

Challenges of integrated variance estimation in emerging stock markets*¹

Josip Arnerić², Mario Matković³

Abstract

Estimating integrated variance, using high frequency data, requires modelling experience and data crunching skills. Although intraday returns have attracted much attention in recent years, handling these data is challenging because of their unique characteristics. When dealing with ultra-high frequency or tick-by-tick observations the enormous amount of data needs to be processed prior to estimation of integrated variance for two reasons: eliminating microstructure noise and finding appropriate unbiased estimator. This paper contributes to the existing literature in a two ways. First, we propose how to handle quality issues of the high frequency data due to non-frequent trading and lower liquidity of emerging markets. Second, we find the optimal sampling frequency at slow time scale that should be used to obtain two-time scale estimator of integrated variance for each emerging market under consideration: Romania, Hungary, Bulgaria and Croatia. Empirical results indicate that intraday returns should be sampled every 7 to 10 minutes at slow time scale while the fast time scale should be fixed at the highest possible frequency. Realized variance estimator at the fast time scale mostly overestimates the integrated variance on all stock markets except Bulgaria; on average between 70% and 90% of the time. Moreover, the robustness of the results with respect to the price jumps has been verified for Romania and Hungary, unlike Croatia and Bulgaria, for which we recommend a robust version of two-time scale estimator of integrated variance within truncation technique. It is

* Received: 26-04-2019; accepted: 18-12-2019

¹ Disclaimer: the views expressed herein are those of the author and do not reflect the views of NN Group & Nationale-Nederlanden.

² Associate Professor at Department of Statistics, University of Zagreb, Faculty of Economics and Business, Trg. J. F. Kennedyja 6, 10000 Zagreb, Croatia. Scientific affiliation: financial econometrics, financial time series models and methods, volatility and cointegration, GARCH and MGARCH models, high-frequency data and stochastic processes. Phone: +385 1 238 3361. E-mail: jarneric@efzg.hr. ORCID: <http://orcid.org/0000-0002-2901-2609> (corresponding author).

³ PhD in Economics, Financial Risk Manager at Risk Models Department, NN Group & Nationale-Nederlanden, Prinses Beatrixlaan 35, 2595 AK Den Haag, Netherlands. Scientific affiliation: financial econometrics, realized volatility and OHLC estimators, risk management. E-mail: mmatkic@hotmail.com.

additionally found that intraday returns should be sampled more frequently in a highly volatile periods. These findings offer valuable information to market participants, as they are able to apply the most accurate ex-post volatility measure, as unbiased and consistent estimate of integrated variance.

Key words: *integrated variance, optimal sampling frequency, microstructure noise, jumps, two-time scale estimator, emerging stock market*

JEL classification: *G1, C13, C14, C58*

1. Introduction

Estimating and forecasting problems in finance include unobservable variables such as integrated variance, i.e. the true but unknown volatility measure. Volatility measurement and forecasting is still, and will remain one of the major research subjects for a long time. Many empirical studies in this field have focused on parametric models for computing ex-post measures of volatility, such as GARCH-type models that have gained a great popularity among practitioners and academics. These models consider only daily closing prices, whereby they neglect the intraday information contained in high frequency observations and thus may suffer from imprecise estimation of the variance. Conversely, due to advances in electronic trading and information technology it has become possible to estimate integrated variance using the available intraday price observations.

Many econometricians, including Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shepard (2002; 2006), have considered employing data sampled at very high frequency to compute the ex-post measure of volatility at a lower frequency, known as realized variance (RV). The realized variance is simply defined as the cumulative sum of equally spaced intraday returns available at trading day t . Andersen and Bollerslev (1998) are the first to point out that RV is a more precise ex-post integrated variance measure than a daily squared return, as the latter one is extremely noisy and will not lead to the same outcome that would be obtained if the true variance were observed. It is well documented that RV estimator becomes biased with the increase of sampling frequency (Andersen and Bollerslev, 1998; Hansen and Lunde, 2005; Bandi and Russell, 2008). The sources of the bias are attributed to the autocorrelation of intraday returns resulting from microstructure noise, i.e. non-synchronous trading, discrete price observations, irregularly spaced intervals and bid-ask bounce (Aït-Sahalia et al., 2011; Bandi and Russell, 2008). Usually, higher sampling frequency will lead to a more significant noise problem and the poor quality of intraday observations, especially in emerging stock markets. For instance, a 1 second sampling frequency would result with many gaps in transaction data and many zero prices. On the other hand, Andersen et al. (2003) prove that 30 minutes returns ensure stable realized variance, but this would leave us only a few intraday observations at hand.

In this paper we discuss unique characteristics of high frequency data, such as prices discreteness, irregularly spaced intervals, bid–ask bounce and diurnal pattern. Related to mentioned characteristics we propose how to handle these data to mitigate microstructure noise as much as possible. In the next step we find realized variance estimator that fits integrated variance the best for each of the analyzed emerging market individually by utilizing two-time scale estimator (Zhang et al., 2005; Zhang, 2011). This estimator reduces the bias of RV even it's variance increases and thus an optimal sparse sampling frequency (at slow time scale) has been chosen to balance between the bias and the efficiency of the estimator. However, subsampling method proposed by Zhang et al. (2005) produces an upward biased estimator of the integrated variance in the presence of jumps (Aït-Sahalia et al., 2011). Accordingly, robustness of two-time scale estimator in the presence of potential jumps will be also examined to check if optimal sampling frequency at slow time scale is stable. For that same reason two-time scale estimator will be truncated as proposed by Andersen et al. (2012).

It should be noted that this paper deals with ex-post or realized measure of volatility but not ex-ante or implied volatility. Forward looking nature of markets derivatives, especially futures and options, is a different approach that cannot be applied as derivative markets are not developed or they do not exist in selected emerging countries.

The rest of the paper is organized as follows. Section 2 refers to recent studies and findings. Section 3 describes methodology. The empirical data and analysis are given in section 4. Section 5 presents results with discussion. Concluding remarks, limitations of the present research and directions for the future research are provided in the last section.

2. Literature review

High frequency data are increasingly being used to compute realized variance as a measure of integrated variance that is robust to microstructure noise. Reducing microstructure noise and finding the most accurate proxy of the true but unobservable integrated variance is not straightforward. The advantage of RV is it's simplicity and usage of enormous amount of data by taking a large number of intraday returns J . Namely, when the time interval between intraday returns is small ($\Delta \rightarrow 0$) there is much information that can be used ($J \rightarrow \infty$), but then the returns are contaminated with market microstructure and often result in biased RV. The problem of choosing the optimal sampling interval Δ has attracted significant interest in the literature (Andersen et al., 2001; Aït-Sahalia et al., 2005; Zhang, 2011; Boudt and Zhang, 2013; Arnerić et al. 2019). For example, a 5 minute sampling would result in 84 price observations for one trading day within a

stock market that opens at 9:00 and closes at 16:00, while a 1 second sampling would result in 25200 price observations. For a very thick time interval, there is “infinite” number of intraday returns, but high frequency data are not available for all markets. In empirical studies, realized variance is usually obtained on various sampling frequencies; from 1 minute up to 30 minute returns. The state-of-the-art in this field would be to employ 5 minute returns (Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002) while Andersen et al. (2003) recommend the use of 30 minutes sampling interval, which represents a reasonable trade-off between minimizing microstructural bias and the sampling error.

Under the absence of autocorrelation of intraday returns, RV is a consistent estimator of unobservable integrated variance (Aït-Sahalia et al., 2005). In practice the assumption of identically and independently distributed intraday returns does not hold and that’s why high frequency data should be properly cleaned and sampled. Thus, in what follows we discuss the effects of microstructure noise in the estimation of integrated variance. LeBaron (1992) studied the negative dependence between autocorrelation and realized variance, also known as the “LeBaron effect”, which manifests at daily and weekly levels. He found that autocorrelation changes over time and that is related to the stock returns. Bianco et al. (2009) showed that there is also a negative dependence between autocorrelation and variance on intraday level; whereas Oomen (2005) shows that a careful choice of the optimal sampling frequency can reduce the impact of autocorrelation on the realized variance. At the 1 minute sampling frequency the bias, due to autocorrelation, was estimated at 35%, while the optimal sampling frequency was between 25 and 35 minutes. Apart from autocorrelation, additional source of microstructure noise is the bid-ask bounce. The bid-ask bounce was initially introduced by Roll (1984) who describes that transaction prices “bounce” back and forth between the bid and ask prices. Even when intraday price observations continuously bounce between the bid and ask prices and never exceed its boundaries it might significantly affect the realized variance. Furthermore, in practice intraday prices are not observed continuously and hence suffer from discretization error.

