Is Jump Robust Two Times Scaled Estimator Superior among Realized Volatility Competitors?

Maria Čuljak [1,†,‡], Josip Arnerić [2,*,†] and Ante Žigman [1,†,‡]

1   Croatian Financial Services Supervisory Agency, Franje Račkoga 6, 10 000 Zagreb, Croatia;
    maria.culjak@hanfa.hr (M.Č.); ante.zigman@hanfa.hr (A.Ž.)
2   Faculty of Economics and Business, University of Zagreb, Trg J. F. Kennedyja 6, 10 000 Zagreb, Croatia
*   Correspondence: jarneric@efzg.hr; Tel.: +385-1-238-3361
†   These authors contributed equally to this work.
‡   The content of this research does not reflect the official opinion of the Croatian Financial Services
    Supervisory Agency.

Abstract: This paper compares the empirical performance of the realized volatility estimators on an extensive high-frequency dataset of stock indices from four developed European markets with thick trading and intensive intraday activity. Even though the proposed estimators have distinctive properties, it is not clear which one has a better performance in terms of unbiasedness and consistency. Some of them are robust to microstructure noise only, and others are robust solely to price jumps, whereas a few of them are robust to both. Therefore, the main purpose is finding a benchmark estimator among alternative competitors, as the best proxy of integrated variance, and empirical demonstration of its superiority. The vast majority of the existing studies largely rely on developed US data or simulation data, but inferences obtained on such data might deviate from European developed markets. This study aims to fill in that niche. In particular, the optimal sampling frequency of proposed benchmark estimator is determined with respect to the trade-off between its bias and the variance of each stock index individually. Afterwards, probability integral transformation, Mincer–Zarnowitz regression and upper tail correlation from appropriate copula function are considered as an adequate pairwise comparison methods. Notable contributions of this paper include unambiguously proven superiority of robust two times scaled estimator for selected European developed markets within the range of optimal slow time frequency from 10 to 30 s. Finally, recommendations for research and practitioners regarding the usage of jump robust two times scaled estimator are given. In fact, asset managers, institutional investors as well as market regulators could benefit from proposed realized volatility benchmark in making long-term investment decisions, leading to sustainable finance.

Keywords: high-frequency observations; realized volatility benchmark; microstructure noise; price jumps; sampling frequency selection

MSC: 62P05; 62P20; 91B02; 91B05; 91B84

1. Introduction

Over the past few decades, trading algorithms have been improved in line with technological advancements and increasing accessibility of high-frequency data, i.e., transaction data recorded in very small successive intervals of time. Usage of such data is extremely beneficial to practitioners in establishing trading strategies and consequently achieving higher possible returns. For the same reason, new statistical and econometric tools for high-frequency data analysis have been developed, taking into account a number of distinct characteristic of these data, such as discreteness, unevenly spaced time intervals, long memory, market microstructure noise and price jumps. These properties have been comprehensively studied in the literature for better understanding the mechanism of the price formation as well as the variance of price returns [1–3].

Estimation and forecasting of the variance of a financial time-series is still, and will remain for a long time, one of the major challenges for researchers and practitioners alike. Despite the enormous existing studies and various proposed approaches [4–6], most of them do not deal with empirical high-frequency or intraday data for estimation of a true but unknown variance of returns, so-called integrated variance (IV), but use simulation data or just ignore some of their distinguishing properties. However, integrated variance can be efficiently estimated by the realized variance (RV) as the sum of equidistant intraday squared returns [7]. It has been documented that realized variance converges in probability to integrated variance [8]. Moreover, RV is a perfect measure of IV when intraday prices are observed continuously [9]. On the contrary, when prices are exposed to a substantial amount of microstructure noise, the realized variance may not obtain consistency and asymptotic unbiasedness [10,11]. Common sources of market microstructure noise are discreteness of price observations, non-synchronous trading, bid–ask bounces, differences in trade sizes, etc. In real life, stock prices are only observed at discrete time points and high-frequency data are not available for all stocks and markets when trading is thin. RV is more contaminated with microstructure noise as sampling frequency increases, or in other words, when the time interval between successive observations declines [12]. On the contrary, when sampling at a lower frequency, RV becomes less biased due to reduction of the noise [13]. The widespread opinion is to sample sparsely from 5 min to 30 min, as suggested in the literature [14]. On the other side, sparsely sampling reduces a significant amount of information, leading to inefficient RV estimation. For example, one might sample every 10 min from transactions data observed every second, which means that every 600th observation is selected. In that context, it is crucial to find an optimal sampling frequency for which the balance between efficiency and potential biasedness of the RV estimator is achieved [14,15]. The trade-off between sampling too frequently and sampling too rarely can be formalized by minimizing the root mean squared error of the estimator. A less formal approach of finding appropriate sampling frequency, but very popular in empirical work, is volatility signature plot [16]. Nonetheless, the first consistent estimator of the IV proposed in the literature is a Two Times Scaled Realized Variance (TSRV). Estimator TSRV exhibits a nice asymptotic properties in the presence of the noise, while keeping all the data at the highest possible frequency. This is accomplished by taking a linear combination of the pre-averaged RV sampled at the two scales: fast time scale and low time scale [15,17]. However, when jumps occur, a two times scaled estimator becomes biased. Many jump robust estimators were established in the literature, but most of them are not robust to microstructure noise [18,19].

