



UNIVERSITY OF WARSAW

Faculty of Economic Sciences

WORKING PAPERS

No. 13/2009 (23)

ROBERT ŚLEPACZUK
GRZEGORZ ZAKRZEWSKI

HIGH-FREQUENCY AND MODEL-FREE
VOLATILITY ESTIMATORS

WARSAW 2009



UNIVERSITY OF WARSAW
Faculty of Economic Sciences

Working Paper No. 13/2009 (23)

High-Frequency and Model-Free Volatility Estimators

(This work was supported by the Foundation for Polish Science)

Robert Ślepaczuk

Faculty of Economic Sciences
University of Warsaw

Grzegorz Zakrzewski

Credit Risk Management Expert
Deutsche Bank PBC S.A.

Abstract

This paper focuses on volatility of financial markets, which is one of the most important issues in finance, especially with regard to modeling high-frequency data. Risk management, asset pricing and option valuation techniques are the areas where the concept of volatility estimators (consistent, unbiased and the most efficient) is of crucial concern. Our intention was to find the best estimator of true volatility taking into account the latest investigations in finance literature. Basing on the methodology presented in Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991), Yang and Zhang (2000), Andersen et al. (1997, 1998, 1999a, 199b), Hansen and Lunde (2005, 2006b) and Martens (2007), we computed the various model-free volatility estimators and compared them with classical volatility estimator, most often used in financial models. In order to reveal the information set hidden in high-frequency data, we utilized the concept of realized volatility and realized range. Calculating our estimator, we carefully focused on Δ (the interval used in calculation), n (the memory of the process) and q (scaling factor for scaled estimators). Our results revealed that the appropriate selection of Δ and n plays a crucial role when we try to answer the question concerning the estimator efficiency, as well as its accuracy. Having nine estimators of volatility, we found that for optimal n (measured in days) and Δ (in minutes) we obtain the most efficient estimator. Our findings confirmed that the best estimator should include information contained not only in closing prices but in the price range as well (range estimators). What is more important, we focused on the properties of the formula itself, independently of the interval used, comparing the estimator with the same Δ , n and q parameter. We observed that the formula of volatility estimator is not as important as the process of selection of the optimal parameter n or Δ . Finally, we focused on the asymmetry between market turmoil and adjustments of volatility. Next, we put stress on the implications of our results for well-known financial models which utilize classical volatility estimator as the main input variable.

Keywords:

financial market volatility, high-frequency financial data, realized volatility and correlation, volatility forecasting, microstructure bias, the opening jump effect, the bid-ask bounce, autocovariance bias, daily patterns of volatility, emerging markets

JEL:

G14, G15, C61, C22

The content of this material reflects personal opinions and views of the authors.
Please contact the authors to give comments or to obtain revised version.
(contact information: rslepaczuk@wne.uw.edu.pl)

1. Introduction

Almost everyone who works within financial markets, no matter if we mean their practical or theoretical aspects, is concerned about the volatility issue. The concept of volatility, especially predicting its future levels and managing the risk coming from its fluctuations, is of crucial importance for a number of reasons.¹ The knowledge of true volatility is necessary when we compute all asset pricing models (CAPM, APT, multi-factor asset pricing models) and when we try to optimize mean-variance interrelation. It also has to be estimated in all kinds of VaR models including stress-testing and worst case scenarios, which try to predict the portfolio loss taking into account given significance level and the distribution of the returns. When we consider the option pricing techniques, volatility is the main input variable, the levels of which affect the final theoretical value the most, no matter which option valuation model we choose. Of course we cannot forget about implied volatility derived from option pricing models, which in practice is traded on the market instead of the value of the underlying contract; and the management of Vega, which is the most sophisticated issue when we consider the risk of actively managed option portfolio. Finally, we have to emphasize the management of portfolios of derivatives (futures, options or swaps), where what matters is not only the direction of the market but also, and most of all, an accurate prediction of future volatility (especially in the case of volatile market with high unexpected jumps between close and open), which is even more important for proper portfolio risk management. We should add here that volatility risk, highly correlated with liquidity risk, especially in the time of market turmoils, is nowadays the main source of risk in financial system, where the hedge funds managing billions of dollars use the financial leverage in almost each trade. This subject is particularly important in the globalized financial markets, where significant turmoils happen more often (May-June 2006, March, August, November 2007 or January 2008) and, what is more important, spread all over the world with ultra-fast speed. Therefore, the observation of modern capital markets prompts us to design the research that pays crucial attention to the ways of risk quantifying.

Contemporary state-of-the-art is that the classical volatility estimator (standard deviation of daily log returns) or implied volatility (used in option modelling) are the measures of volatility most often used in models, whenever we need the true measure of volatility. However, as the latest research revealed (Andersen et al. (2001), Yang and Zhang (2000), Shu and Zhang (2006), Martens and Dijk (2007)) we can find other, more efficient and yet unbiased estimators of true volatility. Among other things, we want to verify this notion in our paper.

Taking into consideration all the above-mentioned issues and trying to find better estimators of true volatility by studying the properties and merits of realized range² and realized volatility³, we calculated the following measures of volatility:

1. classical volatility estimator ($^{annualstd}SD_{\Delta}^n$)
2. realized volatility ($^{annualstd}RV_{\Delta}^n$)
3. realized range ($^{annualstd}RR_{\Delta}^n$)

¹ Investors who did not consider these issues as key factors were taught a very expensive and painful lesson. One of numerous examples is the collapse of prestigious Long Term Capital Management in 1998, which almost destabilized global financial system (Lowenstein, 2000).

² The measure of daily volatility computed by summing high-low ranges for intra-day intervals.

³ The measure of daily volatility computed by summing squared returns for intra-day intervals.

4. Garman-Klass volatility estimator ($^{annual}std GK_{\Delta}^n$)
5. Rogers-Satchell volatility estimator ($^{annual}std RS_{\Delta}^n$)
6. Yang-Zhang volatility estimator ($^{annual}std YZ_{\Delta}^n$)
7. realized volatility with autocovariance correction ($^{annual}std_{AC1} RV_{\Delta}^n$)
8. scaled realized volatility ($^{annual}std_q RV_{\Delta}^n$)
9. scaled realized range ($^{annual}std_q RR_{\Delta}^n$)

After obtaining given estimators of volatility, we compared their distribution with the classical volatility estimator to find out all the details concerning the information content of each of them and specify the main differences between them. Next, we focused on several aspects concerning the way of its measuring, i.e.: number of days for scaling factor ($q=?$), the optimal sampling frequency ($\Delta=?$), and the length of the memory ($n=?$), so as to ensure the extraction of appropriate information from today's and previous quotations. The detailed discussion of parameter n constitutes the important contribution of our paper to volatility literature. Much as the issue is often omitted, we show it to play a crucial role in the performance of volatility estimators, stability of fluctuations of annualized and averaged estimators and the set of information incorporated in their calculation formula.

We noticed the necessity of using high, low, open and close prices instead of only close prices in the process of estimation. It is clearly observed that a substantial part of daily volatility can be revealed only when we base our calculations on the intraday range, revealing its intra-interval fluctuations, as well (Slepaczuk and Zakrzewski, 2008). Not including this information in the close-to-close estimators was the reason, while we added the various concepts of realized range estimators to our comparison (RR, GK, RS, YZ, etc.).

One of our main goals was to find the most efficient estimator, which would take us closer to the *true volatility* estimator. In order to accomplish this goal, we calculated several statistics describing relative performance of our volatility estimators (the variance efficiency ratio, the modified variance efficiency ratio, and the relative variance efficiency ratio). We also discussed the issue of accuracy, however, we leave the detailed investigation of that subject for the future research.

The consecutive idea of this paper was to show that computing volatility on high-frequency data could provide us with valuable information about fluctuations of the market which is not revealed in the data on the daily basis, especially when the market is volatile during the day. That was the reason for calculating our estimators on various sampling frequencies (from 5-minute to daily interval).

Finally, we have to stress that the main objective of this work, which was to find the best estimator of true volatility, was only the step into further investigations, which concern forecasting future volatility levels. What we have to do first in the process of selecting the best econometric models for forecasting is to choose the most appropriate input data (one of the proposed volatility estimators) which significantly influence accurate volatility forecasts.

Taking into account all the above mentioned, we formulated several hypotheses, which will be verified further in this paper:

1. Computing volatility on high-frequency data will provide us with valuable information about fluctuations of the market, which is not revealed in the data on the daily basis. It is

indirectly connected with greater efficiency of estimators calculated on the basis of HF data.

2. During the estimation process we should utilize information included in price range (high, low, open and close prices), in order to obtain the most efficient estimators.
3. The efficiency of volatility estimator is closely connected with the length of the memory (parameter n) used in the process of calculation.
4. The efficiency of the estimator concerning its formula is only revealed when compared to the other estimator calculated on the basis of the same interval δ and parameter n .
5. Various concepts of volatility estimators could over or under-estimate the actual level of volatility depending on their calculation formula.

To accomplish described goals we planned the structure of our paper as follows. Next section describes the recent literature and some stylized facts about volatility. Theoretical background, formulas and the overall context of presented research are in the third part. Fourth section contains the description of data used in the process of analysis. The process of selection of q , δ and n is described in the fifth section. The consecutive part contains the comparison of distributions of volatility estimators. The seventh part focuses on the connection between market turmoil and the behaviour of different volatility estimators. In this section we also describe implications of the results for financial models. The last section concludes and defines paths for further researches.

2. Literature review

The concept of volatility estimators is widely researched in financial literature. Scientists try to find the best estimator of true volatility, which is not observed/rather latent process, through numerous researches on daily, weekly or high frequency financial data.

Contemporarily, the most frequently used estimator is still classical volatility estimator (the sum of squared differences of i_{th} return and the mean return over the analyzed period of time) which is the part of many kinds of financial models (Black-Scholes model, CAPM, APT, input variable in various GARCH and ARCH models, as well as stochastic volatility models, etc.) and which is frequently treated as sufficient estimator of *true volatility* process. Although this estimation is to a large extent successful, we are aware of the fact that it is possible to find better, more efficient, still unbiased and consistent estimators. The most important disadvantage of SD is that it is calculated on the daily basis, not revealing intraday fluctuations and that it is supposed to have low efficiency in comparison with other volatility estimators (e.g. Martens and Dijk, 2007, Yang and Zhang, 2000, Shu and Zhang, 2006).

Since the concept of volatility has grown in importance through the last forty years, many new concepts of volatility estimators focused on gaining on efficiency and being robust to all existing microstructure biases (bid-ask spread, the opening jump effect, non-trading bounce, etc.) have been invented. Therefore we have thoroughly and chronologically studied the most influential works concerning the issue of volatility estimators and their properties, in order to place our research as the natural consequence of the contemporary state-of-the-art and focus on the most important details which were not sufficiently explained in the previous works.

Merton (1980), who was the first to propose realized volatility concept (the sum of squared returns over the analyzed period of time measured in equidistant periods) as the unbiased and consistent estimator of daily variance σ_t^2 on condition that the returns have a zero mean and are uncorrelated. He agreed that RV is the true volatility estimator when

returns are sampled as often as possible. This concept was later heavily researched by Taylor and Xu (1997) and Andersen et al. (1998, 2000, 2001a and 2001b) as well as others, who additionally paid close attention to microstructure bias which unfortunately grows in importance as the sampling frequency increases.

In 1980, Parkinson introduced the new concept, which utilized price range (high price - low price, or consequently their logs) in order to reveal all the information set included in the price fluctuations. This improvement led to range-based volatility estimator which was still unbiased if the mean was equal zero and almost 5 times more efficient than the classical volatility estimator.

Next concept of Garman and Klass (1980) improved Parkinson's estimation by including not only high and low prices but open and close prices, as well. They defined the minimum-variance unbiased estimator for Brownian motion with zero-drift. Moreover, they proved that their range based volatility estimator is eight times more efficient than the classical volatility estimator.

Following the discoveries described above, Rogers and Satchell (1991) proposed a new approach to volatility estimation. Suggested concept was unbiased and independent of the drift term. Their estimator improved the main drawback of Garman-Klass estimator which was biased if used in the case of non-zero drift. However, Rogers-Satchell estimator still assumed no opening jump effect and was unbiased only under this assumption.

The next step into increasing volatility literature, where successive researchers attempted to find the unbiased estimators, which would be independent of both opening jump and the drift term, was Yang and Zhang (2000) with their multiperiod volatility estimator based on high, low, open and close prices. The Yang-Zhang estimator had the following properties: (a) unbiased, (b) independent of the drift, (c) consistent in dealing with the opening jump and (d) smallest variance among all the estimators with similar properties (the typical biweekly Yang-Zhang variance estimation was over 7 times more efficient than the classical variance estimator).

Andersen and Bollerslev started to popularize the notion of realized volatility and correlation in 90s having written the numerous research papers (Andersen and Bollerslev, 1998, 1999a, 1999b) devoted to the techniques focusing on many possible aspects and dimensions of that issue, especially the properties of such estimator calculated on the high frequency data. They noticed that the realized volatility is a more efficient and unbiased estimator of volatility than the popular daily classical volatility estimator. Moreover, it converges to the true underlying integrated variance when the length of the intraday interval goes to zero (Andersen et al. 2001a, 2001b). They found that the efficiency of the daily high-low range is between that of the realized variance computed using 3- and 6- hour returns. Estimating realized volatility of stock returns they noticed that the sampling frequency of 5- and 30-minute intervals strike a balance between the increasing accuracy of higher frequencies and the adverse effects of market microstructure frictions (Andersen et al., 2001a, 2003).