Occurrence of the jumps is yet another problem that may exist. The realized variance does not match the integrated variance in their presence. Barndorff-Nielson and Shepard (2006), Boudt et al. (2012), Boudt and Zhang (2013) as well as Huang and Tauchen (2005) and Corsi et al. (2010) discuss detection of the jumps and propose several jump robust estimators. For example, Barndorff-Nielson and Shepard (2006) have shown that the normalized sum of products of the absolute value of contiguous returns is a consistent estimator for integrated variance, known as the bipower variation (BPV). The bipower variation estimator converges to the same probability limit as realized variance when there are no jumps in the semi-martingale process. The difference between RV and BPV can be helpful in detecting significant jumps that can be used later in realized variance forecasting (Arnerić

et al. 2018). Thus BPV is considered as a robust estimator in the presence of the jumps, but not robust to microstructure noise. Nevertheless, proposed staggered version of BPV estimator, that is robust to both microstructure noise and price jumps, is comprehensively discussed by Huang and Tauchen (2005). According to Huang and Tauchen (2005) technique of sufficiently staggering returns can correct the bias induced by microstructure noise as well. Even so, Huang and Tauchen (2005) considered only developed market high frequency data, i.e. S&P500 index, and didn't provide a strictly defined rule for choosing sufficiently staggering returns. Boudt and Zhang (2013) proposed a jump robust version of the two-time scale estimator of Zhang et al. (2005) and Zhang (2011) that is robust to both microstructure noise and price jumps. Related to this paper, Andersen et al. (2012) have provided evidence that medRV is more efficient and it has improved finite sample performances against BPV. That's why Boudt and Zhang (2013) have used medRV to define a jump robust version of the two-time scale estimator.

Unfortunately, the vast majorities of the studies examine strictly developed financial markets, to a large extent relying on NYSE trading and quotes, or simply use simulated noisy data (Table 1). A fewer studies examine realized variance on the foreign exchange markets, but there is a great lack of studies regarding emerging stock markets.

Table 1: Review of most significant studies with main findings related to the sampling frequency selection and proposed estimators of integrated variance

Author(s)	Data	Sample period	Estimator	Main findings
Andersen at al. 2001	exchange rates for DM/USD and YEN/USD	December 1, 1986 to November 30, 1996	RV	"Sampling frequency of 5 min was chosen, while distribution of realized variance is skewed to the right and leptokurtic"
Aït-Sahalia et al. 2005	Simulated data using Monte Carlo method	10000 samples with length of 1 year	Quadratic variation	"Optimal sampling frequency varies from 4 min to 3 h"
Huang and Tauchen, 2005	S&P500 stock market index	April 21, 1997 to October 22, 2002	Staggered BPV	"Staggering helps to improve the jump detection at high frequency levels of 1 minute and 3 minute returns"
Hansen and Lunde, 2006	Tick-by-tick of 30 Down Jones Industrial Average	January 29, 2001 to December 31, 2004	RV with overnight return	"Sampling frequency of 10 seconds was chosen according to the signature plot of realized variance"

Author(s)	Data	Sample period	Estimator	Main findings
Bundi and Russell, 2008	IBM stock	February 2002 with total 140614 quotes	Bias corrected RV	“Optimal sampling frequency ranges from 2 minutes to 2.7 minutes”
Zhang, 2011	Tick-by-tick stock prices of Microsoft and Google	October 2, 2005 to October 31, 2005	Two scale estimator	“The two scale estimator is stable for large enough sampling frequency (approximately 3.5 min)”
Boudt and Zhang, 2013	Simulated data using Euler scheme	1000 days with 23400 intervals and 1 second increment	Jump robust TSRV	“If trades occur every 2 seconds with jumps, the optimal sampling frequency is half minute”
Arnerić et al. 2019	1 minute returns of 7 emerging markets	January 4, 2010 to April 1, 2016	TTSE	“Optimal sampling frequency varies from 4 min to 12 min and in the lack of intraday prices OHLC estimators should be considered”

Source: Author’s presentation

Hanousek et al. (2013) proposed optimal price jumps indicator with respect to type I and type II errors for Czech Republic, Poland and Hungary, but did not considered finding the most appropriate estimator of integrated variance that is robust to jumps as well as microstructure noise. They concluded that news announcement in selected emerging CEE countries as well as foreign information arriving from U.S. markets affects price jumps. This research mostly relies on previous one (Hanousek et al., 2012) which uses artificial time series only.

This paper also relies on the argument of Brownlees and Gallo (2006), who state that data cleaning is a preliminary necessary condition for moving into the estimation step. Additionally, there is a great disagreement in the literature about the “optimal” sampling frequency for computing “best” proxy of integrated variance. It is more common to propose a higher frequency, for example 1 minute, for liquid assets, while for illiquid assets a much lower frequency should be used. In general, it is suggested in practical implications not to sample too often, i.e. sparse sampling frequency in range from 5 minute intervals to as long as 30 minutes is recommended. Opposite to that, data shouldn’t be sampled too sparsely as it increases the variance of the RV due to discretization error. The different optimal sampling frequencies are found for different assets (Table 1), as proposed by Aït-Sahalia et al. (2005; 2011), Zhang et al. (2005), Hansen and Lunde (2006) and Bandi and Russell (2008). Moreover, the microstructure noise is considered negligible at sampling at lower frequencies than 5 minutes that is common when using a tick-by-tick data. However, tick-by-tick data are not appealing in studies of emerging markets as they are irregularly spaced and poor.

3. Methodology

To determine an appropriate estimator of integrated variance using high frequency data a theoretical background of the underlying data generating process is required. It is common to assume that the logarithmic price $p_t = \log P_t$ follows a continuous stochastic diffusion process known as an Ito process

$$dp_t = \mu_t dt + \sigma_t dw_t \quad (1)$$

where μ_t is the drift, σ_t is a variance process and w_t is a Wiener process, i.e. standard Brownian motion, with independent and stationary increments and also independent from variance process. By taking the integral over a one day interval $[t, t - 1]$ it follows

$$r_t = p_t - p_{t-1} = \int_{t-1}^t \mu_t dt + \int_{t-1}^t \sigma_t dw_t \quad (2)$$

where r_t is continuous return, i.e. a first difference of logarithmic prices. After rearranging the second integral on a right side a daily integrated variance is given as

$$IV_t = \int_{t-1}^t \sigma_t^2 dt \quad (3)$$

From above notation, an integrated variance is usually understood as quadratic variation of semi-martingale process. In practice, however, stock prices are observed in discrete time. There are no price changes observed prior to the opening nor after the closing of any stock market. To estimate the integrated variance a discrete formulation is required. As noted in Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) the integrated variance can be estimated using intraday returns $r_{t,j} = p_{t,\Delta} - p_{t-1,\Delta}$ where Δ is the length of intraday sampling interval and J is a number of non-overlapping intraday returns since $= 1/\Delta$. The cumulative sum of intraday squared returns observed at sampling interval Δ is defined as daily realized variance

$$RV_t^\Delta = \sum_{j=1}^J r_{t,j}^2 \quad (4)$$

Assuming that intervals between observations are equally spaced, commonly referred to as equidistant intervals, a total of J intraday returns can be constructed from the opening to the closing time of the market. The sum of squared intraday returns is an unbiased measure of the integrated variance under certain conditions.