Additionally, a jump robust version of TSRV was proposed (RTSRV) [20]. Both estimators, TSRV and RTSRV, respectively, are constructed to deal with highly liquid assets, e.g., observations sampled every few seconds. Otherwise, if sampling frequency is low to begin with, e.g., every minute, the bias of the estimator might be over-corrected. This is the main reason why this paper considers high-frequency data obtained exclusively from developed European stock exchanges. Therefore, the key objective of the paper is to provide empirical evidence of RTSRV superiority among realized volatility competitors, as the best approximation of IV. For the same reason, the jump robust two times scaled estimator is used as a benchmark in this research. Although, this research does not cover the most recent period of the historical data, obtained findings contribute to academics as well as market participants by indicating an optimal slow time scale frequency that should be used when RTSRV is applied for each analyzed market individually. There is no consensus in the literature on which estimator is better for which market. This research provides an answer to that question. Along with comparison of RTSRV as a benchmark realized volatility against other RV estimators, the purpose is to conclude which markets are more contaminated with the microstructure noise, and in which markets price jumps are more present. Thus, concrete findings offer valuable suggestions to users when employing high-frequency data in financial management. Once again, it should be noted that performance of RV estimators which are robust to microstructure noise, price jumps or

both have been documented in previous studies based on simulation data or data from the US markets [21–23], while there is no research that analyzes the developed European stock markets, as in this paper. Less developed markets are not considered due to the poor quality of high-frequency data which cannot be observed every second, while observations sampled at lower frequency are not welcome as the bias of the benchmark estimator might be over corrected. Moreover, the novelty of the paper is that it empirically demonstrates RTSRV robustness to both microstructure noise and price jumps, i.e., for each market under consideration of an optimal slow (low) sampling, frequency is determined to mitigate the sensitivity to microstructure noise without losing a significant part of intraday observations, while a threshold parameter is attentively selected for truncation purposes to mitigate the price jumps. In total, seven competing estimators are considered for comparison against benchmark volatility proxy of the IV. Pairwise comparison methods incorporate the Kolmogorov–Smirnov test on probability integral transformations, Mincer–Zarnowitz regression, and upper tail correlation from the Gumbel copula.

The rest of the manuscript is outlined as follows: Section 2 presents the employed data and methods. Section 3 provides empirical findings. Section 4 offers discussion with respect to current results, while conclusions are provided in Section 5.

2. Data and Methods

This study utilizes in total more than 195 million observations, over a 7-year period. Data of developed stock markets from 4 January 2010 to 28 April 2017 for Germany, Italy, France and UK were provided by Thomson Reuters Tick History (TRTH) service. Indices MIB, DAX, CAC and FTSE represent benchmarks from Italian, German, French and British stock exchanges. Despite the fact that historical data do not cover the most recent period, due to limited funding from the project which is completed four years ago, an extensive dataset exhibit representativeness with respect to the long period of time and the most liquid European stock markets which is strongly required for computation of proposed benchmark estimator RTSRV.

In Table 1, the number of trading days and the number of 1-second observations are given. These numbers differ across observed European markets. The intraday data taken into consideration were during official trading hours from 9:00 a.m. until 5:30 p.m., five days a week.

Table 1. Summary of high-frequency data observed from 4 January 2010 to 28 April 2017.

