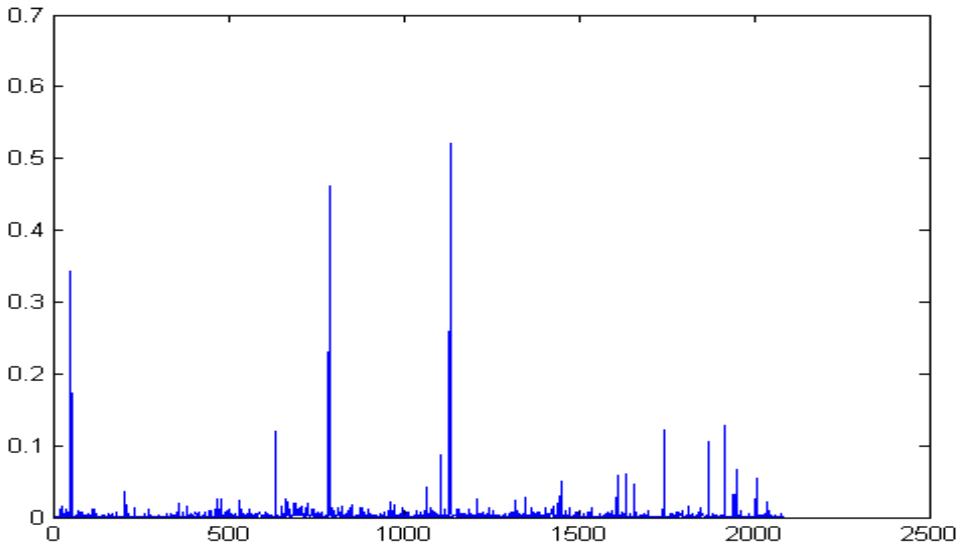
Independent	Explanatory	alpha	beta	r-squared
RV	SV	-5.42E-06	1.1344	0.6031
	EWMA	1.20E-06	0.9651	0.5671
	GARCH	-3.15E-06	1.0640	0.5535
	ARFIMA	-2.30E-07	1.0160	0.6284
	MFV	2.27E-06	0.8167	0.6453
BV	SV	-6.17E-06	1.0514	0.6311
	EWMA	9.31E-08	0.8914	0.5894
	GARCH	-3.91E-06	0.9825	0.5750
	ARFIMA	9.70E-07	0.9833	0.6772
	MFV	9.14E-07	0.7579	0.6770

As can be seen from Table 5, the Bayesian stochastic volatility model provides relatively good fit to the realized variance and the bipower variation. Specifically - it provides the best fit from the group of models utilizing daily data (i.e. SV, EWMA and GARCH). Its fit is however worse than in case of the ARFIMA model which utilizes directly the past history of the realized variance and has also the additional advantage of long-memory. It is also worse than the fit of the Model-Free volatility model (see Britten-Jones and Neuberger 2000) which utilizes option prices. It is also good to mention that all of the models achieved better

performance when predicting the bipower variation then the realized variance which is in accordance with the notion that the continuous component of volatility is more predictable.

Finally, we will compare the Bayesian parametric approach and the power-variation non-parametric approach in their ability to identify jumps. The most obvious difference between the two methods is in the number of jumps they found. As we can see from Table 2 the non-parametric approach identified 1620 alleged jumps at the 50% level and 330 jumps on the 99% probability level. The Bayesian approach on the other hand identified just a single jump with a probability of occurrence larger than 50% and only 7 jumps with probability of occurrence larger than 10%. The Bayesian probabilities of jumps (calculated as the mean of the jump realizations in the last 4500 MCMC iterations) are plotted in the Picture 3.

Picture 3 - Bayesian probabilities of jumps calculated using the MCMC algorithm



Due to the extremely low number of jumps found by the method it is clear that no significant jump-clustering effects could be identified by the Hawkess process. The Bayesian estimates of all of the parameters of the SVJD model are in Table 6.