The theoretical justification for this approach is that when the number of intraday observations goes to infinite RV_t^Δ converges in probability to the quadratic variation of the semi-martingale process. It is therefore a consistent and unbiased estimator of integrated variance. However, the intraday returns must be serially uncorrelated and there should exist no bid-ask bounce or any form of microstructure noise that contaminates the result. Although these conditions perhaps do not seem unreasonable, in practice they are often violated.

Using the additive property of logarithmic returns, it follows that the daily return is an aggregation of J intraday returns and the variance of the daily return equals to the sum of intraday variances of returns if and only if they are independently distributed. Otherwise, when intraday returns are correlated, the variance of the sum of intraday returns equals the sum of the variances plus covariances of intraday returns

$$Var(r_t) = Var\left(\sum_{j=1}^J r_{t,j}\right) = \sum_{j=1}^J Var(r_{t,j}) + 2 \sum_{i=1}^{J-1} \sum_{j=i+1}^J Cov(r_{t,i}, r_{t,j}) \quad (5)$$

The last term on the right hand side denotes the covariances of the intraday returns. When the covariances are not equal to zero the realized volatility becomes biased. Thus in practice the realized variance is very seldom unbiased estimator of the true integrated variance (Zhang et al., 2005; Bandi and Russell, 2008; McAleer and Medeiros, 2008). One approach to reduce the bias is using tick-by-tick intraday returns to estimate the variance of the noise and then optimal sampling interval is obtained by minimizing the mean square errors. This approach doesn't use equidistant price observations at hand and basically means that realized variance is empirically evaluated on various sampling intervals. Since the analyzed emerging markets in this paper are mostly under-developed and characterized by infrequent trading we can expect higher sampling interval than 5-minute.

Another approach is to use two-time scale estimator, i.e. market microstructure noise in the RV can be filtered out by using a subsampling method as proposed in Zhang et al. (2005) and in Ait-Sahalia, Mykland and Zhang (2005). Namely, Zhang et al. (2005) assure that sampling over longer intervals only reduces the microstructure noise, rather than correcting its effect for variance estimation and thus data should not throw away. We can use all of the data and still have unbiased and consistent estimator of the integrated variance which leads to the two-time scale estimator (TTSE). The two-time scale estimator uses the highest possible sampling frequency J (fast time scale) to filter out the magnitude of the noise term by subtracting it from the average RV estimator at a sparse frequency (slow time scale). It combines two estimators; entire sample estimator RV_t^Δ and average sparse estimator $RV_{t,\Delta}^S$. The average sparse $RV_{t,\Delta}^S$ is computed over a total of S non-overlapping subsampled returns. For example, if the slow sampling

frequency is 10 minutes, with 1 minute increment, than the first sparse RV uses the prices sampled at 9:00, at 9:10, etc. The second sparse RV is obtained using prices sampled at 9:01, 9:11, etc. and the last sparse RV is obtained by prices sampled at 9:09, 9:19, etc. The average sparse RV using S subsamples is defined for each trading day as

$$RV_t^{\Delta,S} = \frac{1}{S} \sum_{s=1}^S RV_{t,s}^{\Delta} = \frac{1}{S} \sum_{s=1}^S \sum_{j=1}^{n_s} r_{t,j,s}^2 \quad (6)$$

where $r_{t,j,s}$ is the j -th slow frequency scale return of the subsample s and n_s is the number of returns in each subsample.

Finally, TTSE is computed by subtracting the bias term from the average sparse RV to remove the microstructure noise according to

$$TTSE_t^{\Delta} = RV_t^{\Delta,S} - \frac{\bar{n}}{J} RV_t^{\Delta} = \frac{1}{S} \sum_{s=1}^S \sum_{j=1}^{n_s} r_{t,j,s}^2 - \frac{\bar{n}}{J} \sum_{j=1}^J r_{t,j}^2 \quad (7)$$

where $\bar{n} = \frac{J-S+1}{J}$ presents the average size of subsamples. The above defined estimator is corrected for the bias and still uses all available data in estimation of integrated variance. Furthermore, a small sample adjustment can be applied when the number of returns is not equal across subsamples

$$TTSE_t^{\Delta,adj} = \frac{1}{1 - \frac{\bar{n}}{J}} TTSE_t^{\Delta} \quad (8)$$

Sums of squared returns which are S and J steps apart can be additionally truncated to remove the returns that exceed a certain threshold under the assumption of no jumps (Boudt and Zhang, 2013). Truncation can be easily obtained using indicator function $I(\omega)$ depending on the threshold ω as a real number. Usually, the threshold is set to be 9 which indicates that S and J steps apart returns are larger than three standard deviations from the mean

$$I(\omega) = \begin{cases} 0, & \text{otherwise} \\ 1, & \text{if } \frac{r_{t,j,s}^2}{(\int \sigma_t^2 dt + 2\sigma_{\epsilon}^2)} \leq \omega \end{cases} \quad (9)$$

The last term of indicator function is squared return standardized by the denominator which is equal to the sum of integrated variance, given by integral $\int \sigma_t^2 dt$, and the variance of the noise $2\sigma_{\epsilon}^2$. Noise variance can simply be estimated as follows

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2J} \left(\sum_{j=1}^J r_{t,j}^2 - TTSE_t^{\Delta,adj} \right) \quad (10)$$

As integrated variance is still unknown approximation of medRV is used as initial or starting value of integrated variance, proposed by Anderesn et al. (2012)

$$\text{medRV}_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{J}{J-2} \right) \sum_{j=2}^{J-1} \text{med}(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|)^2 \quad (11)$$

If the fast scale frequency J is set in advance, according to the best available equidistant observations, the major problem that remains is to determine an optimal slow scale frequency. In another words, as J is given we should find an optimal \bar{n}^* and thus an optimal number of subsamples S^* and consequently an optimal slow scale frequency or sparse frequency Δ^* . According to Arnerić et al. (2019) the 1 minute fast time scale is the highest sampling frequency at which we can eliminate zero prices and transaction gaps within equidistant and non-empty time intervals. The optimal slow time scale can be found by minimizing the mean squared error (MSE) of the average sparse RV. As we know that MSE of an estimator is the sum of the squared bias and it's own variance we can use this criteria to balance between the bias and the variance, suggested by Zhang et al. (2005). MSE of the average sparse RV as a function of \bar{n} can be written as

$$\text{MSE}(RV_t^{\Delta,S}) = 4\bar{n}^2 \left(\frac{RV_t^\Delta}{2J} \right)^2 + \frac{4}{3} \left(\frac{J}{3} \frac{RQ_t^\Delta}{\bar{n}} \right) \quad (12)$$

where RQ_t^Δ is realized quarticity multiplied by $J/3$ to obtain consistent estimator of integrated quarticity

$$IQ_t = \int_{t-1}^t \sigma_t^4 dt \quad (13)$$

$$RQ_t^\Delta = \sum_{j=1}^J r_{t,j}^4 \quad (14)$$

By minimizing $\text{MSE}(RV_t^{\Delta,S})$ with respect to \bar{n} an optimal slow time scale frequency can be found. For example, if $J = 420$ intraday price observations within $\Delta = 1$ minute fast time scale sampling frequency and 1 minute increment and if $\bar{n}^* = 60$ then the optimal number of subsamples $S^* = 6,9016$ which means that optimal slow time sampling frequency or sparse frequency is approximately $\Delta^* = 7$ minutes.