European Market	Stock Index	Trading Days	One Second Observations
Italy	MIB	1862	42773317
Germany	DAX	1863	56650069
France	CAC	1878	46659800
UK	FTSE	1850	48920222

Eight estimators of integrated variance, including a benchmark RTSRV, are presented in Table 2. Each of them depends on sampling frequency Δ, i.e., the increment between two successive and equally spaced price observations $\Delta = t_i - t_{i-1}$. The number of intraday returns r_{t_i} for every trading day is n_t. Since $n_t = 1/\Delta$, the lesser information is used when Δ increases. Moreover, realized variance RV_t^Δ becomes unbiased when sampling frequency increases due to reduction of microstructure noise as a consequence of sparse sampling [16]. However, the cost of sparse sampling is a high variance of an estimator (inconsistency) as a small number of observations is left for computational reasons. Not only does the microstructure noise contaminate realized variance, but also price jumps [23–25]. Thus, a second estimator BPV_t^Δ is proposed in the literature as jumps robust estimator. The idea of [7] who introduced a bipower variation, is that multiplication of $|r_{t_i}|$ with the adjacent $|r_{t_{i-1}}|$ will dampen the jump impact if it occurs even for sufficiently small Δ. Similar to bipower variation, two estimators also emerged in the literature, i.e., minimized

and medianized realized variance $MinRV_t^\Delta$ and $MedRV_t^\Delta$, respectively [26]. According to [21], both estimators follow a concept of the nearest neighbor truncation by the use of the minimum operator on blocks of two returns or median operator on blocks of three returns. They have better performance in the finite samples compared to bipower variation although $MinRV_t^\Delta$ suffers from a similar exposure to zero returns as BPV_t^Δ.

Table 2. Competing realized estimators of integrated variance.

Estimator	Formulation
Realized variance	$RV_t^\Delta = \sum\limits_{i=1}^{n_t} r_{t_i}^2$
Bipower variation	$BPV_t^\Delta = \dfrac{\pi}{2} \sum\limits_{i=2}^{n_t} |r_{t_{i-1}}||r_{t_i}|$
Minimized RV	$MinRV_t^\Delta = \dfrac{\pi}{\pi-2} \dfrac{n_t}{n_t-1} \sum\limits_{i=2}^{n_t} min(|r_{t_{i-1}}|, |r_{t_i}|)^2$
Medianized RV	$MedRV_t^\Delta = \dfrac{\pi}{6-4\sqrt{3}+\pi} \dfrac{n_t}{n_t-2} \sum\limits_{i=2}^{n_t-1} med(|r_{t_{i-1}}|, |r_{t_i}|, |r_{t_{i+1}}|)^2$
Average subsampled RV	$ARV_t^{\Delta,k} = \dfrac{n_t}{n_t-k+1} \dfrac{1}{k} \sum\limits_{j=1}^{k} \sum\limits_{i=1}^{n_t} r_{t_{ij}}^2$
Two times scaled RV	$TSRV_t^{\Delta,k} = \dfrac{1}{1-\dfrac{n_t-k+1}{n_t k}} \left(\dfrac{1}{k} \sum\limits_{j=1}^{k} \sum\limits_{i=1}^{n_t} r_{t_{ij}}^2 - \dfrac{n_t-k+1}{n_t k} \sum\limits_{i=1}^{n_t} r_{t_i}^2 \right)$
Robust two times scaled RV	$RTSRV_t^{\Delta,k,\theta} = \dfrac{c_\theta}{1-\dfrac{n_t-k+1}{n_t k}} \left(\dfrac{1}{k} \sum\limits_{j=1}^{k} \sum\limits_{i=1}^{n_t} r_{t_{ij}}^2 I_i(\theta) - \dfrac{n_t-k+1}{n_t k} \sum\limits_{i=1}^{n_t} r_{t_i}^2 I_i(\theta) \right)$
Hayashi–Yoshida RV	$HYRV_t^{\Delta,\theta} = \sum\limits_{i=1}^{n_t} r_{t_i}^2 I_i(\theta)$

A serious drawback of the aforementioned estimators is the lack of data when sampling sparsely. Opposite to that, Ref. [27] proposed how to keep all the data but still have an unbiased and asymptotically consistent estimator of integrated variance. Thus, a two time scaled estimator $TSRV_t^{\Delta,k}$ was introduced, which combines average subsampled realized variance at slow time scale $ARV_t^{\Delta,k}$ and realized variance RV_t^Δ at fast time scale. In other words, when utilizing the two times scaled estimator, parameter Δ is the fast time scale, i.e., the highest possible sampling frequency available to the user, while the parameter k is the slow time scale frequency which defines the number of subgrids. For practical reasons, it is common to keep the fast time scale fixed, while the slow time scale should be determined optimally by minimizing RMSE (root mean square error) of the estimator.