The discussion of the volatility estimation techniques was enriched by the description of distribution of volatility during the normal stock session which we can name daily patterns of volatility (Taylor and Areal, 2000). Basing on five-minute returns they presented the distribution of the volatility of FTSE-100 index focusing on significant jumps of volatility during the day, caused by the announcement of US or UK macro data. They also revealed that the patterns of volatility are considerably diverse through consecutive days of the week. Studying the distribution of the logarithm of volatility and that of returns standardized by realized volatility they confirmed that it is almost exactly normal.

Zhang et al. (2005) went one step further and developed the estimator which combined realized variance estimator obtained from returns sampled at two different frequencies. The realized variance estimator obtained using a certain (low) frequency was corrected for bias due to microstructure noise using the realized variance obtained with the highest available sampling frequency.

Ait-Sahalia et al. (2005) and Hansen and Lunde (2006b) revealed that returns at very high frequencies are distorted such that the realized variance becomes biased and inconsistent. Deriving the theoretical properties for realized range in a world with no market microstructure noise and with continuous trading, Christensen and Podolskij (2005) stated that this estimator is five times more efficient than the realized variance sampled with the same frequency and converges to the integrated variance with the same rate.

When testing the relative performance of various historical volatility estimators that incorporate daily trading range Shu and Zhang (2006) found that the range estimators perform very well when asset price follows a continuous geometric Brownian motion. However, significant differences among various range estimators are detected if the asset return distribution involves an opening jump or large drift. Nonetheless, the empirical result is supportive of the use of range estimators in estimating historical volatility.

Martens and Dijk (2007) tried to develop the concept of realized range by introducing scaled realized range which was additionally robust to microstructure noise. They noticed that realized range with their bias-adjustment procedure was more efficient than realized variance when using the same sampling frequency.

Discussing the issue of volatility estimators we cannot forget about implied volatility derived from the market prices of options with help of adequate theoretical model. The concept, which has been developing successfully from the mid 70s, got the inspiring injection of new theoretical thought after publication of Derman et al. (1999), who explain the properties and the theory of both variance and volatility swaps, deriving volatility directly from option prices basing on model-free and non-parametric approach to volatility estimation. They also design the framework for hedging Vega, showing how a variance swap can be theoretically replicated by a hedged portfolio of the strip of out of the money options (Call and Put) with adequate weights. Assuming that the fair value of the variance swap is the cost of the replicating portfolio, they derive analytic formulas for theoretical fair value in the presence of realistic volatility skew. Nowadays, the above concept lies behind the theoretical formulas designed for VIX and many other volatility indexes (VXO, VXN, VDAX, VDAX-NEW, VSMI, and recently computed VIW20⁴). Moreover, VIX index is even the basis instrument, for derivatives (futures and options) quoted on the CBOE and this issue will be important in the discussion presented in the sections seven and eight.

Before we go to the main part of this paper let us look at some stylized facts established between theoreticians and practitioners dealing with the issue of volatility:

- Volatility time series are *mean-reverting*; Moreover, analyzing the behaviour of VIX we can even say that it is “*minimum reverting process*” (Slepaczuk and Zakrzewski, 2007).
- *Long memory phenomenon* or the *persistence effect* in the volatility time-series, i.e. after negative or positive shock in volatility, the shock dies out very slowly (Baillie et al., 1999 – *fractionally integrated time series*).

⁴ VIW20 is the volatility index for WIG20 index, the main equity index for the Polish stock market (Slepaczuk and Zakrzewski, 2007).

- *Volatility clustering*, i.e. we observe distinct periods when volatility clusters on the high or low level for the long period of time (Andersen et al., 2001a). This effect is closely related to long memory effect, described above.
- *The leverage effect* revealing asymmetric volatility reaction on the shocks in the basis stock market index, i.e. sudden jumps of volatility connected with the sharp downward movement and moderate increase or even decrease in the time of upward movement (Black, 1976; Ebens, 1999; Andersen et al., 2001a). Additionally, Andersen et al. (2001a) showed that this effect is stronger on the aggregated level (market stock indexes) than for individual stocks.
- *Strong negative correlation with the basis index*, which additionally strengthens in the time of market turmoils, contrary to the correlation between normal instruments (stocks, bonds, etc) where initially defined negative correlation disappears when the market is on the edge of crash (Slepaczuk and Zakrzewski, 2007).
- *Volatility is time varying and predictable* to a certain extent (Giot and Laurent (2004) and Martens and Dijk (2007),
- The distribution of variance is characterized by high kurtosis, positive skewness and non-normality, but the logarithm of volatility (realized volatility) is approximately normal (Giot and Laurent (2004), Andersen et al. (2001a and 2001b).
- *Volatility-in-correlation effect*, the strong positive relations between individual stock volatilities and the corresponding strong positive relations between contemporaneous stock correlations (Andersen et al., 2001a). There is a systematic tendency for the variances to move together and for the correlations among the different stocks to be high/low when the variances for the underlying stocks are high/low, and when the correlations among the other stocks are also high/low.
- *Upward and downward sloping term structure of volatility*, especially when we consider implied volatility, which can be easily explained by the mean reverting process. When the short-term implied volatility is relatively high/low then the term structure is downward/upward sloping.

While conducting our research we will compare our results with stylized facts mentioned above.

3. Theoretical background, and the notion behind each formula

On the ground of presented literature we assumed that volatility estimators calculated on the basis of high-frequency data including information about intraday range should be the most efficient ones. Choosing the set of estimator for our research we took this notion into account.

The definitions of our estimators are presented below. First, we present the formula to calculate classical volatility estimator (standard deviation of log returns) which gives us, the average variance over the period of n-days:

$$VAR_{\Delta}^n = \frac{1}{(N_{\Delta} * n) - 1} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} (r_{i,t} - \bar{r})^2 \quad (1)$$

where:

VOL_{Δ}^n – variance of log returns calculated on high frequency data on the basis of last n-days,

$r_{i,t}$ – log return for i_{th} interval on day t with sampling frequency equal Δ , which is calculated in the following way:

$$r_{i,t} = \log C_{i,t} - \log C_{i-1,t} \quad (2)$$

$C_{i,t}$ – close price for i_{th} interval on day t with sampling frequency equal Δ ,

\bar{r} – average of log returns for i_{th} interval on the basis of last n -days with sampling frequency equal Δ , which is calculated in the following way:

$$\bar{r} = \frac{1}{N_{\Delta} * n} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} r_{i,t} \quad (3)$$

N_{Δ} – the amount of Δ intervals during the stock market session,

n – the memory of the process measured in days, used in the calculation of adequate estimators and averages.

One of the first estimators which we choose in our study is the Parkinson estimator (Parkinson, 1980). Initially it was calculated on the basis of daily intervals, but after Martens and Dijk (2007) we decided to use intraday prices, in order to obtain so called realized range. We aggregate high-low ranges for intraday intervals to obtain daily realized range:

$$RR_{\Delta,t} = \sum_{i=1}^{N_{\Delta}} \frac{(h_{i,t} - l_{i,t})^2}{4 \log 2} \quad (4)$$

where:

$RR_{\Delta,t}$ – daily realized range calculated for day t with sampling frequency equal Δ ,

$l_{i,t}$ – log of minimum price ($\log L_{i,t}$) for i_{th} interval on day t with sampling frequency equal Δ ,

$h_{i,t}$ – log of maximum price ($\log H_{i,t}$) for i_{th} interval on day t with sampling frequency equal Δ ,

Next, after Andersen et al. (2001), Taylor and Areal (2000) and Martens and Dijk (2007) we shortly explain the theoretical background behind the concept of realized volatility estimator. Realized volatility computed from high-frequency intraday data is an effective error-free and model-free volatility measure, considering that we choose the optimal sampling frequency. Furthermore, construction of realized volatility is trivial as one simply sums intra-period high-frequency squared log returns (or cross products, for realized covariance⁵), period by period. For example, for a 7-hour market (420 minutes), daily realized volatility based on Δ -minute underlying returns is defined as the sum of the N_{Δ} intra-day squared Δ -minute returns, taken day by day⁶:

$$RV_{\Delta,t} = \sum_{i=1}^{N_{\Delta}} r_{i,t}^2 + 2 \sum_{i=1}^{N_{\Delta}} \sum_{j=i+1}^{N_{\Delta}} r_{j,t} r_{j-i,t} \quad (5)$$

where $C_{i,t}$ denotes the close price of i_{th} interval on day t and $r_{i,t}$ denotes the log-return of i_{th} interval on day t . Assuming that the returns have zero mean and are uncorrelated, and following the discoveries of Andersen, Bollerslev, Diebold and Labys (2001), we can treat

⁵ Having known the big importance of the correlation of returns we decided to leave this subject for consideration in forthcoming papers.

⁶ Andersen, Bollerslev, Diebold and Labys (1999a).

$\sum_{i=1}^{N_{\Delta}} r_{i,t}^2$ as consistent and unbiased estimator of daily variance σ_t^2 and formula for daily realized volatility becomes simpler:

$$RV_{\Delta,t} = \sum_{i=1}^{N_{\Delta}} r_{i,t}^2 \quad (6)$$

The Garman-Klass volatility estimator (Garman-Klass, 1980) which utilizes the open and close price in addition to the high and low prices is calculated as:

$$GK_{\Delta} = \sum_{i=1}^{N_{\Delta}} \left[0.5 * (h_{i,t} - l_{i,t})^2 - (2 * \log 2 - 1) * r_{i,t}^2 \right] \quad (7)$$

Next, we calculated Rogers-Satchell volatility estimator (Rogers and Satchell, 1991):

$$RS_{\Delta} = \sum_{i=1}^{N_{\Delta}} \left[(h_{i,t} - o_{i,t}) * (h_{i,t} - c_{i,t}) + (l_{i,t} - o_{i,t}) * (l_{i,t} - c_{i,t}) \right] \quad (8)$$

where:

$o_{i,t}$ – log of open price ($\log O_{i,t}$) for i_{th} interval on day t with sampling frequency equal Δ ,

$c_{i,t}$ – log of maximum price ($\log C_{i,t}$) for i_{th} interval on day t with sampling frequency equal Δ ,

Consecutive estimator, presented by Yang and Zhang (2000) improved the most important imperfections connected with the previous ones. The formula for Yang-Zhang volatility estimator is as follows:

$$YZ_{\Delta}^n = V_o + kV_c + (1 - k)RS_{\Delta}^n \quad (9)$$

where:

$$V_o = \frac{1}{(N_{\Delta} * n) - 1} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} (ro_{i,t} - \bar{ro})^2, \quad ro_{i,t} = \log O_{i,t} - \log C_{i-1,t}, \quad \bar{ro} = \frac{1}{N_{\Delta} * n} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} o_{i,t} \quad (10)$$

$$V_c = \frac{1}{(N_{\Delta} * n) - 1} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} (rc_{i,t} - \bar{rc})^2, \quad rc_{i,t} = \log C_{i,t} - \log O_{i,t}, \quad \bar{rc} = \frac{1}{N_{\Delta} * n} \sum_{t=1}^n \sum_{i=1}^{N_{\Delta}} rc_{i,t} \quad (11)$$

and the constant k chosen to minimize the variance of this estimator is given by:

$$k = \frac{0.34}{1.34 + \frac{n * N_{\Delta} + 1}{n * N_{\Delta} - 1}} \quad (12)$$

Alternate formula for volatility estimator was presented in Hansen and Lunde (2006b). Their kernel-based estimator tried to remove the bid-ask bounce by adding autocovariances to the realized variance. In our research we included $_{AC1}RV_{\Delta}^n$, which incorporates the first-order autocovariance:

$$_{AC1}RV_{\Delta} = \sum_{i=1}^{N_{\Delta}} r_{i,t}^2 + 2 \sum_{i=2}^{N_{\Delta}} r_{i,t} r_{i-1,t} \quad (13)$$

Two last estimators which were taken into account in our study were the concepts of scaled realized range and scaled realized volatility proposed by Martens and Dijk (2007). They suggested a bias correcting procedure based on scaling the realized range/volatility

with the ratio of the average level of daily range/volatility and the average level of the realized range/volatility over the q previous days:

$${}_qRV_{\Delta} = \frac{\sum_{a=1}^q RV_{daily,t-a}}{\sum_{a=1}^q RV_{\Delta,t-a}} RV_{\Delta,t} \quad (14)$$

$${}_qRR_{\Delta}^n = \frac{\sum_{a=1}^q RR_{daily,t-a}}{\sum_{a=1}^q RR_{\Delta,t-a}} RR_{\Delta,t} \quad (15)$$

where:

- $RV_{\Delta,t}$ - daily realized volatility calculated for day t with sampling frequency equal Δ ,
- $RV_{daily,t-a}$ - daily realized volatility calculated for day t based on daily data,
- $RR_{\Delta,t}$ - daily realized range calculated for day t with sampling frequency equal Δ ,
- $RR_{daily,t-a}$ - daily realized range calculated for day t based on daily data,

The process of scaling was based upon the idea that daily realized range is (almost) uncontaminated by microstructure noise, and thus provides a good indication of the true level of volatility. While implementing this bias adjustment we have to choose the proper number of trading days q used to compute the scaling factor. Martens and Dijk (2007) suggest that if the trading intensity and the spread do not change for the asset under consideration, q may be set as large as possible to gain accuracy. Naturally, in practice both features tend to vary over time, which suggests that only the recent price history should be used and q should not be set too large. When we consider the data for WIG20 index futures utilized in our research, we see that the spread is relatively small (1 point), but the trading intensity varies significantly over time. This variation in the average spread suggests that the ratio of the average level of the daily realized range relative to the average level of realized range changes over time. Therefore, we decided to compute the scaling factor using the previous q trading days, where q was equal: q = {5, 10, 15, 21, 42, 63, 84, 105, and 126}. Then we choose the most appropriate scaling factor on the base of the estimator efficiency.