4. Empirical data and analysis

This paper utilizes historical prices of the main stock market indices of four European emerging countries with EU membership at the time of writing, i.e. Romania, Bulgaria, Hungary and Croatia. All the high frequency prices are denoted in local currency. Data were provided by Thomson Reuters Thick History from January 4, 2010 to April 28, 2017. The total number of price observations differs per stock exchange due to difference in official trading hours, national holidays and overall trading activities. Prior to the analysis data are cleaned and properly filtered. Namely, preparation of the data and the analysis itself can be summarized in the following steps:

1. Removing observations from non-official trading hours and deleting zero prices
2. Selecting the highest possible sampling frequency (fast time scale) to obtain regularly spaced intervals which are non-overlapping and non-empty
3. Finding the lowest sampling frequency (slow time scale) which balances between the bias and the variance by minimizing MSE of average sparse RV
4. Computing daily series of TTSE for each stock market individually
5. Providing daily series of jump robust TTSE for each stock market individually

Steps 1 and 2 serve for data processing and creating relevant data set. The type of errors that may contaminate this type of data are mostly driven by missing intraday prices, e.g. due to lack of trading activities in considered markets, and price observations outside of official trading hours. Using ultra-high sampling frequencies in these markets like for example tick-by-tick prices or even sampled by 1 second would result in predominantly empty fields for most of the indices. Step 1 is easy to compute while step 2 depends on the trading activity of each stock exchange under consideration. Total number of transactions is recorded within trading hours every day from January 4, 2010 to April 28, 2017 and average number of transactions per 1 minute is obtained for each market (Table 2). This information was helpful to calculate average time between transactions that varies from 36 second to 77 seconds.

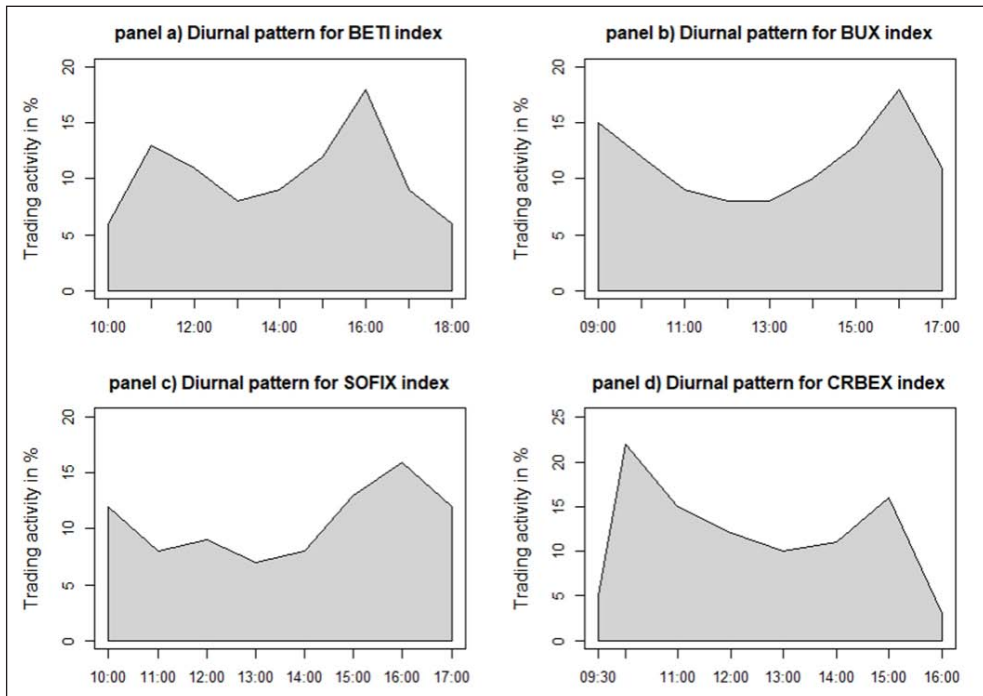
Table 2: Description of intraday observations across emerging stock markets

Stock Exchange	Index Ticker	Local Trading Hours	Average number of transactions per minute	Average time between transactions	Intraday observations after cleaning from January 4, 2010 to April 28, 2017
Bucharest	BETI	10:00 – 18:00	1.65	36 sec	383 469
Budapest	BUX	09:00 – 17:00	1.60	38 sec	737 263
Sofia	SOFIX	10:00 – 17:00	1.02	59 sec	650 932
Zagreb	CRBEX	09:30 – 16:00	0.78	77 sec	339 753

Source: Author's calculation according to data obtained from Thomson Reuters Tick History

It follows that 1 minute is the highest sampling frequency (fast time scale) at which we can eliminate transaction gaps within equidistant and non-overlapping time intervals. This finding is similar to Arnerić et al. (2019). Intraday diurnal pattern has been also found for each market individually, explaining periodicity in trading activities (Figure 1).

Figure 1: Intraday periodicity of trading activity across emerging stock markets



Source: Author's calculation according to data obtained from Thomson Reuters Tick History

Trading activities are more dense in the beginning and closing hours than in the lunchtime. According to the Figure 1 all market indices show a “U” shaped pattern in the middle of the day. The estimates of integrated variance in scope of this research are not considered to be sensitive to intraday diurnal patterns as according to Liu and Maheu (2012) intraday diurnal patterns have a repeating character, but do not tend to manifest in a daily seasonal affect.

Nevertheless, it can be noticed that both CRBEX and BETI start with lower recorded trading activities, while BUX and SOFIX show much higher trading activities very close to the end of trading hours. There is a minor exception of the SOFIX index, which shows “peak” around 12:00 indicating a slight increase in trading activities during lunchtime.

For the reasons outlined we suggest avoiding ultra-high frequencies data, i.e. tick-by-tick prices, when dealing with emerging markets and propose a fairly simple filtering technique to ensure that only non-zero prices within official trading hours are included. Table 2 also presents the number of intraday observations, sampled every 1 minute after cleaning process. Intraday returns sampled every 1 minute are inputs for step 3, 4 and 5. Step 3 is most relevant to filter out the microstructure noise from intraday returns and it is essential to continue with the step 4 and 5. In most cases the realized variance on a fast time scale shows similarities over time across stock markets, even it is contaminated with microstructure noise and thus biased (Figures 2-5 in the Appendix). Moreover, the second half of the observed period 2014-2017 shows to be rather less volatile compared to the first half denoting 2010-2013. A next step (step 3) is to find an optimal sparse sampling frequency (slow time scale) which balances between the bias and the variance by minimizing MSE of average sparse RV (Table 3).

Table 3: MSE of average sparse RV over slow time scale from 2 minute to 30 minutes

Slow time scale	BETI	BUX	SOFIX	CRBEX
2 min	0,00441	0,00573	0,00374	0,00078
3 min	0,00271	0,00350	0,00247	0,00051
4 min	0,00234	0,00190	0,00210	0,00044
5 min	0,00198	0,00143	0,00212	0,00041
6 min	0,00184	0,00119	0,00199	0,00040
7 min	0,00162	0,00215	0,00198	0,00042
8 min	0,00194	0,00151	0,00211	0,00039
9 min	0,00192	0,00263	0,00243	0,00044
10 min	0,00191	0,00114	0,00252	0,00043
11 min	0,00193	0,00185	0,00245	0,00052
12 min	0,00217	0,00133	0,00237	0,00050
13 min	0,00254	0,00128	0,00269	0,00054
14 min	0,00216	0,00153	0,00281	0,00072
15 min	0,00260	0,00131	0,00253	0,00053
20 min	0,00324	0,00176	0,00245	0,00074
25 min	0,00410	0,00254	0,00529	0,00074
30 min	0,00483	0,00263	0,00301	0,00081

Source: Author's calculation (as MSE's are very small numbers they are expressed in 10^{-6})

In the existing literature there is no consensus reached on the optimal sampling frequency for realized variance on a two-time scale. A conclusion that can be drawn from different studies is that the optimal sampling frequency depends on the type of security and its intraday trading activity. The highest possible frequency (fast time scale) can be limited regarding intraday trading activity, as in this paper. Fast time scale is limited at 1 minute because the trading activity becomes substantial at the 1 minute interval, while for example, very poor intraday trading activities at the 1 second interval are observed. Still, a slow time scale sampling frequency should be determined. According to highlighted results in Table 3, an optimal slow time scale frequency is 7 minutes for BETI index, 10 minutes for BUX index, 7 minutes for SOFIX index and finally 8 minutes for CRBEX index. According to selected slow time scale frequencies, two-time scale estimates of integrated variance are presented on Figures 2-5 in the Appendix (panel c). The last step is providing daily series of jump robust TTSE for each stock market individually which are presented on Figures 2-5 in the Appendix (panel d).