As previously highlighted, RV_t^Δ is robust to microstructure noise as well as $ARV_t^{\Delta,k}$ and $TSRV_t^{\Delta,k}$, depending on the parameters Δ and k, while BPV_t^Δ, $MinRV_t^\Delta$ and $MedRV_t^\Delta$ are robust to price jumps only. Robust versions of IV have also been designed in the presence of both jumps and noise. In particular, robust version of two times scaled realized variance $RTSRV_t^{\Delta,k,\theta}$ was designed by [20]. This estimator requires one additional parameter θ which is employed as a threshold with respect to indicator function $I_i(\theta)$. The threshold θ is usually set to 9 indicating if returns are larger than three standard deviations from the mean. If returns are larger than three standard deviations from the mean, the indicator function has value 0 and 1 otherwise. The last estimator which is considered in this paper for comparison is $HYRV_t^{\Delta,\theta}$ designed by [28]. This estimator can be understood as a threshold realized variance which uses jump detection rule $I_i(\theta)$ to truncate the effect of jumps. It has to be mentioned that all estimators are sensitive to the selection of sampling frequency, and criteria for optimal frequency should be considered. Comprehensive study with respect to these criteria can be found in [13,15].

For better visual inspection of the differences between realized volatility measures, their performances are compared in Figure 1, considering DAX index (Frankfurt Stock Exchange) within the selected range of data from 1 March 2017 until 31 March 2017.
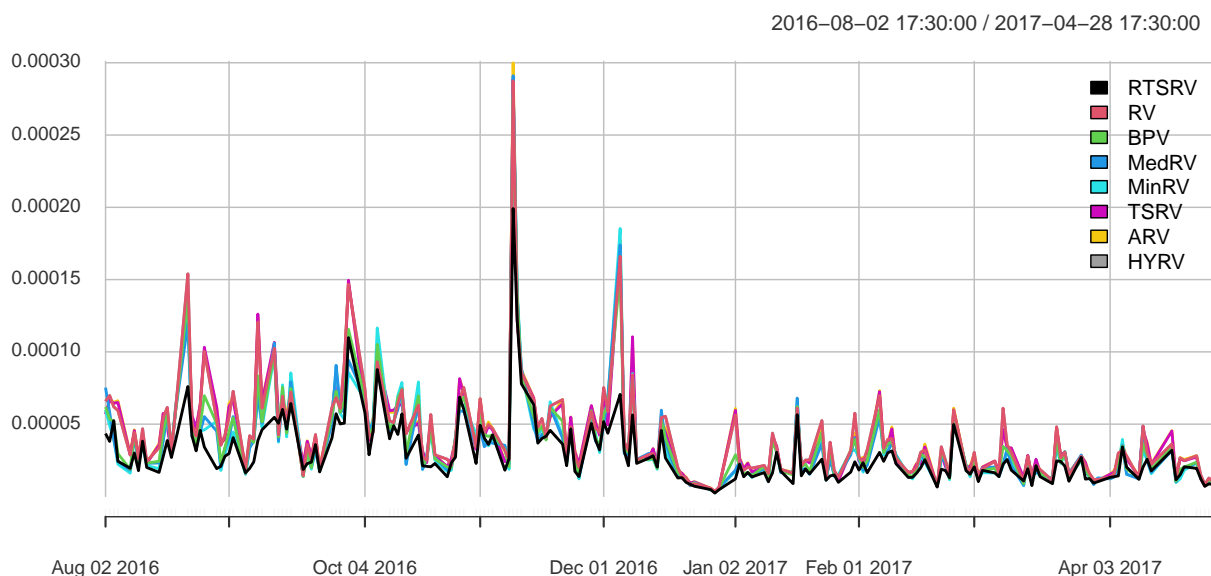

Figure 1. Realized volatility measures within a selected range of data for the DAX index.