Looking at the formulas presented, we can distinguish multi- and one-period estimators, computed in many different ways, thus in the next step we implement two different methods of averaging. Next, we will annualize these estimators assuming that there is 252 working days in a year and we will take the square root of our estimator to obtain standard deviation as a measure of volatility instead of variance.

$$annual_std SD_{\Delta}^n = \sqrt{252 * N_{\Delta} * VAR_{\Delta}^n} \quad (16)$$

$$annual_std YZ_{\Delta}^n = \sqrt{252 * N_{\Delta} * YZ_{\Delta}^n} \quad (17)$$

$$annual_std [vol_est]_{\Delta}^n = \sqrt{252} \sqrt{\frac{1}{n} \sum_{t=1}^n [vol_est]_{\Delta,t}} \quad (18)$$

where:

$$[vol_est] = RR, RV, GK, RS, {}_{AC1}RV, {}_qRV, {}_qRR,$$

$${}^{annual_std}[vol_est]_{\Delta}^n - \text{annualized volatility estimator: } {}^{annual_std}RR_{\Delta}^n, {}^{annual_std}RV_{\Delta}^n,$$

$${}^{annual_std}GK_{\Delta}^n, {}^{annual_std}RS_{\Delta}^n, {}^{annual_std}{}_{AC1}RV_{\Delta}^n, {}^{annual_std}{}_qRV_{\Delta}^n, {}^{annual_std}{}_qRV_{\Delta}^n,$$

Naturally, everywhere we use log we mean natural logarithm. After the presentation of all the formulas for volatility estimators we come to the point of choosing the most adequate n , Δ , and q in order to select the best estimator of true volatility (section six). However, let us describe our data sets first.

4. Data and descriptive statistics

Our empirical analysis is based on high-frequency financial data for WIG20 index futures⁷. WIG20 consists of 20 largest companies quoted on WSE and is computed as a weighted measure of the prices of its components. The 5-minute data, supplied by Information Products Section from WSE⁸, cover the period from June 2, 2003 to July 7, 2007. The number of 5-minute returns for a trading day depends on the trading hours for futures contracts but this have been changed once during our research period. The trading took place from 9:00 a.m. to 4:00 p.m. for the time period from June 2, 2003 until September 30, 2005 and from 9:00 until 4:30 p.m.⁹ for the next two years from October 3, 2005 until July 7, 2007. Thus, we had 84 or 90 five-minute returns for a day in the research period, but in order to conveniently define delta-minute returns, we removed all prices recorded after 4:00 p.m., and as a result we were left with 84 five-minute returns during the day.¹⁰ All returns were computed as the first difference in the regularly time-spaced log prices of WIG20 index futures, with the overnight return included in the first intraday return. After correction for outliers (three on the basis of five-minute intervals and two on the daily basis) we get a total of 1031 trading days and a total of 86414 five-minute intervals. The intraday intervals for different delta (taking daily prices into consideration as well) were obtained from the basic five-minute data set.

Table 4.1 summarizes the descriptive statistics for five-minute data interval divided into two subsets: the original data (returns-I) and the data after modification (returns-II) described above. We include this comparison in order to show that our modification does not significantly change the properties of the data set used in our research. Analyzing both returns, we can see high kurtosis and negative, but small skewness. The average returns are small and are not significantly different from zero. The distribution of the returns is

⁷ We based our study on the continuous time series for futures, where expiring futures contract was replaced by the next series, where the number of open positions achieved the higher value. Described mechanism is one of the most common ways of creating continuous time series for futures. We do not have enough data for the longer period of time because of the short time to expiration of individual future contract, for that reason we had to create continuous futures index.

⁸ Warsaw Stock Exchange.

⁹ In practice, the continuous trading finished at 4:10 p.m., then the close price was settled between 4:10 p.m. and 4:20 p.m., and next investors could trade until 4:30 p.m. only on the basis of close price. Therefore, we could say that we have 86 instead of 90 intervals in the second period.

¹⁰ We adjusted high, low and close price in the last 5-minute interval (from 3:55 p.m. to 4:00 p.m.) with the information included in the intervals from 4:00 p.m. to 4:10 p.m. and the daily close price. Table 4.1 presents the descriptive statistics for our data set (five-minute returns) with and without this restriction as to confirm that this transformation did not influence its properties.

leptokurtic, i.e. it is almost symmetric and has fat tails and a substantial peak at zero. Testing for normality we get the same results for both data sets, i.e. the statistics reveal non-normality of the data sets tested. The most important is that the descriptive statistics for both data sets do not differ significantly, which allows us to use modified data set (returns-II) in the research.

Table 4.1. The descriptive statistics for log returns of analyzed index futures returns for the period from June 2, 2003 to July 7, 2007.^a

		returns-I ^b	returns-II ^c
	N	86414	84818
	Mean	0.000013534	0.000012834
	Median	0	0
	Variance	2.3131977E-6	2.3085858E-6
	Std Dev	0.0015209	0.0015194
	Minimum	-0.0305313	-0.0305313
	Maximum	0.0279413	0.0279413
	Kurtosis	24.2092143	23.9819602
	Skewness	-0.1556275	-0.1774762
	P1	-0.0041335	-0.0041140
	P5	-0.0021031	-0.0021224
	P10	-0.0014276	-0.0014461
	P90	0.0014535	0.0014724
	P95	0.0021146	0.0021330
	P99	0.0042061	0.0041754
Test for Normality			
Kolmogorov-Smirnov	Statistic	0.105333	0.105758
	p-value	<0.0100	<0.0100
Cramer-von Mises	Statistic	379.7725	356.5796
	p-value	<0.0050	<0.0050
Anderson-Darling	Statistic	2122.372	1988.245
	p-value	<0.0050	<0.0050

^a The table contains the descriptive statistics for five-minute returns for original data and the data with prices recorded between 9:00 and 16:00 only. The statistics presented above are: number of observation, mean, median, variance, standard deviation, minimum, maximum, kurtosis, skewness and normality tests. ^b the original data. ^c the modified data with prices recorded between 9:00 and 16:00 only.

Following the discoveries of Ait-Sahalia et al. (2005) we wanted to answer the question if we really need the data sampled as often as possible or 5-, 10-, 15- or even 30-minute intervals are enough. Therefore, we tried to choose the optimal sampling frequency (Δ parameter), i.e. the best interval when we consider maximizing the efficiency and minimizing the microstructure bias (bid-ask bounce and infrequent trading bias) combined with autocovariance bias. Areal and Taylor (2000) agreed that five-minute returns are the highest that avoid distortions from microstructure effect such as bid-ask bounce. On the other hand, Oomen (2001) suggests that the use of equally-spaced thirty-minute returns strikes satisfactory balance between the accuracy of the continuous-record and the market microstructure frictions. We wanted to answer the same question on the basis of the data being researched. Finally, we decided to choose a range of sampling frequencies in the process of calculating the various volatility estimators and leave the moment of selecting it for the next section, where we focus on the efficiency of the estimator.

The point of our interest in the next sections will be to find the most efficient estimator with regard to the estimation parameters: q , n and $delta$. In the process of calculation we used $delta$ which are divisors of 420 (the number of minutes during the normal stock market session), i.e. $delta = \{5, 10, 15, 30, 60, 105, 210, 420\}$. As for the memory parameter (n) and the scaling factor for scaled estimator (q) we establish our set of possibilities basing on the time intervals, i.e. equivalent of one day, week, month, etc., naturally taking into account only working days. Therefore, we calculated volatility estimators for the following parameters: $n = \{1, 5, 10, 15, 21, 42, 63, \text{ and } 126\}$ and $q = \{5, 10, 15, 21, 42, 63, 84, 105, \text{ and } 126\}$.

5. The optimal sampling selection and the length of the memory of the process.

In order to solve the problem of optimal sampling selection, we had to calculate the volatility estimators for all values of q , n , and $delta$ parameters. Given that we used 9 different possible values of q and 8 different possible values of n and $delta$, in result we got a quite substantial number of combinations for each volatility estimator:

1. 64 possible values for: $^{annual_std}RR_{\Delta}^n$, $^{annual_std}RV_{\Delta}^n$, $^{annual_std}GK_{\Delta}^n$, $^{annual_std}RS_{\Delta}^n$, $^{annual_std}_{AC1}RV_{\Delta}^n$, $^{annual_std}YZ_{\Delta}^n$, $^{annual_std}SD_{\Delta}^n$, which depend on n and $delta$ parameters (8x8 possibilities).
2. 576 possible values for scaled estimators: $^{annual_std}_qRR_{\Delta}^n$, $^{annual_std}_qRV_{\Delta}^n$, which depend on n , q and $delta$ parameters (8x9x8 possibilities), which totaled in 1600 volatility estimators.

Having calculated these estimators, we had to choose the adequate methodology for measuring the efficiency of the estimator. We agreed with Yang and Zhang (2000) that the variance/standard deviation of an estimator measures the uncertainty of the estimation, i.e. the smaller the variance/standard deviation, the more efficient the estimation. They concluded that, from the point of view of both theoretical consideration and practical application, it is desirable to find the minimum variance/standard deviation unbiased estimator. Garman and Klass (1980) who defined the efficiency of a variance estimator to be the ratio of the variance of the classical daily volatility estimator to that of current estimator, claimed that the higher the ratio, the more efficient the estimator for a given number of periods. Primal formula for variance efficiency ratio was:

$$Eff = \frac{Var(VAR_{daily}^n)}{Var([var_est]_{\Delta}^n)} \quad (19)$$

where:

Eff - variance efficiency ratio,

$Var(VAR_{daily}^n)$ - variance of the classical variance estimator,

$Var([var_est]_{\Delta}^n)$ - variance of the specific variance estimator,

The puzzling thing in the above formula was that it compared the estimators calculated on the basis of different intervals ($delta$ parameter) and various price histories (parameter n), what could be misunderstandable in the process of parameters selection. When we use different n or $delta$, as a result we do not get the most efficient estimator with regard to its formula but the outcome efficiency is conditional on parameter n and parameter $delta$. We

tried to avoid this imperfection while estimating the efficiency of our estimators. Therefore, basing on the definition of Garman and Klass we define new indicator: modified volatility efficiency ratio (mEff) as the ratio of standard deviation of annualized SD (calculated on the basis of specific interval Δ for the last n days) to the standard deviation of the specific volatility estimator (calculated on the basis of the same Δ and n parameters):

$$mEff = \frac{std^{(annual_std)}(SD_{\Delta}^n)}{std^{(annual_std)}[vol_est]_{\Delta}^n} \quad (20)$$

where:

mEff - modified volatility efficiency ratio,

$std^{(annual_std)}(SD_{\Delta}^n)$ - standard deviation of annualized classical volatility estimator,

$std^{(annual_std)}[vol_est]_{\Delta}^n$ - standard deviation of annualized volatility estimator,

We utilized mEff to choose the optimal parameter q , with regard to the efficiency of scaled estimators (table 5.1-5.2 and figure 5.1-5.2 contained the results for RR and table 5.3-5.4 and figure 5.3-5.4 present the results for RV). Unfortunately, the construction of mEff makes it relatively difficult to compare our estimators with respect to different Δ and n , independently of their absolute values. Therefore, we decided to calculate the relative volatility efficiency ratio (rEff) where numerator and denominator were divided by mean of adequate volatility estimator in order to correct imperfection mentioned above. We used rEff in the process of choosing the optimal frequency Δ and the most appropriate price history (parameter n). Naturally, line of reasoning remained the same. The higher ratio indicated the more efficient the estimator.

$$rEff = \frac{std^{(annual_std)}(SD_{\Delta}^n) / mean^{(annual_std)}(SD_{\Delta}^n)}{std^{(annual_std)}[vol_est]_{\Delta}^n / mean^{(annual_std)}[vol_est]_{\Delta}^n} \quad (21)$$

where:

rEff - relative volatility efficiency ratio,

$mean^{(annual_std)}(SD_{\Delta}^n)$ - mean of annualized classical volatility estimator,

$mean^{(annual_std)}[vol_est]_{\Delta}^n$ - mean of annualized volatility estimator,

We were not able to present the results for all possible combinations of Δ parameter, so we chose Δ equal 5- and 30-minute interval, as the ones most often chosen in the literature. As for the lack of free place we present here only the results for two Δ s but all the remaining lead us to the same conclusion concerning the selection of q . The figures 5.1-5.2 and tables 5.1-5.2 present the results for scaled RR with Δ equal 5 and 30 minute for all n and all q .

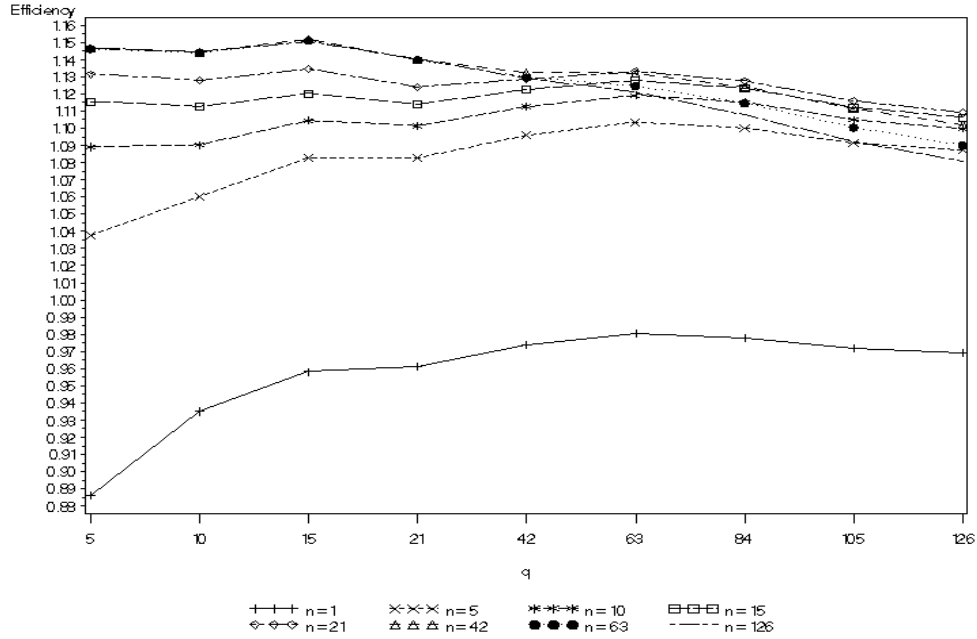
The efficiency ratios for volatility estimators presented in this section were calculated for the period from June 3, 2004 to July 7, 2007. We have to leave 252 daily data (from June 3, 2003 to July 7, 2007) because such amount is required to calculate the first value of $^{annual_std}_{126}RR_{\Delta}^{126}$ and $^{annual_std}_{126}RV_{\Delta}^{126}$ (126 daily data for scaled parameter and 126 in the process of averaging and annualizing).