This research utilizes two-time scale estimator TTSE, as well as jump robust version of TTSE, on four different stock markets from emerging European economies: Romania, Hungary, Bulgaria and Croatia respectively. Reason for using this type of estimator is the ability to keep all intraday returns and still having unbiased and consistent estimator of integrated variance and robust to price jumps. The optimal slow frequency is determined for each considered market by minimizing the mean squared error (MSE) of the average sparse RV estimator (Table 3). This criteria balance between the bias and the variance of the proposed estimator. Empirical findings indicate that intraday returns should be sampled every 7 to 10 minutes at slow time scale. Otherwise, realized variance remains biased. The difference between RV at the highest possible frequency, limited at 1 minute, and TTSE can be attributed to the microstructure noise. Fortunately, this noise can be filtered out using the TTSE estimator. The RV bias can be verified and tested by the Mincer-Zarnowitz regression

$$TTSE_t = \alpha + \beta RV_t + u_t \quad (15)$$

where α and β are the parameters that should be estimated using ordinary least squares (OLS) and u_t are independently and identically distributed error terms. Mincer-Zarnowitz regression is a popular method for out-of-sample performance evaluation of competing variance models against variance benchmark. However, in this paper the Mincer-Zarnowitz regression is used to determine the significance of the bias, testing the null hypothesis $H_0: \alpha = 0, \beta = 1$. If the null hypothesis is rejected then a significant bias exists in RV due to microstructure noise. Both parameters can be tested jointly using Wald test, i.e. multiple restrictions test. A Wald statistic has an asymptotic chi-square distribution with degrees of freedom equal to the number of restriction; in this case 2 restrictions are imposed. Moreover, a negative and significant intercept α means that RV on average overestimates TTSE. The significant RV bias is found for all emerging stock markets according to the Wald statistic (Table 4). Additionally, it is not negligible that all intercepts are negative.

Table 4: Mincer-Zarnowitz regression with Wald test in comparison of TTSE against RV

Index Ticker	BETI	BUX	SOFIX	CRBEX
Dependent variable	TTSE 7 min	TTSE 10 min	TTSE 7 min	TTSE 8 min
Constant	-0.00001*** (0.000001)	-0.00005*** (0.000002)	-0.00001*** (0.000001)	-0.00004*** (0.000001)
RV (1 min)	0.88699*** (0.01139)	1.16493*** (0.01100)	1.15878*** (0.01094)	1.04517*** (0.01159)
Observations	1847	1825	1811	1828
R2	0.76675	0.86028	0.86105	0.81657
Adjusted R2	0.76663	0.86021	0.86097	0.81646
Residual Std. Error	0.00004	0.00004	0.00003	0.00002
Wald Statistic	462.660***	1047.20***	231.450***	59.101***
Underestimated %	14.19%	10.08%	51.89%	30.71%
Overestimated %	85.81%	89.92%	48.11%	69.29%

Source: Author's calculation (standard errors are in the brackets and significance levels are denoted as * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$)

By counting the points below or above the Mincer-Zarnowitz regression we can obtain the percentage of days when variance was overestimated or underestimated. Most of the time integrated variance was overestimated, on average between 70% and 90%, for all stock markets except Bulgaria (Table 4).

Moreover, the same Mincer-Zarnowitz regression is applied to verify contribution of the jumps when comparing TTSE against jump robust TTSE, i.e. JRTTSE

$$JRTTSE_t = \alpha + \beta TTSE_t + u_t \quad (16)$$

If the null hypothesis $H_0: \alpha = 0, \beta = 1$ is not rejected one can conclude that TTSE is already robust to the price jumps as TTSE is equal to the JRTTSE. Results of this testing procedure are given in Table 5.

Table 5: Mincer-Zarnowitz regression with Wald test in comparison of JRTTSE against TTSE

Index Ticker	BETI	BUX	SOFIX	CRBEX
Dependent variable	JRTTSE 7 min	JRTTSE 10 min	JRTTSE 7 min	JRTTSE 8 min
Constant	0.0000059*** (0.00000069)	0.00001*** (0.000001)	0.00001*** (0.000001)	0.000002*** (0.0000004)
TTSE	0.996870 (0.80196030)	0.96332* (0.527268)	0.28202*** (0.00967)	0.46716 (0.00893)
Observations	1847	1825	1811	1828
R2	0.97114	0.95228	0.31983	0.59966
Adjusted R2	0.97106	0.85220	0.31945	0.59944
Residual Std. Error	0.00002	0.00003	0.00004	0.00003
Wald Statistic	3.98751	5.0362*	850.612***	735.090***

Source: Author's calculation (standard errors are in the brackets and significance levels are denoted as * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$)

According to results in Table 5, the null hypothesis of the Wald statistic is not rejected for two of four analyzed emerging stock markets at the significance level of 5%. Thus, we can conclude that TTSE estimator is already robust to price jumps when comparing against JRTTSE for Romania and Hungary. For Bulgaria and Croatia TTSE and JRTTSE differs significantly which indicates a significant price jumps in these emerging markets. This findings confirm that TTSE estimator is robust to microstructure noise as well as price jumps not for all countries and for that same reason JRTTSE should be applied at the slow time scale frequency obtained by minimizing the mean square error of TTSE, while the fast time scale is held fix at the highest possible frequency.

Another aspect to consider is the extent to which the optimal slow sampling frequency is time invariant, i.e. the optimal sampling frequency might change over time due to variations of the underlying security. Therefore, entire observation period is divided into two parts; high variance sub-period (04.01.2010 - 31.12.2013) and low variance sub-period (02.01.2014 - 28.04.2017). For each sub-period an optimal slow time scale frequency is obtained by minimizing MSE of average sparse RV (Table 6).

Table 6: Determination of optimal slow time scale frequency in two sub-periods

Observation sub-period	BETI	BUX	SOFIX	CRBEX
04.01.2010 - 31.12.2013	7 min	8 min	6 min	8 min
02.01.2014 - 28.04.2017	11 min	10 min	7 min	9 min

Source: Author's calculation according to data obtained from Thomson Reuters Tick History

From Table 6 it is evident that in a highly volatile sub-period one should sampled intraday returns more frequently, i.e. smaller sampling interval is optimal at the slow time scale. Contrary to that one should sampled intraday returns less frequently in a less volatile sub-period. Nevertheless, sampling intervals do not differ more than two minutes except for Romania (BETI index). Stability of the sampling intervals over time is important for robustness checking of slow time scale frequency, which is confirmed.

5. Results and discussion

To the best of the authors' knowledge, previous studies have not considered finding appropriate unbiased estimator of the unknown integrated variance in Romania, Hungary, Bulgaria and Croatia. There is a single attempt to use the two-time scale estimator and to search for optimal slow time scale in European emerging markets (Arnerić et al. 2019), but one would expect that optimal slow time scale frequency varies not only across markets but also varies over time. The main result is that intraday prices should be sampled from 7 and 10 minutes at slow time scale while the fast time scale should be fixed at 1 minute to keep as much information as possible when obtaining two-time scale estimates of integrated variance. This leads us to conclusion that slow time scale frequency shouldn't be chosen arbitrarily or too sparsely. This conclusion is opposite to developed markets findings of 5 minute sampling (Andersen et al., 2001; Hansen and Lunde, 2006; Bundi and Russell, 2008). Moreover, in more volatile period slow time scale varies from 7 to 11 minutes, i.e. slightly lower compared to less volatile period. The optimal slow frequency is determined by minimizing the mean squared error (MSE) of the average sparse RV estimator as it balances between the bias and efficiency. However, the fast time scale frequency less than 1 minute is not appropriate due to not so frequent trading and lower liquidity in considered emerging markets. Additionally, the present paper verifies robustness of proposed two-time scale estimator to price jumps considering truncation. This study also enriches the literature in framework of high frequency data cleaning and filtering. Most of the time integrated variance was overestimated on fast time scale, on average between 70% and 90%, for all stock markets except Bulgaria. We also found that TTSE estimator is already robust to price jumps when comparing against JRTTSE for Romania and Hungary. For Bulgaria and Croatia TTSE and JRTTSE differs significantly which indicates a significant price jumps in these emerging markets.