After the computation of realized volatility measures with respect to optimal sampling frequency, three comparison methods are conducted in order to determine which volatility estimator fits the most to $RTSRV_t^{\Delta,k,\theta}$ as a benchmark. These three comparison methods are Mincer–Zarnowitz regression, probability integral transformation (PIT) test, and Gumbel copula upper tail dependence. Mincer–Zarnowitz regression is based on the overall performance of the realized volatility estimators. The PIT test was performed as a density goodness of fit procedure in order to test how the examined realized volatility estimators preform in comparison to $RTSRV_t^{\Delta,k,\theta}$. The upper tail dependence was used because it examines what happens in the extreme values or tails. The $RTSRV_t^{\Delta,k,\theta}$ is defined as a benchmark because it is robust to market microstructure noise as well as price jumps and non-synchronous trading in the intraday stock price series, while other estimators do not possess these features. These advantages were found in a simulation study only [20], but there is no strong empirical evidence of superiority of $RTSRV_t^{\Delta,k,\theta}$ considering developed European stock markets.

3. Empirical Results

Each comparison method is comprised of fitting seven realized volatility estimators to the benchmark $RTSRV_t^{\Delta,k,\theta}$. Before pairwise comparison, fixed parameters Δ and θ are set in advance, as previously argued $\Delta = 1$ and $\theta = 9$, while the optimal slow time scale sampling frequency k was selected for each European market based on minimizing the RMSE of the benchmark.

3.1. Sampling Frequency Selection

The fast time scale sampling frequency $\Delta = 1$ second is determined in front, according to data availability of the observed financial markets within the shortest, nonempty and equidistant intervals. In order to define the optimal slow time scale sampling frequency k, the root mean square error (RMSE) of the benchmark $RTSRV_t^{\Delta,k,\theta}$ is used. The RMSE of the benchmark for each stock index was calculated as the sum of its squared bias and its variance, and afterwards the RMSE was minimized with respect to slow time scale

frequency *k*, which corresponds to the optimal number of non-overlapping subsamples or subgrids, over which the slow time scales sum of squared returns is averaged [29].

Namely, it is suggested to use returns that are sampled as often as possible because one can get the maximum amount of information from data. However, if the sampling frequency is as high as it can be, it leads to a bias problem due to microstructure noise. There is a trade-off between the bias and efficiency while determining an optimal sampling frequency. One must establish the optimal sampling frequency for an observed financial market in order to reduce the bias but still keep the efficiency of the benchmark estimator. In the presence of jumps, there will also be bias in practical applications. Thus, it will have an effect on sampling frequency, which is also influenced by market structure, liquidity and microstructure noise. The way to increase the efficiency of estimators is to sub-sample (taking the average of an estimator across all possible sub-samples).

Even the robustness of the $RTSRV_t^{\Delta,k,\theta}$ to the choice of *k* has been theoretically demonstrated by [30], it is found at which optimal slow frequency a benchmark should be computed for each of four developed European markets (Table 3).

Table 3. Selection of sampling frequencies on both scales for every market individually.

European Market	Fast Time Scale Frequency Δ	Optimal Slow Time Scale Frequency *k*	$RMSE(RTSRV_t^{\Delta,k,\theta})$
Italy	1 s	13 s	0.00038
Germany	1 s	20 s	0.00019
France	1 s	10 s	0.00041
UK	1 s	30 s	0.00016

As can be seen in Figure 2, the optimal sampling frequency for Germany is 20 s; for the UK, it is 30 s; for Italy, it is 13 s; and, for France, it is 10 s. Figure 2 shows the root mean squared error (RMSE) of the robust two times scaled estimator for each stock market index against the number of subsamples.
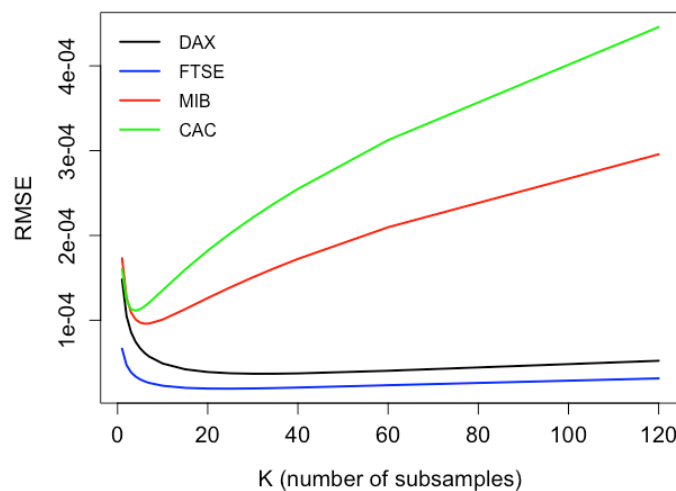

Figure 2. Root mean square error of the benchmark estimator $RTSRV_t^{\Delta,k,\theta}$ with respect to the number of subsamples (slow time scale frequencies).