Table 5.1. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RR^n_q$ for all n under investigations.^a

$\frac{n}{q}$	1	5	10	15	21	42	63	126
5	0.886	1.038	1.089	1.116	1.132	1.146	1.146	1.147
10	0.935	1.060	1.090	1.113	1.128	1.144	1.144	1.145
15	0.958	1.083	1.105	1.120	1.135	1.152	1.151	1.151
21	0.961	1.083	1.102	1.114	1.124	1.140	1.140	1.140
42	0.974	1.096	1.113	1.123	1.129	1.132	1.130	1.129
63	0.981	1.104	1.119	1.128	1.133	1.132	1.125	1.121
84	0.978	1.100	1.115	1.123	1.128	1.124	1.115	1.108
105	0.972	1.091	1.105	1.112	1.116	1.111	1.101	1.092
126	0.969	1.087	1.100	1.106	1.109	1.102	1.090	1.081

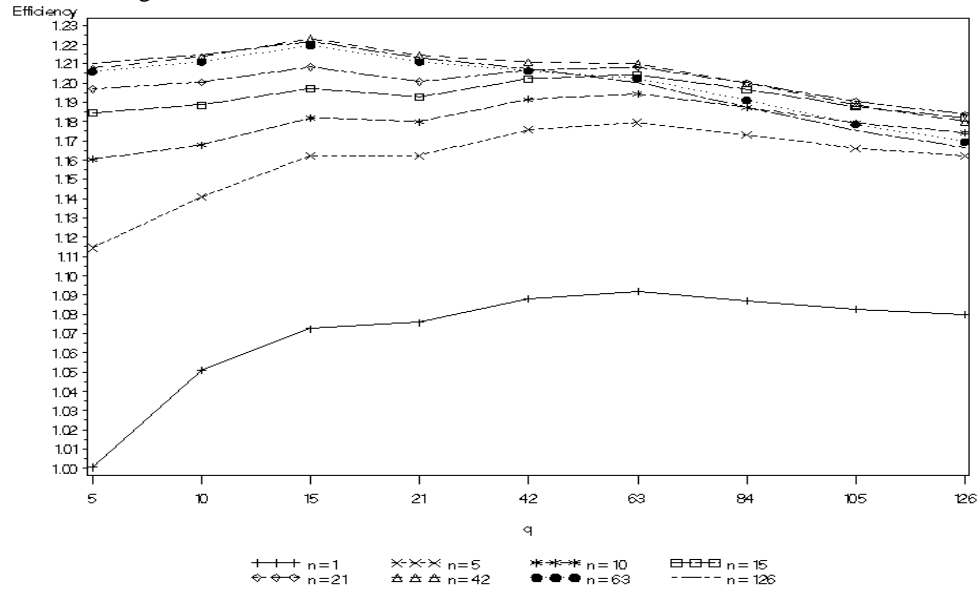
^a scaled RR is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $\delta = 5$ -minute interval.

Figure 5.1. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RR^n_q$ for all n under investigations.^a



^a scaled RR is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $\delta = 5$ -minute interval.

Figure 5.2. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RR^n_{30}$ for all n under investigations.^a



^a scaled RR is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for 30-minute interval.

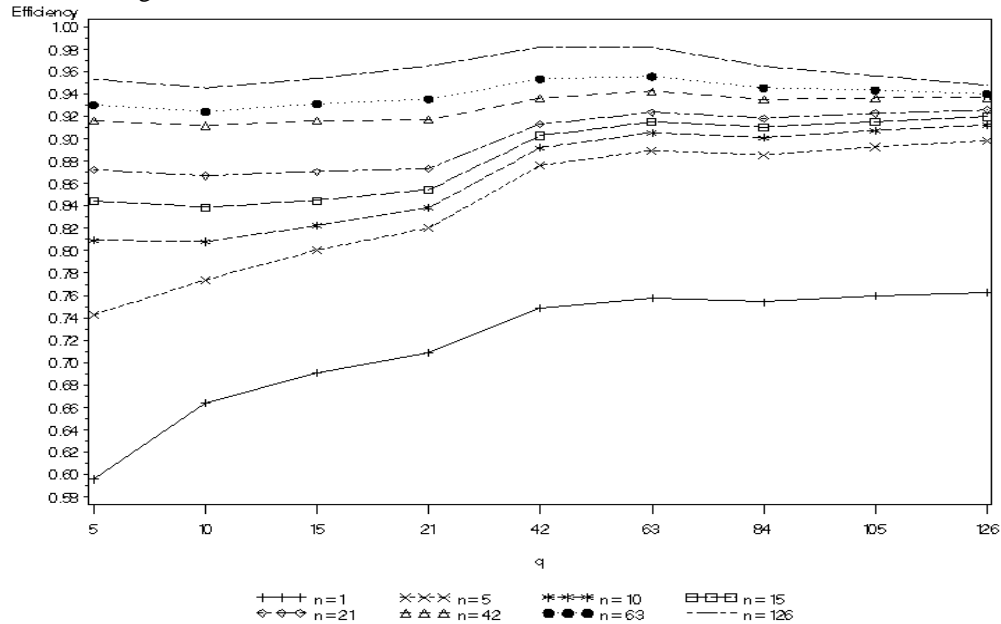
Table 5.2. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RR^n_{30}$ for all n under investigations.^a

$\begin{matrix} n \\ q \end{matrix}$	1	5	10	15	21	42	63	126
5	1.001	1.115	1.161	1.185	1.197	1.208	1.206	1.210
10	1.051	1.141	1.168	1.189	1.201	1.214	1.211	1.215
15	1.072	1.162	1.182	1.197	1.208	1.223	1.220	1.222
21	1.076	1.162	1.180	1.193	1.201	1.214	1.211	1.213
42	1.088	1.176	1.192	1.202	1.207	1.211	1.206	1.207
63	1.092	1.180	1.194	1.204	1.208	1.210	1.202	1.200
84	1.087	1.173	1.187	1.197	1.200	1.200	1.191	1.187
105	1.082	1.166	1.179	1.188	1.190	1.189	1.179	1.175
126	1.080	1.162	1.174	1.182	1.184	1.180	1.170	1.166

^a scaled RR is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $\delta = 30$ -minute interval.

Analyzing figures 5.1-5.2 and tables 5.1-5.2 we choose $q=63$ days as the parameter maximizing the modified variance efficiency ratio (for the largest number of parameter n) what implies minimizing the variance of the estimator with chosen q .

Figure 5.3. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RV^n_q$ for all n under investigations.^a



^a scaled RV is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $\delta = 5$ -minute interval.

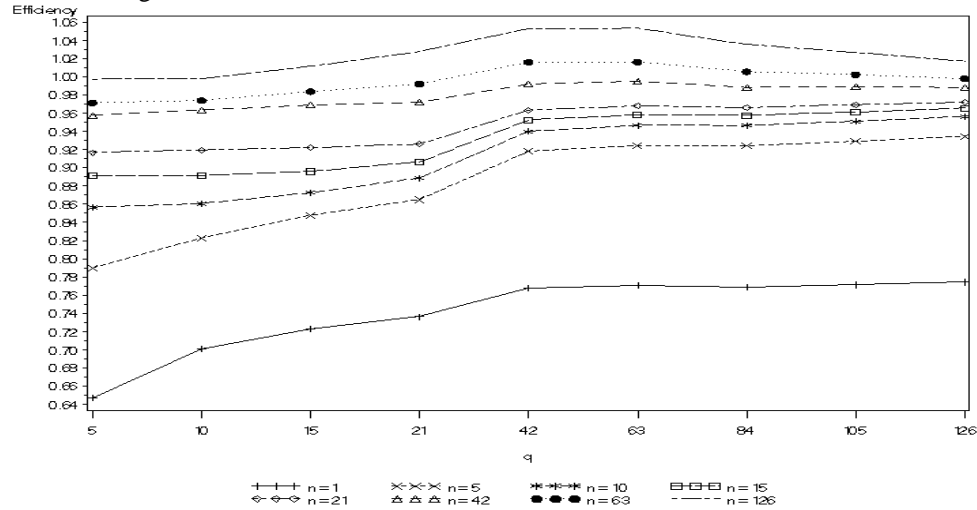
Table 5.3. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RV^n_q$ for all n under investigations.^a

$\frac{q}{N}$	1	5	10	15	21	42	63	126
5	0.596	0.743	0.809	0.844	0.872	0.916	0.930	0.953
10	0.664	0.774	0.808	0.839	0.867	0.912	0.924	0.945
15	0.691	0.801	0.822	0.845	0.871	0.916	0.931	0.954
21	0.709	0.820	0.838	0.854	0.873	0.917	0.936	0.965
42	0.749	0.876	0.892	0.903	0.913	0.936	0.953	0.981
63	0.758	0.890	0.905	0.915	0.924	0.943	0.956	0.982
84	0.754	0.886	0.901	0.910	0.918	0.935	0.946	0.965
105	0.759	0.893	0.907	0.915	0.923	0.936	0.943	0.956
126	0.763	0.898	0.912	0.920	0.926	0.937	0.940	0.948

^a scaled RV is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $\delta = 5$ -minute interval.

The figures 5.3-5.4 and tables 5.3-5.4 present the results for scaled RV with δ equals 5 and 30 minute for all n and all q . The selection of parameter q for scaled RV is the same as for scaled RR, i.e. $q=63$, what additionally confirms our previous selection.

Figure 5.4. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RV^n_{30}$ for all n under investigations.



^a scaled RV is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta=30-minute interval.

Table 5.4. The modified volatility efficiency ratio (mEff) for $^{annual}_{std}RV^n_{30}$ for all n under investigations.

$\frac{n}{q}$	1	5	10	15	21	42	63	126
5	0.648	0.790	0.857	0.891	0.917	0.958	0.971	0.997
10	0.701	0.823	0.861	0.892	0.919	0.963	0.974	0.998
15	0.723	0.848	0.873	0.896	0.922	0.969	0.984	1.012
21	0.737	0.865	0.889	0.907	0.926	0.972	0.992	1.027
42	0.768	0.918	0.940	0.952	0.963	0.992	1.016	1.053
63	0.771	0.925	0.947	0.958	0.968	0.995	1.016	1.053
84	0.769	0.924	0.947	0.958	0.966	0.988	1.006	1.036
105	0.772	0.929	0.951	0.961	0.969	0.989	1.002	1.027
126	0.774	0.935	0.956	0.966	0.972	0.988	0.998	1.017

^a scaled RV is computed according to formula (18) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta = 30-minute interval.

After selecting q we focused on parameter n which was responsible for the length of the memory of the process. We had to solve some kind of the optimizing problem between long memory (high n) - smoothed volatility estimators with hardly any noise but low accuracy and short memory (low n) - volatile estimators which were highly infected by the noise factor but additionally characterized with high accuracy. We introduce the relative volatility efficiency ratio (rEff - formula (21)) in order to choose the best n maximizing the efficiency of the estimator regardless of delta parameter.

We base the process of selecting of parameter n on interval data equal 5- and 30-minute, which were found in the theoretical literature as the best compromise between:

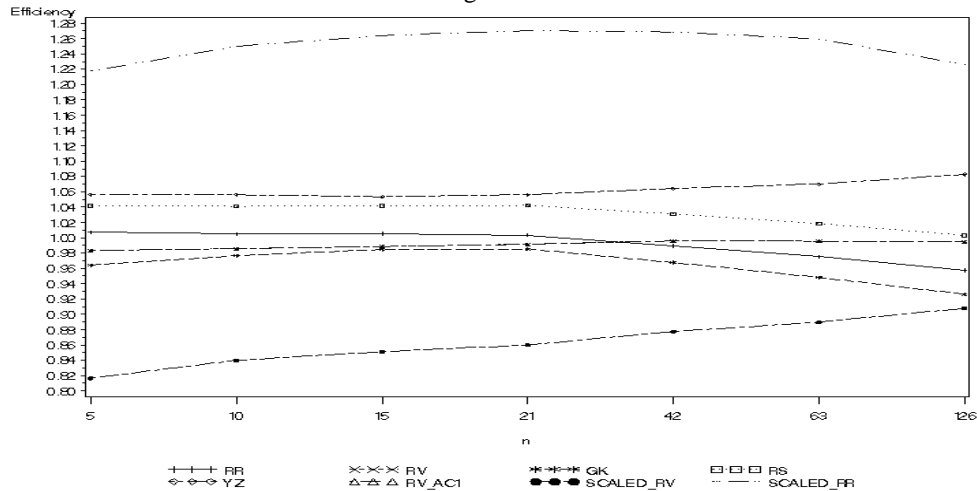
1. maximizing the frequency of the data in order to get the best estimator of true volatility,
 2. controlling the microstructure bias which increases significantly when we base our calculation on the interval sampled with the frequency which is too high.
- We base the process of selection of n on the rEff instead of mEff because we want to make the comparison robust to the absolute values of standard deviation of our estimators.