These findings offer valuable information to market participants, as they are able to apply the most accurate integrated variance estimates. Monitoring and forecasting of these estimates can be helpful in designing trading strategies in real time. Moreover, the optimal sampling frequency provides information how frequently stocks are traded and therefore it can be used as a liquidity indicator. This kind of

information is valuable not only to portfolio managers, but also public authorities to understand market trading on high frequency level.

6. Concluding remarks

Extreme events on stock markets, for example 2010 Flash Crash, impose the need for deeper understanding of stock price variations in real time. For the same necessity a great attention has been given to high frequency returns observed on very small time intervals within a single day. Despite the importance of intraday returns, not only in estimation of integrated variance, but also in determining market liquidity and designing trading strategies, they introduce new challenges to practitioners, especially in emerging stock markets. These intraday returns are observed only at discrete time points disabling any variance estimator to be unbiased. Discretization error is not the only source of the bias. All other sources, introduced in the first two sections of the paper, are collectively summarized as a microstructure noise.

When a microstructure noise is present realized variance is biased measure of the true integrated variance, which is not computable in practical applications. The significance of the bias is confirmed and tested using Mincer-Zarnowitz regression. Therefore, it was necessary to find unbiased and consistent estimator of integrated variance. More specifically, we suggested using two-time scale estimator with optimal slow time scale frequency between 7 and 10 minutes. Selection of optimal slow time scale frequency wasn't arbitrary or straightforward, especially when dealing with emerging markets under consideration: Romania, Hungary, Bulgaria and Croatia. Trading in these markets is not so frequent and thus pre estimation process of cleaning the data is also proposed. We strongly suggest not to use tick-by-tick data and then search for optimal fast and slow time scale frequencies, but to determine fast time scale frequency in advance, depending on trading activity. The fast time scale is limited at 1 minute which enables to use as much information as possible within regularly spaced, non-overlapping and non-empty intervals included in official trading hours. Unlike the fast time scale, optimal slow time scale is found by minimizing the mean squared error (MSE) of the average sparse RV.

It is additionally found that intraday returns should be sampled frequently in more volatile period. The time invariance of slow time scale also has been proven. Moreover, the robustness of the results with respect to the price jumps has been verified for Romania and Hungary, unlike Croatia and Bulgaria, for which we recommend using a robust version of two-time scale estimation of integrated variance within truncation technique.

The limitation of the present research refers to only four emerging markets, although chosen by the principle of developing neighboring countries for which

high frequency data were available and provided from Thomson Reuters Tick History. Direction for the further research is to examine if multiple time scales reduce price jumps and microstructure noise better than two-time scales.

References

- Aït-Sahalia, Y., Jacod, J. (2009) “Testing for jumps in a discretely observed process”, *The Annals of Statistics*, Vol. 37, No. 1, pp. 184–222, <http://doi.org/10.1214/07-aos568>.
- Aït-Sahalia, Y., Mykland, P., Zhang, L. (2005) “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise”, *Review of Financial Studies*, Vol. 18, No. 2, pp. 351–416, <http://doi.org/10.1093/rfs/hhi016>.
- Aït-Sahalia, Y., Mykland, P., Zhang, L. (2011) “Ultra high frequency volatility estimators with dependent microstructure noise”, *Journal of Econometrics*, Vol. 160, No. 1, pp. 160–175, <http://doi.org/10.1016/j.jeconom.2010.03.028>.
- Andersen, T. G., Bollerslev, T. (1998) “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts”, *International Economics Review*, Vol. 39, No. 4, pp. 885–905, <http://doi.org/10.2307/2527343>.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001) “The Distribution of Realized Exchange Rate Volatility”, *Journal of the American Statistical Association*, Vol. 96, No. 453, pp. 42–55, <http://doi.org/10.1198/016214501750332965>.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003) “Modeling and Forecasting Realized Volatility”, *Econometrica*, Vol. 71, No. 2, pp. 579–625, <http://doi.org/10.1111/1468-0262.00418>.
- Andersen, T. G., Dobrev, D., Schaumburg, E. (2012) “Jump-robust volatility estimation using nearest neighbor truncation”, *Journal of Econometrics*, Vol. 169, No. 1, pp. 75–93, doi: 10.1016/j.jeconom.2012.01.011.
- Arnerić, J., Matković, M., Sorić, P. (2019) “Comparison of range-based volatility estimators against integrated volatility in European emerging markets”, *Finance Research Letters*, Vol. 28, pp. 118–124, <http://doi.org/10.1016/j.frl.2018.04.013>.
- Arnerić, J., Poklepović, T., Teai, J. W. (2018) “Neural Network Approach in Forecasting Realized Variance Using High-Frequency Data”, *Business System Research*, Vol. 9, No. 2, pp. 18–34, <http://doi.org/10.2478/bsrj-2018-0016>.
- Bandi, F. M., Russell, J. R. (2008) “Microstructure Noise, Realized Volatility, and Optimal Sampling”, *Review of Economic Studies*, Vol. 75, No. 2, pp. 339–369, <http://doi.org/10.1111/j.1467-937x.2008.00474.x>.
- Barndorff-Nielsen, O. E., Shephard, N. (2002) “Estimating quadratic variation using realized volatility”, *Journal of Applied Econometrics*, Vol. 17, No. 5, pp. 457–477, <http://doi.org/10.1002/jae.691>.

- Barndorff-Nielsen, O. E., Shephard, N. (2006) “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation”, *Journal of Financial Econometrics*, Vol. 4, No. 1, pp. 1–30, <http://doi.org/10.1093/jjfinc/nbi022>.
- Bianco, S., Corsi, F., Reno, R. (2009) “Intraday LeBaron effects”, *Proceedings of the National Academy of Sciences*, Vol. 106, No. 28, pp. 11439–11443, <http://doi.org/10.1073/pnas.0901165106>.
- Boudt, K., Cornelissen, J., Croux, C., Laurent, S. (2012) “Non-parametric tests for intraday jumps: Impact of periodicity and microstructure noise”. In Bauwens, L., Hafner, C., and Laurent, S. (Eds.) *Handbook of Volatility Models and Their Applications*, New Jersey: John Wiley & Sons, <http://doi.org/10.1002/9781118272039.ch18>.
- Boudt, K., Zhang, J. (2013) “Jump robust two time scale covariance estimation and realized volatility budgets”, *Quantitative Finance*, Vol. 15, No. 6, pp. 1041–1054, <http://doi.org/10.1080/14697688.2012.741692>.
- Brownlees, C. T., Gallo, G. M. (2006) “Financial econometric analysis at ultra-high frequency: Data handling concerns”, *Computational Statistics & Data Analysis*, Vol. 51, No. 4, pp. 2232–2245, <http://doi.org/10.1016/j.csda.2006.09.030>.
- Corsi, F., Pirino, D., Reno, R. (2010) “Threshold bipower variation and the impact of jumps on volatility forecasting”, *Journal of Econometrics*, Vol. 159, No. 2, pp. 276–288, <http://doi.org/10.1016/j.jeconom.2010.07.008>.
- Hanousek, J., Kočenda, E., Novotný, J. (2012). The identification of price jumps. *Monte Carlo Methods and Applications*, Vol. 18, No. 1, pp. 53–77, <http://doi.org/10.1515/mcma-2011-0019>.
- Hanousek, J., Kočenda, E., Novotný, J. (2013). Intraday price behavior during information arrival in emerging markets. In H. K. Baker, & H. Kiyamaz (Eds.), *Market microstructure in emerging and developed markets*, pp. 445–462, John Wiley & Sons: New York, <http://doi.org/10.1002/9781118681145.ch24>.
- Hansen, P. R., Lunde, A. (2005) “A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data”, *Journal of Financial Econometrics*, Vol. 3, No. 4, pp. 525–554, <http://doi.org/10.1093/jjfinc/nbi028>.
- Huang, X., Tauchen, G. (2005), The relative contribution of jumps to total price variance, *Journal of Financial Econometrics*, Vol. 3, No. 4, pp. 456–499, <http://doi.org/10.1093/jjfinc/nbi025>.
- LeBaron, B. (1992) “Some Relations Between Volatility and Serial Correlations in Stock Market Returns”, *The Journal of Business*, Vol. 65, No. 2, pp. 199–219, <http://doi.org/10.1086/296565>.
- Lee, S., Mykland, P. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics, *Review of Financial Studies*, Vol. 21, No. 6, pp. 2535–2563, <http://doi.org/10.1093/rfs/hhm056>.