3.2. Mincer–Zarnowitz Regression

After finding an optimal slow frequency, $RTSRV_t^{\Delta,k,\theta}$ can be easily computed for each developed market, so that the effect of microstructure noise becomes attenuated. Consecutively, the Mincer–Zarnowitz regression was estimated using GLS rather than OLS to improve the power and the size of the test on how well the realized volatility competitor $\hat{\sigma}_t^2$ fits to the benchmark:

$$RTSRV_t^{\Delta,k,\theta} = \beta_0 + \beta_1 \hat{\sigma}_t^2 + \epsilon_t \tag{1}$$

The Minzer–Zarnowitz regression approach is of great benefit for discriminating between competing estimators because it examines the bias [31,32]. In particular, it was jointly tested two restrictions on the parameters within a single null hypothesis $H_0 : \beta_0 = 0$ and $\beta_1 = 1$. The results, including χ^2 statistics and estimated parameters, are presented in Table 4.

Table 4. Results of Mincer–Zarnowitz regression.

Statistic	RV_t^{Δ}	BPV_t^{Δ}	$MinRV_t^{\Delta}$	$MedRV_t^{\Delta}$	$ARV_t^{\Delta,k}$	$TSRV_t^{\Delta,k}$	$HYRV_t^{\Delta,\theta}$
MIB	8375.1 ***	5442.6 ***	2298.0 ***	3449.3 ***	7530.1 ***	8841.3 ***	8274.6 ***
$\hat{\beta}_0$	0.000008	0.000001	0.000002	0.000003	0.000007	0.000006	0.000008
$\hat{\beta}_1$	0.508353	0.739918	0.799547	0.779631	0.518399	0.504905	0.508959
DAX	6517.3 ***	3404.7 ***	1853.3 ***	2281.7 ***	6732.8 ***	8001.1 ***	6517.1 ***
$\hat{\beta}_0$	−0.000008	−0.000003	−0.000001	−0.000002	−0.000009	−0.000008	−0.000008
$\hat{\beta}_1$	0.766997	0.820539	0.856405	0.850240	0.782201	0.763453	0.766998
CAC	4611.3 ***	2776.1 ***	3739.2 ***	2359.3 ***	6792.5 ***	1853.2 ***	470.67 ***
$\hat{\beta}_0$	0.000004	0.000008	0.000011	0.000010	0.000005	0.000002	0.000005
$\hat{\beta}_1$	1.006286	1.355853	1.426721	1.275828	1.012190	0.792073	1.006807
FTSE	7845.9 ***	5365.1 ***	4039.8 ***	4296.5 ***	8446.9 ***	9819.5 ***	7845.7 ***
$\hat{\beta}_0$	−0.000002	−0.000001	0.000001	0.000001	−0.000002	−0.000002	−0.000003
$\hat{\beta}_1$	0.724623	0.769927	0.777687	0.782853	0.730009	0.713995	0.724622

*** Indicates the significance at 1% level.

For all four European markets under consideration and seven volatility estimators, the null hypothesis was rejected at a significance level of 1%. This showed how seven observed volatility estimators do not fit well to the benchmark. For illustration purposes, Figure 3a,b are scatter plots indicating that both bipower variation BPV_t^{Δ}, which is jump robust only, and average subsampled realized variance $ARV_t^{\Delta,k}$, which is only robust to microstructure noise, underestimate the benchmark.
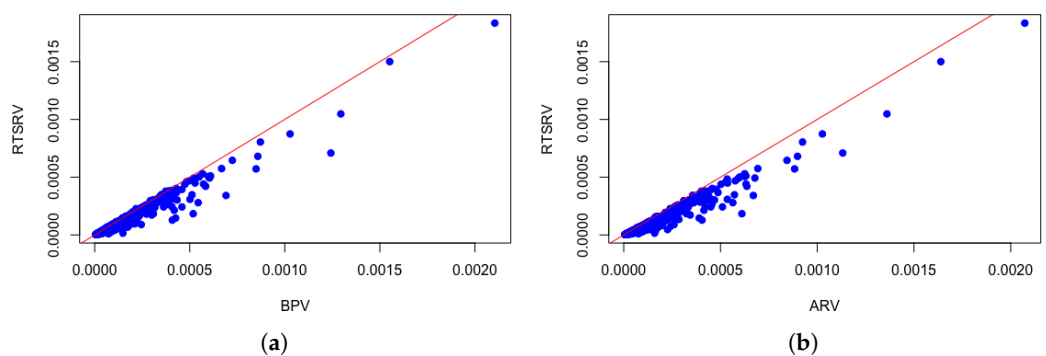