Table 6 - Bayesian estimates of the parameters of the SVJD model

mui	muiJ	sigmaJ	alpha	beta	gamma	thetaJ	betaJ	gammaJ
0.0002	-0.0055	0.0212	-0.0522	0.9950	0.0748	0.0048	0.3670	0.0470

The value of betaJ is simply too low for any clustering effects to occur. In addition to that the Bayesian confidence intervals for betaJ and gammaJ are enormous (not depicted here).

Nevertheless, even with the extremely low number of jumps identified by the Bayesian method we can still ask if the probabilities of jump occurrences inferred by the two methods are at least correlated. In order to find out, we calculated the Spearman rank correlation coefficient between the Bayesian probabilities of jump occurrences and the variable $Z(t, \Delta)$ from equation (13), which determines the non-parametric probabilities of jump occurrences. The results can be seen in the second cell of the rank correlation matrix in Table 7.

Table 7 - Rank correlation matrix of the jump probabilities and other variables

	B.prob	Z	RV	RV-BV	R2(day)	max(r2(15))
B.prob	**1.0000**	0.0148	0.5184	0.1857	0.7248	0.4571
Z	0.0148	**1.0000**	0.0223	0.8698	0.0365	0.2607
RV	0.5184	0.0223	**1.0000**	0.3574	0.4523	0.8348
RV-BV	0.1857	0.8698	0.3574	**1.0000**	0.1888	0.5234
R2_day	0.7248	0.0365	0.4523	0.1888	**1.0000**	0.4273
max(r2_15)	0.4571	0.2607	0.8348	0.5234	0.4273	**1.0000**

As can be seen from the matrix, the rank correlation coefficient between the Bayesian jump probabilities and the variable $Z(t,\Delta)$ is close to zero (0.0148) which is very surprising. In order to find a possible explanation for this phenomenon we calculated rank correlations with some of the other variables as well. As we can see, the Bayesian jump probabilities are most strongly correlated with the squared daily returns R2(day) which is to be expected because the probabilities are inferred from these returns. There is further a relatively strong correlation with realized variance and with the maximum 15 minute squared return during the given day denoted as max(r2(15)). Considering the variable $Z(t,\Delta)$, we can see that it is most strongly correlated with the variable RV-BV (which is equal to JV_0%) and that it is also somewhat correlated with the maximum 15 minute squared return, but surprisingly - almost uncorrelated with the squared daily returns as well as with the realized variance.

So it seems, that the jumps identified using the shrinkage estimator do depend almost exclusively on the behavior of the price at high-frequencies and are virtually indistinguishable from the continuous volatility component on the daily frequency. This is in a clear contradiction with the applied Bayesian approach for jump identification which assumes that the jumps are large enough in order to have some effect on the daily frequency.

In order to evaluate how likely the jumps are to mix with the daily continuous volatility, we calculated the ratios of the moments of the jump component and the continuous volatility component both estimated using the shrinkage estimator with different probability levels and equation (15). The results are in Table 8.

Table 8 - Ratios of the moments of the daily jump and continuous volatility

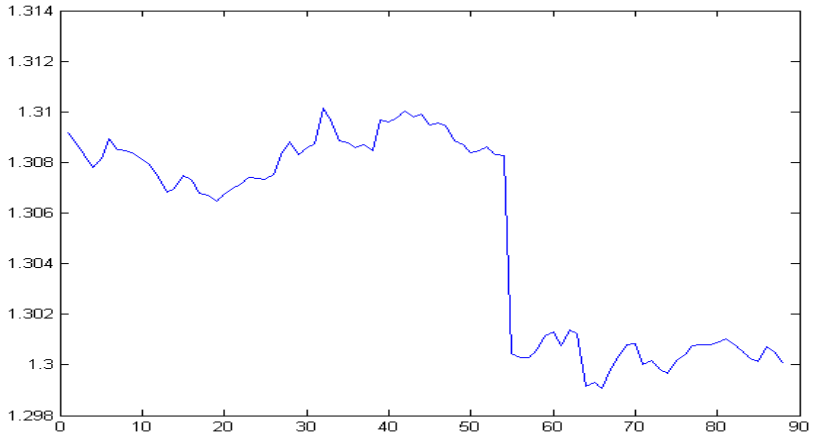
	mean	st.dev	skew	kurt
JV_0%	11.21%	21.74%	134.71%	199.06%
JV_50%	16.22%	22.84%	142.53%	197.27%
JV_90%	24.97%	28.20%	120.33%	135.52%
JV_95%	27.25%	28.17%	100.88%	92.42%
JV_99%	32.17%	33.10%	95.46%	78.06%
JV_99.9%	37.30%	28.39%	57.63%	33.25%
JV_99.99%	44.76%	31.89%	56.16%	29.74%