- Liu, C., Maheu, J. M. (2012) “Intraday dynamics of volatility and duration: Evidence from Chinese stocks”, *Pacific-Basin Finance Journal*, Vol. 20, No. 3, pp. 329–348, <http://doi.org/10.1016/j.pacfin.2011.11.001>.
- McAleer, M., Medeiros, M. C. (2008) “Realized Volatility: A Review”, *Econometric Reviews*, Vol. 27, No. 1-3, pp. 10–45, <http://doi.org/10.1080/07474930701853509>.
- Oomen, R. C. A. (2005) “Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes”, *Journal of Financial Econometrics*, Vol. 3, No. 4, pp. 555–577, <http://doi.org/10.1093/jjfinec/nbi027>.
- Roll, R. (1984) “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market”, *The Journal of Finance*, Vol. 39, No. 4, pp. 1127–1139, <http://doi.org/10.1111/j.1540-6261.1984.tb03897.x>.
- Zhang, L., Mykland, P., Aït-Sahalia, Y. (2005) “A Tale of Two Time Scales”, *Journal of American Statistical Association*, Vol. 100, No. 472, pp. 1394–1411, <http://doi.org/10.1198/016214505000000169>.
- Zhang, L. (2011) “Estimating covariation: Epps Effect, microstructure noise”, *Journal of Econometrics*, Vol. 160, No. 1, pp. 33–47, <http://doi.org/10.1016/j.jeconom.2010.03.012>.

Izazovi u procjeni integrirane varijance dioničkih tržišta u nastajanju¹

Josip Arnerić², Mario Matković³

Sažetak

Procjena integrirane varijance pomoću visoko frekventnih podataka zahtijeva iskustvo modeliranja i vještine sažimanja podataka. Iako su intra-dnevni prinosi posljednjih godina privukli veliku pozornost, upravljanje tim podacima izazovno je zbog svojih jedinstvenih karakteristika. Kada se radi o ultra visoko frekventnim opažanjima, tj. tick-by-tick podacima, potrebno ih je prije same procjene obraditi iz dva razloga: uklanjanje mikrostrukturnog šuma i pronalaženje odgovarajućeg nepristranog procjenitelja integrirane varijance. Zbog ne čestog trgovanja i smanjene likvidnosti tržišta u nastajanju, u ovom se radu razmatraju problemi vezani uz kvalitetu visoko frekventnih podataka. Dva su doprinosa postojećoj literaturi. Prvo, detaljno se raspravlja o problemima s transakcijskim podacima te se predlaže kako postupati s tim podacima. Drugo, pronalazi se optimalna frekvencija uzorkovanja na sporij vremenskoj skali, koja se treba koristiti za dobivanje dvostruko vremenski skalirane procjene integrirane varijance svakog od tržišta u nastajanju koje se razmatra: Rumunjska, Mađarska, Bugarska i Hrvatska. Empirijski rezultati pokazuju da se intra-dnevni podaci trebaju uzorkovati svakih 7 do 10 minuta na sporij vremenskoj skali, dok je brza vremenska skala fiksirana na najvišu moguću i prikladnu frekvenciju. Realizirana volatilnost na brzom vremenskoj skali obično precjenjuje integriranu varijancu na svim dioničkim tržištima osim Bugarske; u prosjeku između 70% i 90% vremena. Štoviše, robusnost rezultata s obzirom na prisutnost cjenovnih skokova, potvrđena je za Rumunjsku i Mađarsku, dok se za Hrvatsku i Bugarsku preporuča koristiti robusnu verziju dvostruko vremenski skalirane procjene integrirane varijance, pomoću tehnike „sakačenja“. Dodatno je utvrđeno da se intra-dnevni prinosi trebaju češće uzorkovati u vrlo volatilnim periodima. Ovi nalazi nude vrijedne informacije sudionicima na tržištu, jer mogu primijeniti najtočnije procjene ex-post mjere volatilnosti kao nepristrane i konzistentne procjene integrirane varijance.

Ključne riječi: integrirana varijanca, optimalna frekvencija uzorkovanja, mikrostrukturni šum, skokovi, procjenitelj dvostruke vremenske skale, dioničko tržište u nastajanju

JEL klasifikacija: G1, C13, C14, C58

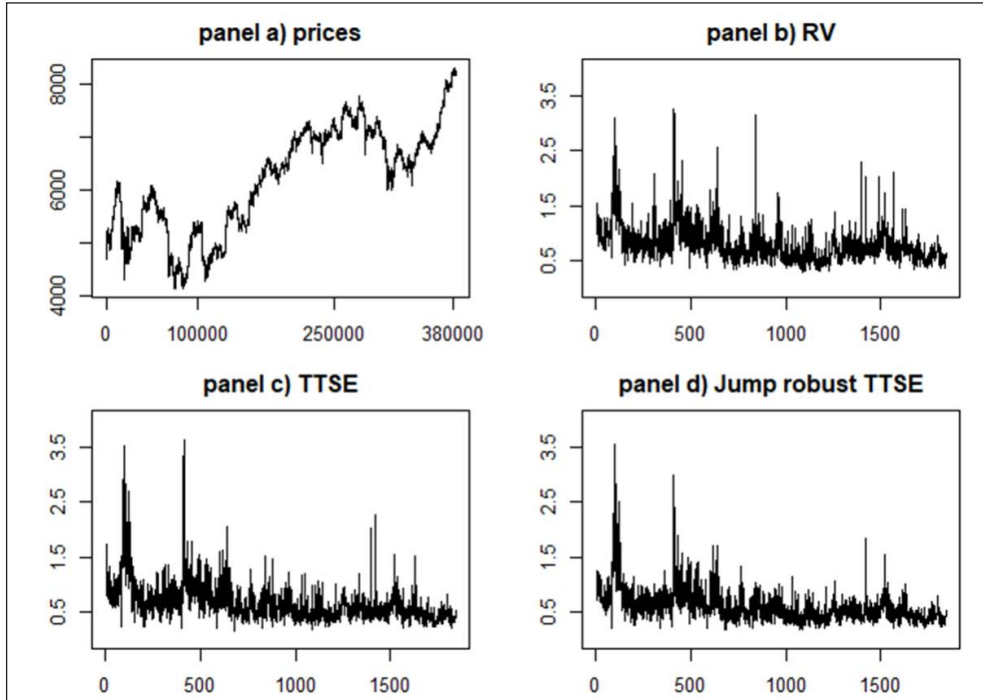
¹ Izricanje odgovornosti: ovdje prikazani stavovi su autorovi i ne odražavaju stavove NN Group & Nationale-Nederlanden.

² Izvanredni profesor na Katedri za statistiku Sveučilišta u Zagrebu, Ekonomski fakultet, Trg. J. F. Kennedyja 6, 10000 Zagreb, Hrvatska. Znanstveni interes: financijska ekonometrija, modeli i metode financijske vremenske serije, volatilnost i kointegracija, modeli GARCH i MGARCH, visokofrekventni podaci i stohastički procesi. Tel.: +385 1 238 3361. E-mail: jarneric@efzg.hr. ORCID: <http://orcid.org/0000-0002-2901-2609> (corresponding author).

³ Doktor ekonomskih znanosti, menadžer financijskog rizika na Odjelu za modele rizika, NN Group & Nationale-Nederlanden, Prinses Beatrixlaan 35, 2595 AK Den Haag, Nizozemska. Znanstveni interes: financijska ekonometrija, ostvarena volatilnost i OHLC procjenitelji, upravljanje rizikom. E-mail: mmatkic@hotmail.com.