Table 5.5. The relative volatility efficiency ratio (rEff) for volatility estimators for delta = 5-minute interval and for all n under investigations.^a

$\text{vol_est} \backslash n$	1	5	10	15	21	42	63	126
$\text{annual_std } RR_5^n$	1.006	1.007	1.005	1.005	1.003	0.989	0.976	0.958
$\text{annual_std } RV_5^n$	0.925	0.983	0.986	0.989	0.992	0.996	0.995	0.995
$\text{annual_std } GK_5^n$	0.912	0.964	0.977	0.985	0.985	0.968	0.948	0.926
$\text{annual_std } RS_5^n$	1.035	1.042	1.041	1.042	1.042	1.031	1.019	1.003
$\text{annual_std } YZ_5^n$	1.130	1.056	1.056	1.054	1.057	1.064	1.070	1.083
$\text{annual_std } RV_{AC1}^n$	0.925	0.983	0.986	0.989	0.992	0.996	0.995	0.995
$\text{annual_std }_{63} RR_5^n$	0.958	1.218	1.250	1.264	1.271	1.269	1.260	1.226
$\text{annual_std }_{63} RV_5^n$	0.621	0.817	0.840	0.851	0.860	0.878	0.890	0.908

^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta = 5-minute interval.

Figure 5.5. The relative volatility efficiency ratio (rEff) for volatility estimators for delta = 5-minute interval and for all n under investigations.^a



^a rEff is computed from volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta = 5-minute interval.

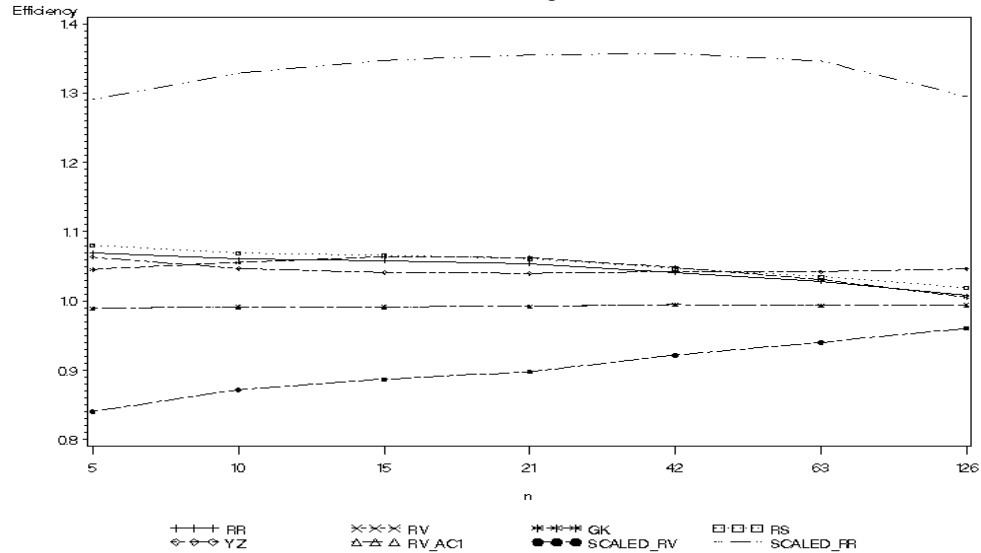
Tables 5.5-5.6 and figures 5.5-5.6 present the comparison criteria (rEff) necessary to select the most adequate n parameter. Basing on the same notion as before, for the consecutive comparisons, we chose $n=42$ and $n=63$ as the parameters maximizing the rEff and therefore minimizing standard deviation of our estimators.

Table 5.6. The relative volatility efficiency ratio (rEff) for volatility estimators for delta = 30-minute interval and for all n under investigations.^a

$\text{vol_est} \backslash n$	1	5	10	15	21	42	63	126
$\text{annual_std } RR_{30}^n$	1.193	1.070	1.060	1.058	1.054	1.041	1.027	1.008
$\text{annual_std } RV_{30}^n$	0.968	0.990	0.992	0.991	0.992	0.995	0.993	0.994
$\text{annual_std } GK_{30}^n$	1.128	1.046	1.056	1.064	1.063	1.048	1.031	1.005
$\text{annual_std } RS_{30}^n$	1.230	1.080	1.069	1.066	1.061	1.047	1.035	1.019
$\text{annual_std } YZ_{30}^n$	1.260	1.063	1.047	1.041	1.040	1.042	1.042	1.046
$\text{annual_std } RV_{AC1}^n$	0.968	0.990	0.991	0.991	0.992	0.995	0.993	0.994
$\text{annual_std } RR_{63}^n$	1.005	1.291	1.329	1.347	1.355	1.357	1.346	1.295
$\text{annual_std } RV_{63}^n$	0.602	0.840	0.872	0.887	0.898	0.922	0.940	0.961

^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta = 5-minute interval.

Figure 5.6. The relative volatility efficiency ratio (rEff) for volatility estimators for delta=30-minute interval and for all n under investigations.^a



^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for delta = 30-minute interval.

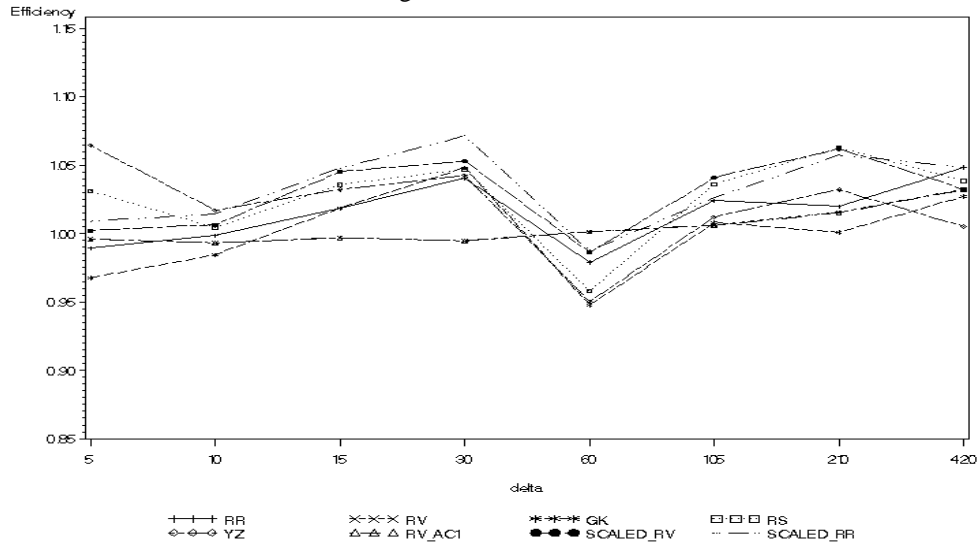
Having selected n and q parameters, finally, we came to the last step, i.e. the optimal sampling frequency selection (δ parameter). We based this step on the volatility estimators calculated for $n=42$ and $n=63$ (chosen in the last stage). The tables 5.8-5.9 and figures 5.8-5.9 present the relative volatility efficiency ratio (rEff) for $n=42$ and $n=63$ for all δ intervals under investigations.

Table 5.8. The relative volatility efficiency ratio (rEff) for volatility estimators for $n=42$ and for all δ interval under investigations.^a

δ	5	10	15	30	60	105	210	420
vol_est								
$annua_std\ RR_{\Delta}^{42}$	0.989	0.998	1.018	1.041	0.979	1.024	1.020	1.049
$annua_std\ RV_{\Delta}^{42}$	0.996	0.993	0.997	0.995	1.001	1.006	1.016	1.032
$annua_std\ GK_{\Delta}^{42}$	0.968	0.984	1.019	1.048	0.948	1.008	1.001	1.027
$annua_std\ RS_{\Delta}^{42}$	1.031	1.004	1.036	1.047	0.958	1.036	1.063	1.039
$annua_std\ YZ_{\Delta}^{42}$	1.064	1.017	1.032	1.042	0.950	1.012	1.032	1.005
$annual_std\ RV_{\Delta}^{42}$ AC1	0.996	0.993	0.997	0.995	1.001	1.006	1.015	1.032
$annual_std\ RR_{\Delta}^{42}$ 63	1.009	1.014	1.048	1.071	0.987	1.026	1.057	1.049
$annual_std\ RV_{\Delta}^{42}$ 63	1.002	1.007	1.045	1.053	0.986	1.041	1.062	1.032

^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $n=42$.

Figure 5.8. The relative volatility efficiency ratio (rEff) for volatility estimators for $n=42$ and for all δ interval under investigations.^a



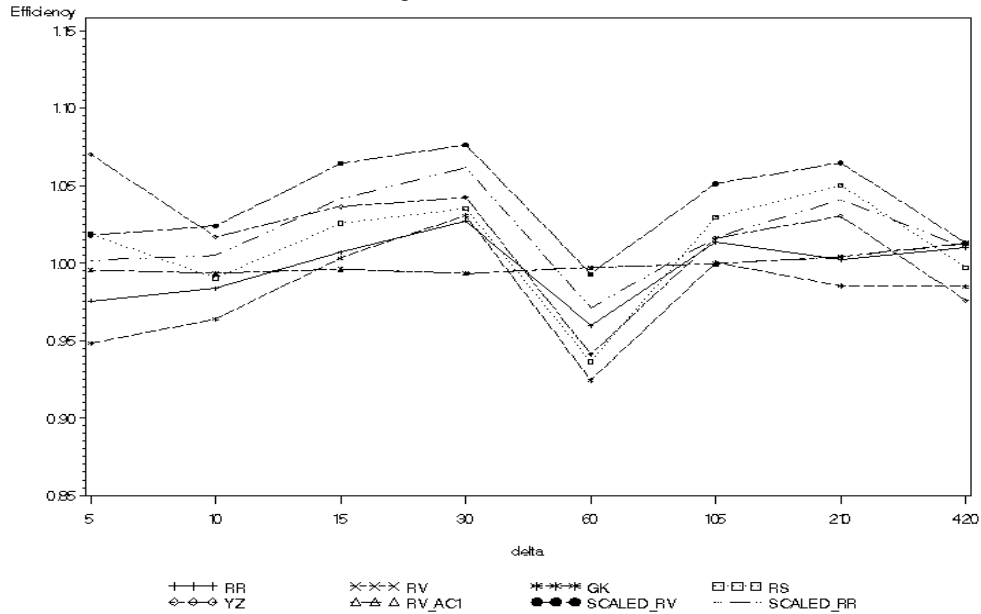
^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for $n=42$.

Table 5.9. The relative volatility efficiency ratio (rEff) for volatility estimators for n=63 and for all delta interval under investigations.^a

<i>vol_est</i>	<i>delta</i>	5	10	15	30	60	105	210	420
<i>annua_lstd</i> RR_{Δ}^{63}		0.976	0.984	1.007	1.027	0.960	1.014	1.002	1.010
<i>annua_lstd</i> RV_{Δ}^{63}		0.995	0.994	0.996	0.993	0.997	1.000	1.004	1.013
<i>annua_lstd</i> GK_{Δ}^{63}		0.948	0.964	1.003	1.031	0.924	1.000	0.985	0.985
<i>annua_lstd</i> RS_{Δ}^{63}		1.019	0.990	1.026	1.035	0.936	1.030	1.050	0.997
<i>annua_lstd</i> YZ_{Δ}^{63}		1.070	1.017	1.037	1.042	0.941	1.016	1.030	0.976
<i>annual_std</i> RV_{Δ}^{63}		0.995	0.994	0.996	0.993	0.997	1.000	1.004	1.013
<i>annual_std</i> RR_{Δ}^{63}		1.002	1.005	1.042	1.062	0.971	1.017	1.041	1.010
<i>annual_std</i> RV_{Δ}^{63}		1.018	1.024	1.064	1.076	0.993	1.051	1.065	1.013

^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for n=63.

Figure 5.9. The relative volatility efficiency ratio (rEff) for volatility estimators for n=63 and for all delta interval under investigations.^a



^a rEff is computed for volatility estimators according to formula (21) for the data covering all period under investigation (from June 2, 2003 to July 7, 2007) for n=63.

We chose $\delta = 30$ (the parameter which maximizes rEff) as the best periodicity for the last stage of our research, i.e. the selection of the best estimator of true volatility. We noticed that the differences between efficiencies of consecutive estimators are not as

significant as it was presented in the previous papers (Martens and Dijk, 2007, Yang and Zhang, 2000, Garman-Klass, 1980, etc.) which introduced the new concepts of volatility estimators. The reason for this is that we compare the estimators with the same n , q and δ focusing only on their formulas.

We are also aware of the fact that the results show the relative efficiency in relation to standard deviation of annualized SD. Taking into account that we have the same benchmark (numerator in formula for mEff and rEff) while comparing our volatility estimators, such comparison should present stable results with respect only to the calculation formula employed.

After detailed process of selection we got nine estimators with two different values of n (42 and 63, which in practice does not influence significantly the properties of final estimator, see Tables 6.1-6.2 in the next section) and two different sampling frequencies ($\delta = 5, 30$), which were let for consecutive comparison in order to accomplish our superior aim.

Before we come to the next section we have to stress a few interesting conclusions coming from the analysis presented above:

1. There is only a slight difference between the values of comparison criteria, which is in contrast to the results obtained in the process of previous researches indicating significant influence on efficiency of the estimator.
2. The amplitude of fluctuations of each comparison criterion differs significantly when we consider the selection of optimal n , q or δ , which additionally reveals the influence of each parameter on the final efficiency of the estimator.
3. We identified parameter n and δ as the most important for the final level of volatility.