As can be seen from the table, the mean size of the jump component as well as its standard deviation are even for the most significant jumps much smaller than the relevant moments of

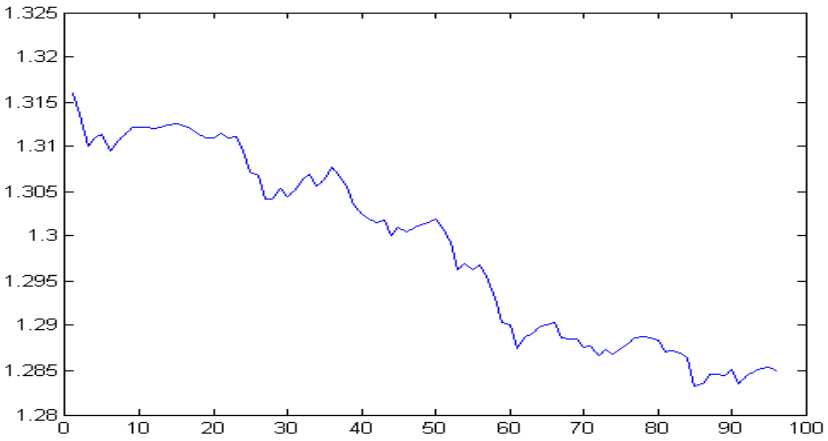
the continuous volatility component. It is then easy for them to get lost in the overall daily variability and become completely indistinguishable from the daily returns.

To illustrate the point we are showing here in Picture 4 the day with the highest value of the $Z(t,\Delta)$ statistics (9.48) and in Picture 5 the day with the highest Bayesian probability of jump occurrence (54.74%). As we can see, the return on the day with the $Z(t,\Delta)$ of 9.48 was not enormously large, but instead contained a single abrupt extremely large 15 minute return. This could of course not be seen on the daily frequency and the Bayesian method actually assigned a mere 0.28% jump probability to that day. On the other hand if we look at the returns on Picture 5 to which the Bayesian method assigned 54.74% probability of a jump occurrence, we can see that there were no visible discontinuities in the price evolution on that day and that the price merely trended the whole day achieving an enormous return during the whole period. Indeed - the value of the $Z(t,\Delta)$ statistic on that day was only 0.4455 meaning that no jump was estimated by any of the JV estimators with the exception of the JV_0%.

Picture 4 - The day with the largest value of the Z statistics (Z = 9.48) (5.1.2007)



Picture 5 - The day with the largest Bayesian probability of jump occurrence (p=54.74%) (11.8.2010)



Conclusion

We compared two different approaches to stochastic volatility and jump estimation. The first approach uses power-variation measures calculated from high-frequency data in order to non-parametrically estimate the continuous and discontinuous components of price variability. In this respect we used the realized variance to estimate the quadratic variation, the bipower variation to estimate the integrated variance and the shrinkage estimator to estimate the jump volatility on several different probability levels. The second approach, on the other hand, specifies the underlying volatility and jump processes parametrically and then estimates their parameters as well as the latent state variables (i.e. the unobservable processes) using Bayesian inference methods. We used a log-variance Stochastic-Volatility Jump-Diffusion model with self-exciting jumps governed by the Hawkes process as our underlying parametric model and we estimated its parameters using an MCMC algorithm.

The comparison of the methods was performed on the past history of the EUR/USD exchange rate series during the period between 3.2.2009 and 15.4.2014.