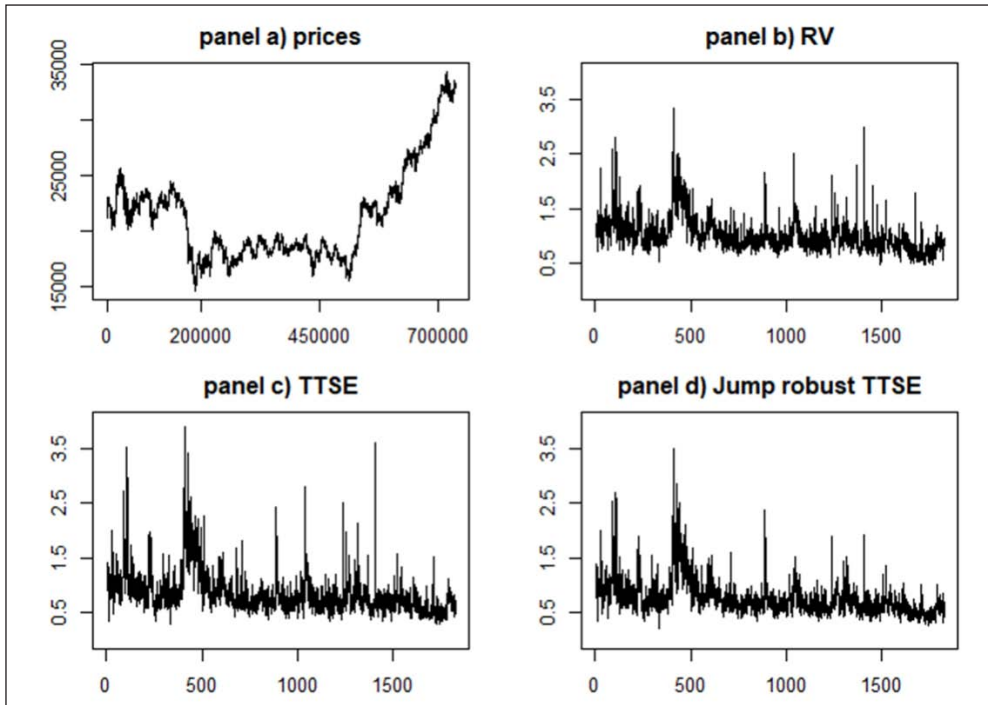
Appendices

Figure 2: Intraday prices of the BETI index after cleaning (panel a), realized variance on a fast time scale (panel b), two-time scale estimates (panel c) and jump robust two-time scale estimates (panel d)



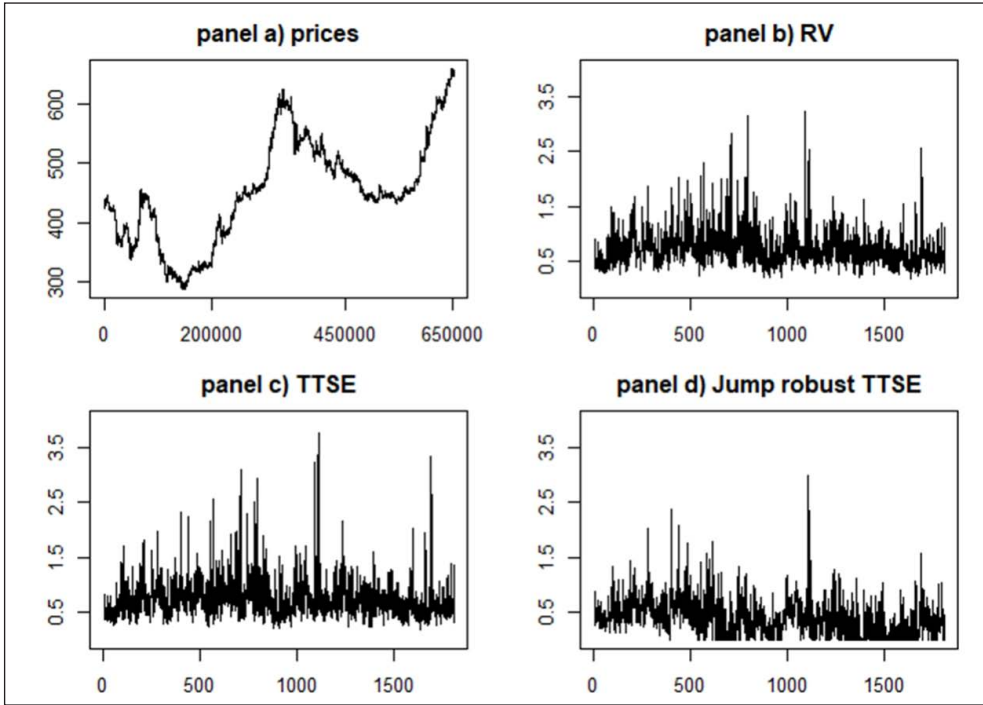
Source: Author's calculation according to data obtained from Thomson Reuters Tick History

Figure 3: Intraday prices of the BUX index after cleaning (panel a), realized variance on a fast time scale (panel b), two-time scale estimates (panel c) and jump robust two-time scale estimates (panel d)



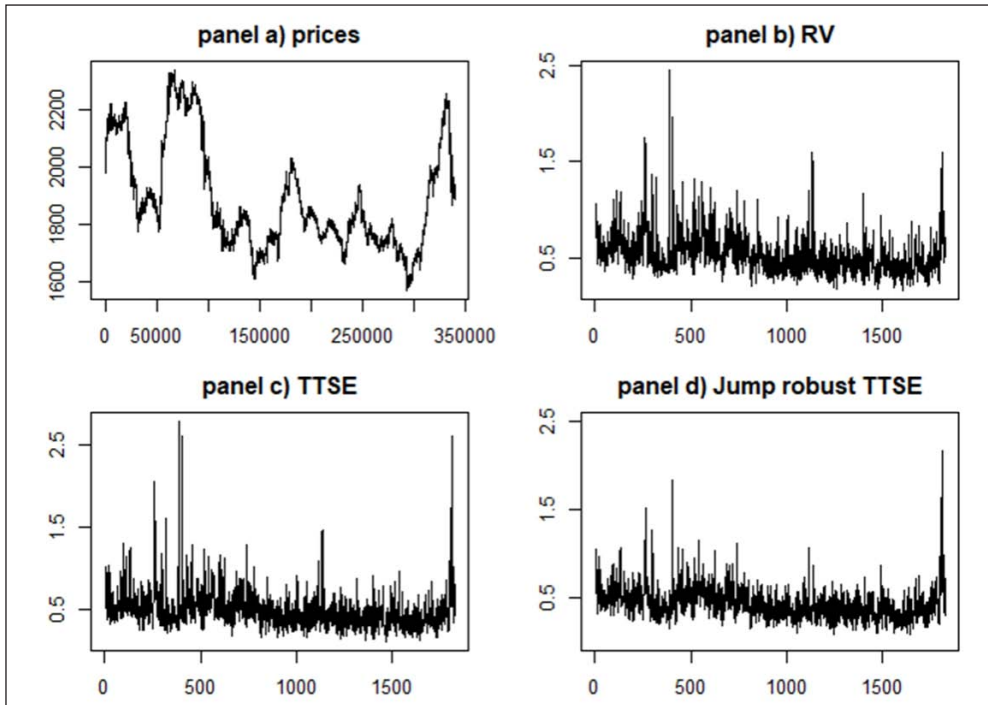
Source: Author's calculation according to data obtained from Thomson Reuters Tick History

Figure 4: Intraday prices of the SOFIX index after cleaning (panel a), realized variance on a fast time scale (panel b), two-time scale estimates (panel c) and jump robust two-time scale estimates (panel d)



Source: Author's calculation according to data obtained from Thomson Reuters Tick History

Figure 5: Intraday prices of the CRBEX index after cleaning (panel a), realized variance on a fast time scale (panel b), two-time scale estimates (panel c) and jump robust two-time scale estimates (panel d)



Source: Author's calculation according to data obtained from Thomson Reuters Tick History