Our results confirm the ones presented in financial literature where 5- to 30- minute intervals turn out to be the optimal sampling frequency. What is more important, the selection of parameter q ($q=63$) was similar to the presumptions of Martens and Dijk (2007). However, the most important value added of our research is the discussion of the optimal length of the memory of the process (parameter n), the issue which was not examined in the previous papers. This importance is even confirmed by the range of fluctuations of our comparison criteria when we try to choose the most appropriate n .

6. The distribution of different volatility measures

After the process of selection of essential parameter in the previous chapter we will carefully describe the properties of distribution of nine selected volatility measures, calculated on the ground of different theoretical notions, with respect to two different n and δ . Moreover, we added the results for SD daily in order to reference our results to the benchmark widely used in volatility literature.

Tables 6.1-6.6 present the standard statistics of distributions, only for the parameters selected in the previous chapter. Additionally, we compare the results for $\delta=5$, being the higher available frequency, which is the most connected to the true volatility which is instantaneous process.

The descriptive statistics for volatility estimators presented in this section were calculated for the period from November 28, 2003 to July 7, 2007. We have to leave 126 daily data (from June 3, 2003 to November 27, 2003) because such amount is required to calculate the first value of ${}^{annualstd}_{63}RR^n_{\Delta}$ and ${}^{annualstd}_{63}RV^n_{\Delta}$ (63 daily data for scaled parameter and 63 in the process of averaging and annualizing).

Table 6.1. The descriptive statistics for nine different volatility estimators calculated for n=63 and delta=5.^a

		RR^b	RV	GK	RS	YZ	_ΔRV	_qRR	_qRV	SD	SD_d
N		905	905	905	905	905	905	905	905	905	905
Mean		0,1689	0,21	0,1499	0,1678	0,1905	0,21	0,1886	0,2248	0,2103	0,22
Median		0,1642	0,2056	0,1429	0,1641	0,1867	0,2056	0,1799	0,2232	0,2058	0,2165
Variance		0,0021	0,003	0,0018	0,0019	0,0021	0,003	0,0025	0,0033	0,003	0,0031
Std Dev		0,0458	0,0547	0,0423	0,0437	0,046	0,0547	0,0503	0,0575	0,0546	0,0554
Minimum		0,111	0,1366	0,0987	0,1119	0,1268	0,1366	0,1225	0,1375	0,1374	0,1408
Maximum		0,2965	0,3686	0,2635	0,2879	0,3235	0,3686	0,3275	0,3824	0,3686	0,3714
Range		0,1967	0,2394	0,1796	0,1971	0,2161	0,2394	0,2005	0,244	0,2389	0,2306
Kurtosis		0,293	0,5467	0,1034	0,1892	0,4028	0,5468	0,3126	0,2576	0,5409	-0,1633
Skewness		0,9757	0,9829	0,9676	0,9287	0,9076	0,9829	0,9908	0,7539	0,987	0,6789
P1		0,1134	0,1393	0,1002	0,114	0,1286	0,1393	0,1254	0,1397	0,1406	0,1448
P5		0,1167	0,1467	0,1025	0,1174	0,1362	0,1467	0,1305	0,147	0,1473	0,1513
P95		0,27	0,3302	0,242	0,2634	0,2887	0,3302	0,2975	0,3476	0,3306	0,3305
P99		0,2926	0,3644	0,2594	0,2838	0,3195	0,3644	0,3234	0,3778	0,3644	0,3644
CV		0,2712	0,2605	0,2822	0,2604	0,2415	0,2605	0,2667	0,2558	0,2596	0,2518
Test for Normality											
Kolmogorov-Smirnov	Statistic	0,901	0,916	0,892	0,904	0,906	0,916	0,926	0,935	0,920	0,926
	p-value	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
Cramer-von Mises	Statistic	0,117	0,093	0,115	0,121	0,143	0,093	0,142	0,130	0,141	0,129
	p-value	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01

^a The table contains the statistics on the daily basis (for the period from November 28, 2003 to July 7, 2007) calculated for delta and n already selected: number of observation, mean, median, variance, standard deviation, minimum, maximum, range, kurtosis, skewness, CV and normality tests. ^b RR stands for $^{annual\ std}RR_{\Delta}^n$, then RV - $^{annual\ std}RV_{\Delta}^n$, GK - $^{annual\ std}GK_{\Delta}^n$, RS - $^{annual\ std}RS_{\Delta}^n$, YZ - $^{annual\ std}YZ_{\Delta}^n$, _ΔRV - $^{annual\ std}_{ACI}RV_{\Delta}^n$, _qRR - $^{annual\ std}_{63}RR_{\Delta}^n$, _qRV - $^{annual\ std}_{63}RV_{\Delta}^n$, SD - $^{annual\ std}SD_{\Delta}^n$, SD_d - $^{annual\ std}SD_{daily}^n$. CV – coefficient of variation is defined as the ratio of the standard deviation to the mean.

Presented tables (6.1 and 6.2) enable us to describe the properties of the distributions of selected volatility estimators, compare them with SD daily and formulate below mentioned conclusions.

First of all, focusing on mean and median values, we observe significant differences (up to 50%, setting GK and ${}_qRV$ on two opposite extremes). Obviously, mean or median values enable us to reference the concept of efficiency and accuracy, which were not discussed in details in this paper. However, significant differences between tested volatility estimators suggest to us that this subject cannot be omitted in the future research.

Secondly, the analysis of variance and std reveals similar patterns to the previous statistics (the difference in volatility is up to 50% with GK and ${}_qRV$ on two opposite extremes). On the other hand, the coefficient of variation (CV):

$$CV = \frac{StdDev}{Mean} \quad (22)$$

which relates to the relative volatility, is more appropriate for comparison. The comparison of efficiency based on this criterion (CV) informs us that differences in relative efficiency are not significant (less than 10%) and, what is more important, the results have changed, moving GK on the last and ${}_qRV$ on the first position with respect to relative efficiency. Additionally, our results show that referencing to relative efficiency, SD daily has similar efficiency to the remaining volatility estimators. Given observations do not confirm the results of Garman and Klass (1980), Yang and Zhang (2000) or Martens and Dijk (2007).

Thirdly, analysing minimum, maximum and range we noticed that extremes are strictly connected with the previous observation for mean values (i.e. for GK and ${}_qRV$).

Fourthly, kurtosis, skewness and tests for normality indicate departure from normality, which brings our results closer to the previous papers (e.g. Andersen et al. 2001a, 2001b).

Results for different delta ($\Delta=30$ in Table 6.2) and different n ($n=42$, not presented in this paper) confirm and additionally strengthen our conclusions.

In the next step we visualized the fluctuations of volatility estimators in order to refer not only to their efficiency but accuracy as well. Figures 6.1, 6.3 and 6.5 presenting the comparison of RR, RV, GK, YZ and SD_d with $n=63$, 21 and $\Delta=5$, 30 add additional observation to the discussion of accuracy of our estimators. Figures 6.2, 6.4 and 6.6 presenting the fluctuations of RS, ${}_qRR$, ${}_qRV$, SD and SD_d refer to the same issue. Taking into account that the fluctuations of RV_{AC1} do not differ significantly from RV, we presented only RV on the figures below.

Detailed comparison of Figures 6.1-6.6 informs us about significant differences between tested volatility estimators and enables us to formulate the following conclusions:

1. SD_d together with YZ (Figures 6.1 and 6.3) and scaled RV (Figures 6.2 and 6.4) are characterized by maximum fluctuations while GK and RR (Figures 6.1 and 6.2) or RS and scaled RR (Figures 6.2 and 6.4) are characterized by minimum fluctuations. This observation has crucial importance when we consider the appropriateness of specific volatility estimator from the point of view of the over- or under-estimation of actual volatility.
2. The reaction to the volatility jump (especially considering its different magnitude) varies significantly between tested volatility estimators (March 2005 or February and March 2007), which is especially important considering the speed of response of specific volatility estimator to volatility shock.
3. Stability of observed fluctuations changes significantly with respect to different parameter n (Figures 6.1 and 6.5 or 6.2 and 6.6). On the other hand, we observe only slight differences in stability of fluctuations with respect to different parameter delta (Figures 6.1 and 6.3 or 6.2 and 6.4). These observations are additionally confirmed by the statistics presented in Tables 6.1-6.2.

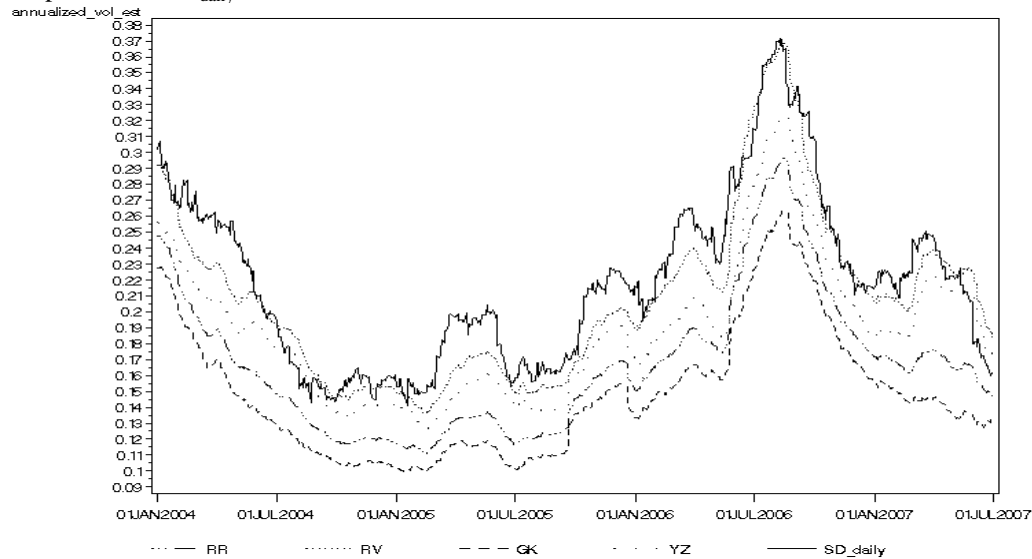
Table 6.2. The descriptive statistics for nine different volatility estimators calculated for n=63 and delta=30.^a

		RR^b	RV	GK	RS	YZ	_ARV	_qRR	_qRV	SD	SD_d
N		905	905	905	905	905	905	905	905	905	905
Mean		0,1799	0,2095	0,1668	0,1814	0,1957	0,2095	0,188	0,2246	0,2097	0,22
Median		0,1742	0,2077	0,1566	0,1755	0,1914	0,2077	0,1797	0,2239	0,208	0,2165
Variance		0,0023	0,0031	0,0021	0,0023	0,0025	0,0031	0,0024	0,0031	0,0031	0,0031
Std Dev		0,0481	0,0558	0,0456	0,0483	0,0503	0,0558	0,0492	0,0561	0,0556	0,0554
Minimum		0,1186	0,1329	0,1108	0,1202	0,1256	0,1329	0,1222	0,132	0,1332	0,1408
Maximum		0,3153	0,3723	0,2904	0,3173	0,3417	0,3723	0,3227	0,376	0,3721	0,3714
Range		0,1855	0,232	0,1648	0,176	0,1967	0,232	0,205	0,2449	0,2312	0,2306
Kurtosis		0,3467	0,5977	0,1856	0,3665	0,5285	0,5979	0,2307	0,1422	0,5844	-0,1633
Skewness		0,9859	0,9141	1,0204	1,0116	0,9404	0,9142	0,9505	0,6295	0,9096	0,6789
P1		0,1216	0,134	0,112	0,1224	0,1284	0,134	0,1266	0,1336	0,135	0,1448
P5		0,1245	0,1394	0,1154	0,1258	0,1363	0,1394	0,1302	0,1445	0,14	0,1513
P95		0,2832	0,3336	0,2631	0,2876	0,3054	0,3336	0,292	0,3409	0,3329	0,3305
P99		0,3111	0,3672	0,2868	0,3139	0,3388	0,3672	0,3179	0,3705	0,3671	0,3644
CV		0,2674	0,2663	0,2734	0,2663	0,257	0,2663	0,2617	0,2498	0,2651	0,2518
Test for Normality											
Kolmogorov-Smirnov	Statistic	0,904	0,935	0,887	0,900	0,919	0,935	0,927	0,937	0,928	0,926
	p-value	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
Cramer-von Mises	Statistic	0,119	0,0907	0,1268	0,1294	0,1374	0,0907	0,146	0,1254	0,1232	0,1295
	p-value	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01

^a The table contains the statistics on the daily basis (for the period from November 28, 2003 to July 7, 2007) calculated for delta and n already selected: number of observation, mean, median, variance, standard deviation, minimum, maximum, range, kurtosis, skewness, CV and normality tests. ^b RR stands for $annua\lstd RR_{\Delta}^n$, then RV - $annua\lstd RV_{\Delta}^n$, GK - $annua\lstd GK_{\Delta}^n$, RS- $annua\lstd RS_{\Delta}^n$, YZ - $annua\lstd YZ_{\Delta}^n$, _ARV - $annua\lstd_{ACI} RV_{\Delta}^n$, _qRR - $annua\lstd_{63} RR_{\Delta}^n$, _qRV - $annua\lstd_{63} RV_{\Delta}^n$, SD - $annua\lstd SD_{\Delta}^n$, SD_d - $annua\lstd SD_{daily}^n$. CV – coefficient of variation is defined as the ratio of the standard deviation to the mean.