Regarding the continuous stochastic volatility, both of the methods provided very similar results. Actually the log-variance stochastic volatility model achieved the best fit to the realized variance from all of the benchmark models using daily data (log-SV, EWMA and GARCH), the fit was however worse than in the case of an ARFIMA model (using directly the realized variance) and the Model-Free volatility (using option prizes).

In the case of the jump component the results of the two methods were markedly different. Firstly - the parametric Bayesian approach using daily returns identified significantly less jumps when compared to the non-parametric approach using 15 minute returns. More importantly - the inferred probabilities of jump occurrences by the two methods were virtually no rank-correlated, meaning that the two methods do not identify jumps on the same days. Surprisingly - the non-parametrically inferred jump probabilities were not correlated even with the daily realized variance and the daily returns, which implies that the high-frequency discontinuous price changes (jumps) may not be high enough in order to be distinguishable on the daily frequency. This sheds doubts on empirical validity of the JDSV model. One possible direction which could be investigated in further research is that the identified daily jumps are in fact caused rather by jumps in volatility than by jumps in returns. In that case one would have to estimate and compare jumps in realized volatility with jumps in stochastic volatility.

As an additional result we found that the jumps identified using the shrinkage estimator exhibit significant dependency with regards to their size as well as regarding the intensity of their occurrence. The size dependency (autocorrelation with possibly long memory) is the strongest for jumps identified using low probability levels in the shrinkage estimator, it decreases with the higher probability levels and completely vanishes at probability levels of 99.9% and 99.99%. The dependency in the jump occurrences was identified using the self-exciting Hawkes process and seems to be present on all of the probability levels for jump identification with the possible exception of the last one (99,9%). The dependence indicates that in the days following a jump occurrence the probability of subsequent jumps is increased. Again, the jump clustering effect could be also explained by occurrence of jumps in volatility which we intend to investigate in further research.

Literature

[1] **ANDERSEN, T.G. and BOLLERSLEV, T., (1998).** "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, November 1998,Vol. 39, No.4, p.885-905.

[2] **ANDERSEN, T.G., BOLLERSLEV, T., CHRISTOFFERSEN, P.F. and DIEBOLD, F.X., (2005).** "Volatility Forecasting", *National Bureau of Economic Research*, March 2005, NBER Working Paper No. 11188.

[3] **ANDERSEN, T.G., BOLLERSLEV, T. and DIEBOLD, F.X., (2007).** "Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility", *Review of Economics and Statistics*, November 2007, Vol. 89, No. 4, p.701-720.

[4] **ANDERSEN, T.G., DOBREV, D. and SCHAUMBURG, E., (2010).** "Jump-Robust Volatility Estimation using Nearest Neighbor Truncation", *Federal Reserve Bank of New Yor*k, August 2010, Staff reports, No.465, (p.37)

[5] **BAKSHI, G. and KAPADIA, N., (2003).** "Delta-hedged gains and the negative market volatility risk premium", *Review of Financial Studies*, February 2003, Vol.16, No.2, p.527-566

[6] **BARNDORFF-NIELSEN, Ole E., HANSEN, P.R., LUNDE, A., SHEPARD, N., (2008).** "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise", *Econometrica*, November 2008, Vol.76, p.1481–1536.

[7] **BARNDORFF-NIELSEN, Ole E. and SHEPHARD Neil, (2004).** "Power and Bipower Variation with Stochastic Volatility and Jumps", *Journal of Financial Econometrics*, 2004, Vol.2, No.1, p.1-48

[8] **BRITTEN-JONES, Mark and NEUBERGER, Anthony, (2000).** "Option Prices, Implied Price Processes and Stochastic Volatility", *Journal of Finance*, April 2000, Vol. 55, No. 2, p.839-866

[9] **CARR, Peter and WU, Liuren, (2009).** "Variance Risk Premiums", *The Review of Financial Studies*, March 2009, Vol.22, No.3 , p.1311-1341

[10] **CHEN, Ke and POON, Ser-Huang, (2013).** "Variance Swap Premium under Stochastic Volatility and Self-Exciting Jumps", *Manchester Business School, University of Manchester*, January 2013, Working paper, p.1-50,

[11] **CORSI, F., PIRINO, D. and RENO, R., (2010).** "Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting", *Sant´Anna School of Advanced Studies*, LEM Working Paper Series, July 2010, No.11, p.1-33.