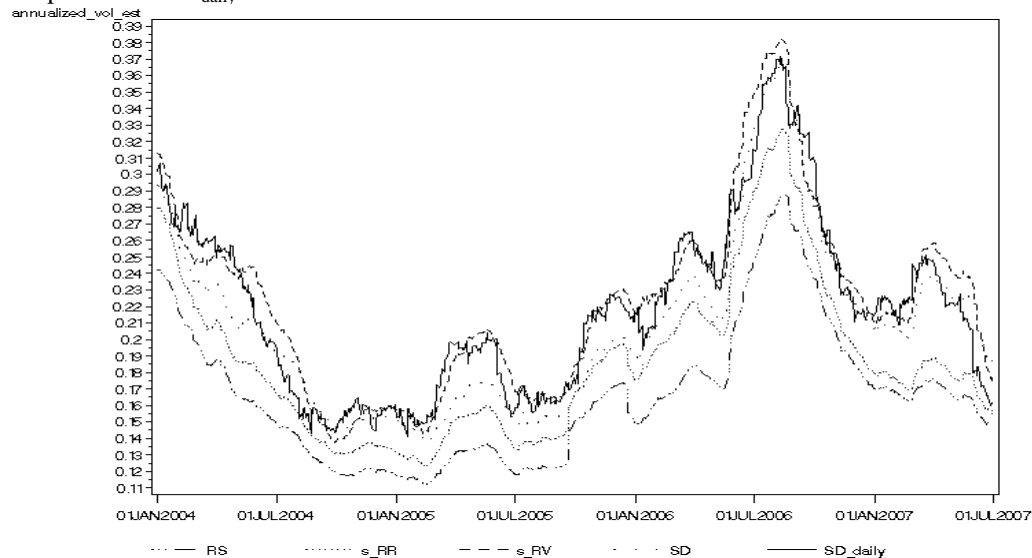
Taking into account conclusions mentioned above and the fact that values of analyzed volatility estimators could differ from 1 to even more than 10 b.p., we regarded the question of accuracy as a very important but sophisticated issue. Therefore, we decided to leave detailed investigation of that subject for consecutive research.

Figure 6.1. The fluctuations of various volatility estimators calculated for $n=63$ and $\delta=5$, compared with SD_{daily} .^a



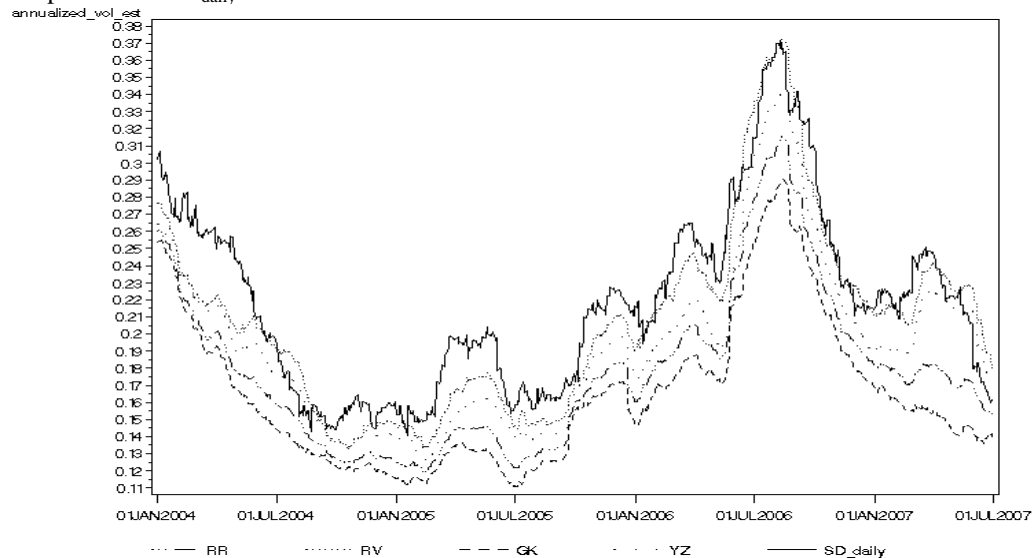
^a The above chart presents volatility estimators: RR, RV, GK and YZ, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=5$ and $n=63$.

Figure 6.2. The fluctuations of various volatility estimators calculated for $n=63$ and $\delta=5$, compared with SD_{daily} .^a



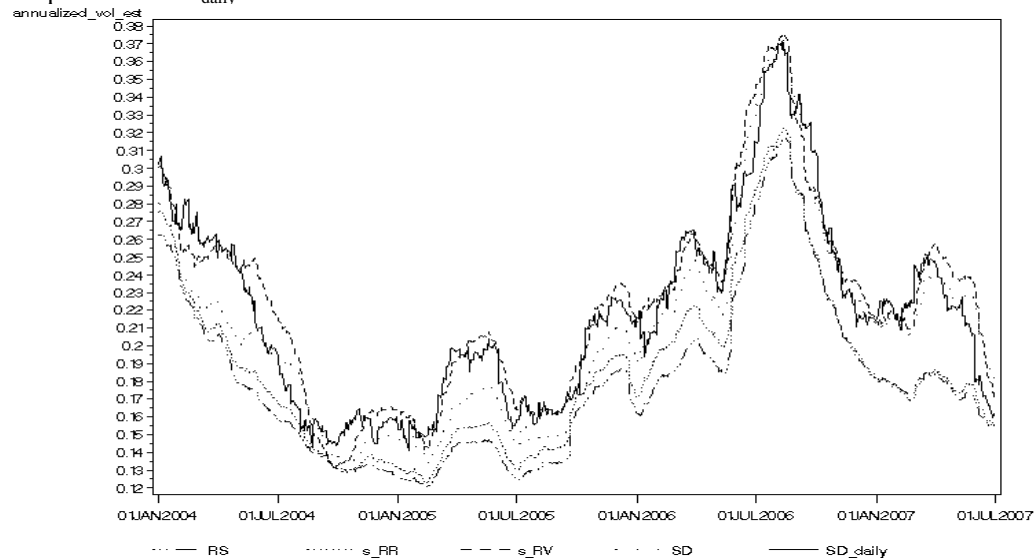
^a The above chart presents volatility estimators: RS, scaled RR, scaled RV and SD, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=5$ and $n=63$.

Figure 6.3. The fluctuations of various volatility estimators calculated for $n=63$ and $\delta=30$, compared with SD_{daily} .^a



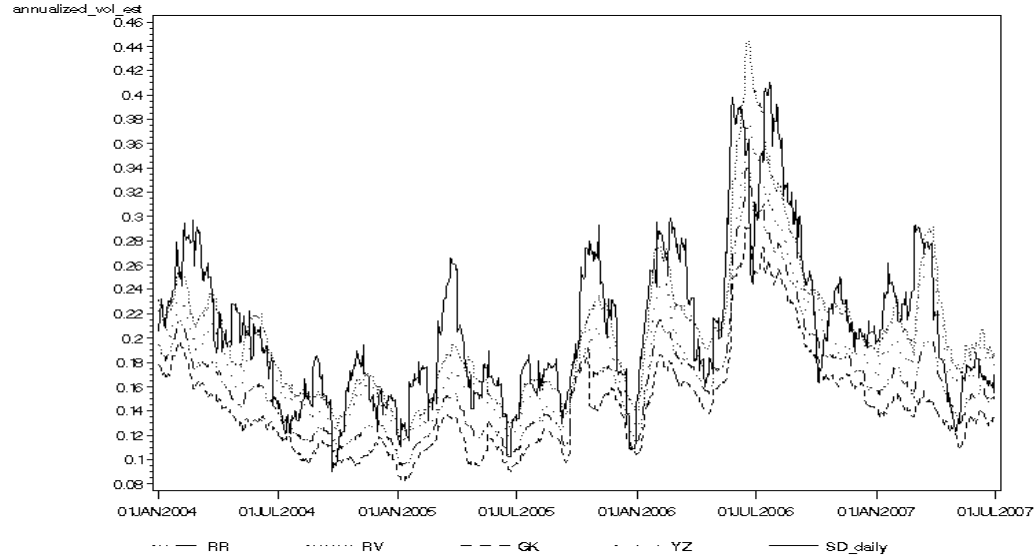
^a The above chart presents volatility estimators: RR, RV, GK and YZ, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=30$ and $n=63$.

Figure 6.4. The fluctuations of various volatility estimators calculated for $n=63$ and $\delta=30$, compared with SD_{daily} .^a



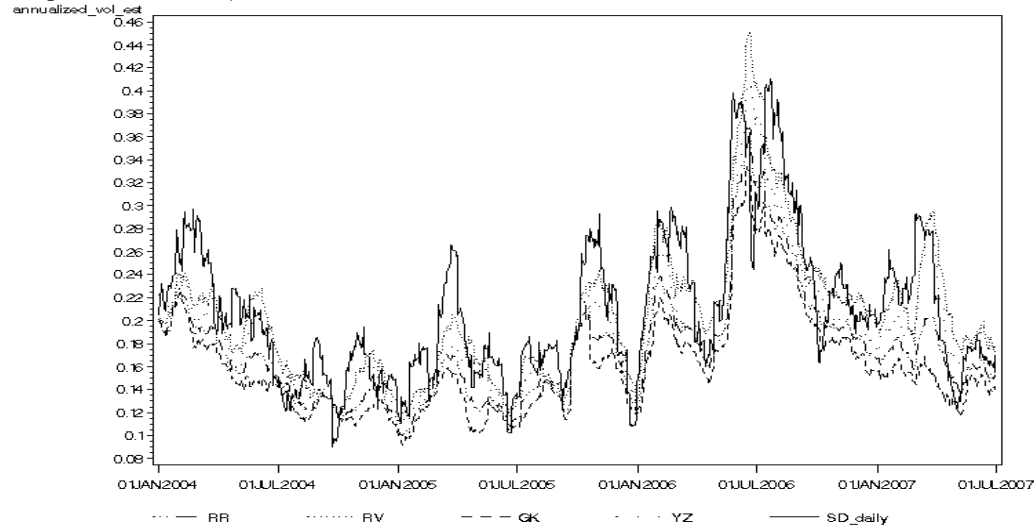
^a The above chart presents volatility estimators: RS, scaled RR, scaled RV and SD, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=5$ and $n=63$.

Figure 6.5. The fluctuations of various volatility estimators calculated for $n=21$ and $\delta=5$, compared with SD_{daily} .^a



^a The above chart presents volatility estimators: RR, RV, GK and YZ, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=5$ and $n=21$.

Figure 6.6. The fluctuations of various volatility estimators calculated for $n=21$ and $\delta=5$, compared with SD_{daily} .^a



^a The above chart presents volatility estimators: RR, RV, GK and YZ, calculated for the period from January 1, 2004 to July 7, 2007 and SD_{daily} for comparison. The calculations were made for $\delta=5$ and $n=21$.

One could argue that at this stage we should also calculate various statistical measures evaluating estimation error (e.g. ME, MSE, RMSE, MAE, MAPE or any other additionally penalizing for over- or under-estimation) in order to accomplish our task which focuses on selecting the best estimator of volatility. We decided to leave the issue of accuracy for consecutive researches since such statistics are biased by definition because we do not know the *true volatility* also called *actual volatility*. Naturally, we can assume e.g. realized volatility

(Andersen et al. 2001a, 2001b) or daily squared returns as the benchmark for actual volatility but than we even do not know the error included in our final statistics. For this reason, we base the process of selection only on the relative comparison, leaving absolute one for future investigation.

7. Market turmoils versus volatility estimators and the implications of the results for financial models

At the end we would like to focus on the interdependence between the turmoils of the market and the level and fluctuations of the volatility. After previous research (Slepaczuk and Zakrzewski, 2007) revealing strong negative dependence between volatility and index futures, we considered the issue of volatility indexes revealing the actual level of risk associated with equity investment as very important subject of future investigations. Analysing the behaviour of volatility indexes (strictly connected with implied volatility concept) we noticed that every downward movement of the market is associated with big jump in volatility. Results for volatility indexes from our previous work show also significant decrease in volatility during upward movement of the market. The phenomenon described, which is called the leverage effect, is distinctly revealed in volatility estimator based on implied volatility rather than various volatility estimators tested in this paper. Emphasized remark, concerning the characteristics of volatility estimators as an input data in different financial models, could be responsible for various results of prediction and consecutive over- or under-prediction of actual volatility, which was deeply investigated by Poon and Granger (2003).

Assuming that this phenomenon exists in all capital, emerging and developed markets we came to the important question: do emerging markets really need sophisticated volatility measures in order to develop further? If yes, then how can we use different concepts of volatility estimators in order to create the benchmark volatility index (possibly similar to existing VIX¹¹ index) robust to existing microstructure biases, enabling to estimate the actual risk of investment on HF data during the stock market session? VIX is one of the best instruments to hedge volatility risk not considering the changes in market price, however, its formula makes it impossible to calculate it for the emerging markets, which do not have liquid and sufficiently developed derivatives market. Therefore, the question arises: can we use any other measure of volatility in order to reflect the fluctuations of actual volatility? Much as the opportunity and the possibility of hedging many different investment strategies with use of the instruments based on volatility index which is negatively correlated with market index prone us to focus on that subject, we consider that this issue requires additional and more detailed investigations.

Focusing on the type of estimator from the point of view of data required in the process of calculation, we can distinguish two main estimator types: periodical (i.e. volatility estimators presented in this paper (RV, RR, GK, etc.) and imminent estimator (volatility index based on implied volatility concept). Periodical estimator requires the data from n last days in order to calculate its value, while volatility indexes require only the information from one point in time, calculating the volatility from the actual prices of options. Discussing the properties of different estimators of volatility but focusing on the cost-benefit analysis, we have to stress that although their fluctuations do not differ significantly, we cannot forget that we need much less data to calculate volatility index in comparison to any other volatility estimator tested in this paper.