[12] **CRAINE, R., LOCHSTOER, L.A. and Syrtveit K., (2000).** "Estimation of a Stochastic-Volatility Jump-Diffusion Model", Economic Analysis Review, 2000, Vol.15, No.1, p.61-87, ISSN 0716-5927

[13] **ERAKER, Bjorn, (2004).** "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices." *The Journal of Finance*, June 2004, Vol. 59, No. 3, p. 1367-1403.

[14] **ERAKER, Bjorn, (2009).** "The Volatility Premium", *Wisconsin School of Business, University of Wisconsin*, 2009, NCCR Working paper

[15] **ERAKER, B., JOHANNES, M. and POLSON, N. G., (2003).** "The Impact of Jumping equity index Volatility and Returns." *The Journal of Finance*, June 2003, Vol.58, No.3, pp. 1269-1300.

[16] **FULOP, Andras., LI, Junye and YU, Jun, (2014).** "Self-Exciting Jumps, Learning, and Asset Pricing Implications", *Singapore Management University*, Working paper, June, 2014, No.02-2014, p.1-54,

[17] **GONZÁLES A., NOVALES, A. and RUBIO, G., (2011).** "Estimation of Stochastic Volatility Models with Jumps in Returns for Stock Market Indices", August 2011, Working Paper, p.1-28.

[18] **HESTON, S., (1993).** "Closed-form Solution of Options with Stochastic Volatility with Application to Bond and Currency Options." *Review of Financial Studies*, 1993, Vol.6, pp. 327-343.

[19] **ISHIDA, Isau and WATANABE, Toshiaki, (2009).** "Modeling and Forecasting the Volatility of the Nikkei 225 Realized Volatility Using the ARFIMA-GARCH Model", *University of Tokyo , Hitotsubashi University*, October 2008 (January 2009), CIRJE-F-608 Discussion Paper, p.1-32.

[20] **JACQUIER, E., JOHANNES, M. and POLSON, N., (2007).** "MCMC Maximum Likelihood for Latent State Models.", *Journal of Econometrics*, 2007, Vol.137, p. 615–640.

[21] **JACQUIER, E., POLSON, N. and ROSSI, P. (1994).** "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business & Economic Statistics*, October 1994, Vol.12, No.4, p.69-87.

[22] **JOHANNES, M. and POLSON, N. (2009).** "MCMC Methods for Financial Econometrics.", In Ait-Sahalia, Hansen, L. P., eds., *Handbook of Financial Econometrics*, p.1-72.

[23] **LANNE, Markku, (2006).** "Forecasting Realized Volatility by Decomposition", *European University Institute*, 2006, Economics Working Papers, ECO2006/20, p.1-26.

[24] **SHEPHARD, Neil**, **(2005).** "*Stochastic Volatility: Selected Readings (Advanced Texts in Econometrics)*", Oxford: Oxford University Press, May 2006, Edited by Neil Shephard

[25] **SHI, Peng, (2009).** "Correcting Finite Sample Biases in Conventional Estimates of Power Variation and Jumps", *Duke University, Durham NC, 27708*, April 2009, Final Paper for Econ201FS, p.1-22

[26] **STÁDNÍK Bohumil, (2014**). "The Riddle of Volatility Clusters", *Verslas: Teorija ir Praktika / Business: Theory and Practice*, 2014, Vol.15, No.2, p.140-148, ISSN 1822-4202.

[27] **TODOROV, Viktor, (2009).** "Variance Risk Premium Dynamics: The Role of Jumps", *The Review of Financial Studies*, January 2010, Vol.23, No.1, p.345-383

[28] **WITZANY, Jiří, (2013).** "Estimating Correlated Jumps and Stochastic Volatilities", *Prague Economic Papers*, 2013, Vol.2013, No.2, p.251-283

[29] **YSUSI, Carla, (2006).** "Detecting Jumps in High-Frequency Financial Series Using Multipower Variation", *Banco de Mexico*, September 2006, Working paper, No.2006-10, p.1-35.

[30] **ZHANG, L., MYKLAND, Per A. and AIT-SAHALIA, Y., (2005).** "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data", Journal of the American Statistical Association, December

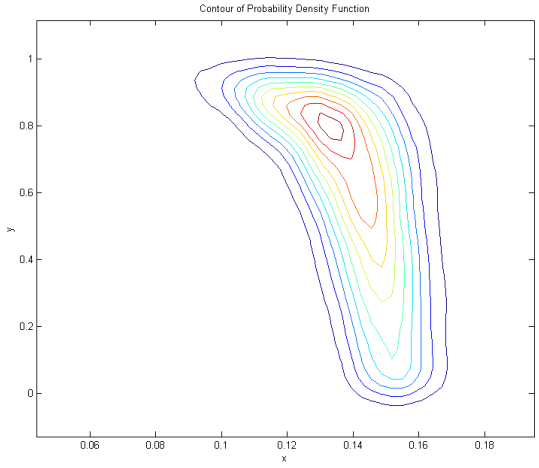
Abstrakt: Porovnáváme dva přístupy pro odhad stochastické volatility a skoků měnového kurzu EUR/USD, ne-parametrický přístup založený na power-variation estimátorech a vysokofrekvenčních datech, a parametrický Bayesovský přístup (MCMC odhad SVJD modelů) vycházející z denních dat. Zjišťujeme, že obě metody podávají podobné odhady spojité stochastické volatility avšak velmi odlišné odhady skokové složky. Zaprvé - skoky odhadnuté ne-parametricky z vysokofrekvenčních dat jsou daleko početnější než skoky odhadnuté Bayesovsky z denních dat. Především však zjišťujeme, že pravděpodobnosti nastání skoků pro jednotlivé dny odhadnuté pomocí obou zmíněných metod nevykazují téměř žádnou pořádkovou korelaci (Spearmanův korelační koeficient je 0,0148), což znamená, že obě metody neidentifikují skoky ve stejných dnech. Krom toho bylo zjištěno, že pravděpodobnosti skoků odvozené ne-parametricky z vysokofrekvenčních dat nejsou příliš korelovány ani s denní realizovanou volatilitou a denními čtvercovými výnosy, což znamená, že nespojité změny ceny pozorovatelné na vysokých frekvencích zřejmě nelze na denní frekvenci rozlišit od spojité volatility. Dále zjišťujeme, že skoky identifikované ne-parametricky (pomocí shrinkage estimátoru) vykazují silnou závislost co do absolutní velikosti, jakož i určité shlukování (modelováno pomocí Hawkesova procesu).

AMS/JEL klasifikace: C11, C14, C15, C22, G1

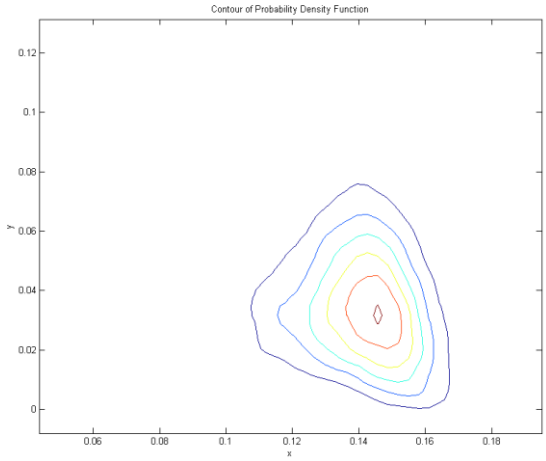
Klíčová slova: Stochastická volatilita, Bayesovský odhad, MCMC, Realizovaný rozptyl, Bipower variation, Shrinkage estimátor, Shlukování skoků, Self-Exciting skoky, Hawkesův proces

Appendix 1

Picture 6 - Posterior density of the parameters ThetaJ (x) and BetaJ (y)



Picture 7 - Posterior density of the parameters ThetaJ (x) and GammaJ (y)



Picture 8 - Posterior density of the parameters BetaJ (x) and GammaJ (y)