The issue of appropriate estimation of actual volatility is of crucial importance to many financial models, so we can directly or indirectly reference our results to:

¹¹ Volatility index for American equity market based on S&P500 index option calculated by CBOE. CBOE introduced volatility derivatives (futures in 2004 and options in 2006), which are heavily traded by market participants. It informs us that there exists a distinctive need to introduce financial instruments which enable to hedge volatility risk independently from market price risk.

- *Asset pricing models (CAPM, APT, etc.)*; where appropriate estimation of actual volatility and its future values has crucial impact on the validity of the results. Over- or under-estimation of volatility results in significant errors in the process of pricing.
- *VaR models*; the main input variable is volatility no matter which type of VaR estimation techniques we choose; inappropriate selection of volatility estimator results in misspecification of risk of actual investments.
- *Option valuation techniques*; in this case the proper estimates and prediction of volatility are so important, that in practice we often say we trade the volatility not the option price.
- *Volatility risk -> Vega hedging -> Risk management*; the importance of *true volatility* measures as the basis instruments for volatility derivatives, essential in further development of option markets and option based strategies (e.g. structured products).
- *Financial stability*, often taken into account in establishing and designing monetary policy, where contemporary financial markets play crucial role, especially when we consider recent turmoils (e.g. subprime mortgage crisis), which are always associated with significant moves in volatility of equity markets.
- *Forecasting volatility*, as the main important subject concerning the volatility issue. Various types of econometric models are used in order to produce the best forecast of future volatility. However, in order to accomplish this task and predict volatility on the basis of any type of GARCH model or historical volatility models, we should first choose the proper volatility estimator as an input variable. Poteshman (2000) indicates the crucial importance of such selection for the final result of the process of forecasting.

8. Conclusions

After detailed comparison of volatility estimators calculated on high-frequency data we verified our initial hypotheses and now we can formulate the following conclusions:

1. The efficiency of volatility estimators, regarding its calculation formula, especially when we consider the relative comparison, does not differ significantly. This is contrary to the previous researches described in the third section. Obviously, we observed some differences resulting in the selection of optimal $n=63$, $\delta=5$ or 30 and $q=63$ for the most efficient estimator but we did not regard the improvement in efficiency as very significant.
2. The appropriate selection of data interval (parameter *delta*) and what is even more important the memory of the process (parameter *n*) plays crucial role in final estimates of actual volatility (described in section 6).
3. The selection of adequate formula for volatility estimator is of crucial importance for the issue of accuracy of actual volatility estimates (Tables 6.1-6.2 and Figures 6.1-6.6), which results in over- or under-estimation.
4. The careful selection of parameter *n* significantly influences the stability of fluctuations of the final level of volatility.

Presented characteristics of distribution of tested volatility estimators in conjunction with conclusions concerning the issue of accuracy of various volatility estimates should enable to build the benchmark volatility index for emerging equity market.

Our value added to this paper concentrates on the discussion of the length of the memory process (parameter *n*) and the observation that the formula of volatility estimator is not as important in comparison of efficiency as the process of selection of the optimal parameter *n* and *delta*.

Consecutive research of this subject should concentrate on higher frequencies (1, 2 minutes and tick data) in order to thoroughly explain the conjunction between gaining on efficiency and increasing various microstructure biases. Moreover, cost-benefit analysis requires detailed investigation (e.g. the degree of improvement in portfolio optimization process when one uses more efficient measure of risk than standard deviation of daily returns). Furthermore, in future

research we will focus on the issue of estimator accuracy, paying special attention to the results of over- or under-estimation of actual level of volatility for different types of investors interested in unbiased estimates and prediction of volatility. Finally, all the issues mentioned above will bring us closer to the main important issue of volatility forecasting, which we will investigate in consecutive papers with respect to accuracy and efficiency of volatility estimator (used as an input variable) influenced by its formula as well as optimal selection of intraday interval and memory of the process (parameter n).

References

1. Ait-Sahalia, Y., Mykland, P.A., Zhang, L., 2005, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* 18, 351–416.
2. Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 47, 1047-1092.
3. Andersen, T.G., 2000, Some reflections on analysis of high-frequency data, *Journal of Business & Economic Statistics* 18, 146-153.
4. Andersen, T., Bollerslev T., 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, 4, 115-158.
5. Andersen T.G., Bollerslev T., 1998, Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts", *International Economic Review*, 39, No.4, 885-905.
6. Andersen T., Andersen T.G., Bollerslev T., Diebold F.X and Labys P., 1999a, Realized volatility and correlation, Manuscript in progress.
7. Andersen T.G., Bollerslev T., Diebold F.X and Labys P., 1999b, Microstructure Bias and Volatility Signatures, Manuscript in progress.
8. Andersen, T., Bollerslev T., Diebold F., and P. Labys, 2000, Exchange rate returns standardized by realized Volatility are (nearly) Gaussian, *Multinational Finance Journal* 4, 159-179.
9. Andersen T.G., Bollerslev T., Diebold F.X, Ebens H., 2001a, The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, 61, 43-76.
10. Andersen, T., Bollerslev T., Diebold F., and P. Labys, 2001b, The Distribution of Realized Exchange Rate Volatility, *Journal of American Statistical Association* 96, 42-55.
11. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003, Modeling and Forecasting Realized Volatility, *Econometrica*, 71, 579-625.
12. Andersen, T.G., Bollerslev, T., Diebold, F.X., Wu J., 2004, Realized beta: persistence and predictability, CFS Working Paper No. 2004/16, Center for Financial Studies.
13. Andersen, T.G., Bollerslev, T., Christoffersen P.F., Diebold, F.X., 2005, Volatility Forecasting, NBER Working Paper 11188, Cambridge.
14. Arak M., Mijid N., 2006, The VIX and VXN volatility measures: Fear gauges or forecasts?, *Derivatives Use, Trading & Regulation* 12, 14-27.
15. Baillie R.T., Bollerslev T., Mikkelsen H.O., 1996, Fractionally integrated generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 74, 3-30.
16. Balaban E., Bayar A., Faff R., 2004, Forecasting stock market volatility: evidence from fourteen countries, Working Paper 02.04, University of Edinburgh.
17. Bali, T.G., Weinbaum, D., 2005, A comparative study of alternative extreme-value volatility estimators, *Journal of Futures Markets* 25, 873–892.
18. Bakshi G., Kapadia N., Madan D., 2003, Stock return characteristics, skew laws and differential pricing of individual equity options, *Review of Financial Studies* 16, 101-143.
19. Bandi, F.M., Russell, J.R., 2005, Microstructure noise, realized variance, and optimal sampling, Graduate School of Business, University of Chicago, Working Paper.
20. Bandi, F.M., Russell, J.R., 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655–692.
21. Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B* 64, 253–280.
22. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2004, Regular and modified kernel-based estimators of integrated variance: the case of independent noise, Oxford Financial Research Centre Economics Series 2004-FE-20.

23. Barone-Adesi, G., Whaley, R., 1987, Efficient Analytic Approximation of American Option Values, *Journal of Finance* 41, 301-320.
24. Becker R., Clements A. E., White S.I., 2006, On the informational efficiency of S&P500 implied volatility, *North American Journal of Economics and Finance*, 17, 139-153.
25. Black, F., and Scholes, M., 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-659.
26. Black F., 1976, Studies of stock market volatility changes, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
27. Bollerslev T., 1986, Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31, 307-327.
28. Bollerslev, T., 1987, "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.
29. Brailsford, T. J., Faff R.W., 1996, An evaluation of volatility forecasting techniques, *Journal of Banking and Finance*, 20, 419-438.
30. Brandt, M.W., Diebold, F.X., 2006, A no-arbitrage approach to range-based estimation of return covariances and correlations. *Journal of Business* 79, 61-74.
31. Campbell, J., Hentschel L., 1992, No news is good news: an asymmetric model changing volatility in stock returns, *Journal of Financial Economics*, 31, 281-318.
32. Chen Z., Daigler R., Parhizgari A., 2006, Persistence of volatility in futures markets, *The Journal of Futures Markets* 26, 571-594.
33. Christensen, B.J., Prabhala, N.R., 1998, The relation between implied and realized volatility, *Journal of Financial Economics*, 50, 125-150.
34. Christensen, K., Podolskij, M., 2005, Realized range-based estimation of integrated variance, *Aarhus School of Business*, manuscript.
35. Corrado, C., and Miller, T., 2005, The forecast quality of CBOE implied volatility indexex, *Journal of Futures Markets*, 25, 339-373.
36. Cox J.C., Ross S., Rubinstein M., 1979, Option pricing. A simplified approach, *Journal of financial Economics* 7, 229-263.
37. Derman E., Demeterfi K, Kamal M., Zou J., 1999, More than you ever wanted to know about volatility swaps, *Quantitative Strategies Research Notes*, Goldman Sachs.
38. Derman E., Kani I., 1994, The Volatility and Its Implied Tree, *Quantitative Strategies Research Notes*, Goldman Sachs.
39. Ebens H., 1999, Realized stock index volatility, Working Paper no. 420, Department of Economics, John Hopkins University, Baltimore.
40. Ederington, L.H., Guan, W., 2000, Measuring implied volatility: Is an average better? Working Paper, University of Oklahoma.
41. Engle R., 1982, Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987-1007.
42. Engle, R. F., and G. G. L. Lee, 1999, A Long-Run and Short-Run Component Model of Stock Return Volatility, in *Cointegration, Causality, and Forecasting*, ed. by Engle, and White. Oxford University Press.
43. Garman M., Klass M., 1980, On the estimation of security price volatilities from historical data, *The Journal of Business* 53, 67-78.
44. Giot P., Laurent S., 2004, Modelling daily Value-at-Risk using realized volatility and ARCH type models, *Journal of Empirical Finance*, vol. 11(3), 379-398.
45. Giot P., 2005, Implied Volatility Indexes and Daily Value at Risk Models” *The Journal of Derivatives*, 54-64.
46. Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, (1993), On the Relation Between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks, *Journal of Finance* 48, 1791-1801.
47. Gospodinov N., Gavala A., Jiang D., 2006, Forecasting Volatility, *Journal of Forecasting* 25, 381-400.
48. Guide to volatility indices of Deutsche Borse, 2007, version 2.4, Deutsche Borse.
49. Hansen, P.R., Lunde, A., 2005, A realized variance for the whole day based on intermittent high-frequency data, *Journal of Financial Econometrics* 3, 525-554.
50. Hansen, P.R., Lunde, A., 2006a, Consistent ranking of volatility models, *Journal of Econometrics* 131, 97-121.

51. Hansen, P.R., Lunde, A., 2006b, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–218 (with comments and rejoinder).
52. Heston, S.L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327-343.
53. Hull J., White A., 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281-300.
54. Kim J., Kim W., Kim K., 2006, Implied volatility using variance decomposition method, MPRA Paper 936/27, Munich.
55. Lowenstein R., 2000, When genius failed. The rise and fall of Long-Term Capital Management, Random House, New York.
56. Martens, M. and J. Zein, 2004, Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility, *Journal of Futures Markets*, 24, 1005-1028.
57. Martens M., Dijk van D., 2007, Measuring volatility with the realized range, *Journal of Econometrics* 138, 181-207.
58. Merton R., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141-183.
59. Merton, R., 1980, On estimating the expected return on the market. An exploratory investigation, *Journal of Financial Economics*, 8, 323-361.
60. Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59: 347-370.
61. Oomen R.C., 2001, Using high frequency stock market index data to calculate, model and forecast realized return variance, EUI Working Paper ECO No.2001/06.
62. Pagan, A. R., Schwert G. W., 1990, Alternative models for conditional stock volatility, *Journal of Econometrics*, 45, 267-290.
63. Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *Journal of Business* 53, 61–65.
64. Patton, A., 2005, Volatility forecast evaluation and comparison using imperfect volatility proxies, London School of Economics, manuscript.
65. Poon S., Granger C.W.J., 2003, Forecasting volatility in financial markets: A review, *Journal of Economic Literature* 41, 478-539.
66. Poteshman A.M., 2000, Forecasting future volatility from option prices, Working Paper, University of Illinois at Urbana-Champaign.
67. Rogers L.C.G., Satchell S.E., 1991, Estimating variance from high, low and closing prices, *The Annals of Applied Probability* 1, 504-512.
68. Shu J., Zhang J., 2003, The relationship between implied and realized volatility of S&P 500 index, *Wilmott Magazine*.
69. Shu J., Zhang J., 2006, Testing range estimators of historical volatility, *The Journal of Futures Markets* 26, 297-313.
70. Slepaczuk R., Zakrzewski G., 2007, VIW20 – Volatility Index for Polish Equity Market, e-Finance 4/2007, Rzeszow.
71. Slepaczuk R., Zakrzewski G., 2008, The daily volatility patterns for realized volatility, manuscript in progress.
72. Taylor S., Areal N.M.P.C., 2002, The realized volatility of FTSE-100 futures prices, *Journal of Future Markets* 22, 627-648.
73. Taylor S.J., Xu X., 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317-340.
74. VIX White Paper, 2003, CBOE Volatility Index, Chicago.
75. Yang D., Zhang Q., 2000, Drift-independent volatility estimation based on high, low, open and close prices, *The Journal of Business* 73, 477-491.
76. Zhang, L., Mykland, P.A., Art-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.
77. Zhang J., Zhu Y., 2006, VIX Futures, *The Journal of Futures Markets* 26, 521-531.



FACULTY OF ECONOMIC SCIENCES
UNIVERSITY OF WARSAW
44/50 DŁUGA ST.
00-241 WARSAW
WWW.WNE.UW.EDU.PL