(**a**) (**b**)

Figure 3. Jump robust two times scaled realized variance against other realized volatility estimators. (**a**) benchmark versus bipower variation; (**b**) benchmark *versus* average subsampled RV.

3.3. Probability Integral Transformation and Kolmogorov–Smirnov Test

The probability integral transformation (PIT) was used to check whether the difference between $RTSRV_t^{\Delta,k,\theta}$ and other competing volatility estimators is uniformly distributed. Namely, uniform transformed volatility estimates are obtained by empirical distribution function of the same estimates from the given estimator. The pairwise differences between two uniform transformed estimates are used as an input in testing the null hypothesis about uniform distribution of these differences by the Kolmogorov–Smirnov test.

The results, including Kolmogorov–Smirnov test statistics, are presented in Table 5. For all European markets, the null hypothesis was rejected at a significance level of 1%, indicating a significant divergent performance of each competing estimator towards a benchmark.

Table 5. Kolmogorov–Smirnov tests using probability integral transformations.

Statistic	RV_t^Δ	BPV_t^Δ	$MinRV_t^\Delta$	$MedRV_t^\Delta$	$ARV_t^{\Delta,k}$	$TSRV_t^{\Delta,k}$	$HYRV_t^{\Delta,\theta}$
MIB	9577 ***	14,373 ***	16,639 ***	14,399 ***	10,161 ***	8801 ***	9554 ***
DAX	12,953 ***	16,349 ***	15,272 ***	15,123 ***	11,598 ***	10,790 ***	12,955 ***
CAC	11,580 ***	10,692 ***	10,152 ***	10,636 ***	11,624 ***	17,964 ***	11,579 ***
FTSE	13,824 ***	18,095 ***	12,495 ***	20,727 ***	13,099 ***	14,016 ***	13,824 ***

*** Indicates the significance at 1% level.

3.4. Upper Tail Dependence

This research utilizes the upper tail dependence coefficient, a result of the Gumbel copula function, which is found useful for measuring extreme dependence, i.e., a dependence above a high quartile. While the Mincer–Zarnowitz regression and Kolmogorov–Smirnov test have not demonstrated the preference of a specific estimator, the copula-based approach can be a powerful and suitable for comparison [29]. The paper focuses on a Gumbel copula as an extreme value copula that is not elliptical, due to the similar behavior of volatility estimator over time, and sometimes takes on extremely large values. The Gumbel copula function is given by:

$$C(u,v) = exp\left(- \left[-ln(u)^2 - ln(v)^\delta \right]^{\frac{1}{\delta}} \right) \tag{2}$$

where $u, v \in [0,1]$ are in fact PIT transformations computed as an empirical cumulative distribution function, with parameter $1 \le \delta < \infty$ that controls the upper tail dependence accordingly $\lambda_u = 2 - 2^{\frac{1}{\delta}}$. A Gumbel copula is fitted to each pair of volatility estimators of which one is always $RTSRV_t^{\Delta,k,\theta}$, and Gumbel parameter is estimated using the pseudo likelihood method (PLM).

The results given in Table 6 present the upper tail dependence coefficient λ_u based on estimated δ from the Gumbel copula function. The results indicate that, for Italy, Germany, and the UK, $RTSRV_t^{\Delta,k,\theta}$ has the highest upper tail dependence with BPV_t^Δ, $MinRV_t^\Delta$, and $MedRV_t^\Delta$ volatility estimators.

Table 6. Upper tail dependence coefficients from the Gumbel copulas.

λ_u	RV_t^Δ	BPV_t^Δ	$MinRV_t^\Delta$	$MedRV_t^\Delta$	$ARV_t^{\Delta,k}$	$TSRV_t^{\Delta,k}$	$HYRV_t^{\Delta,\theta}$
MIB	0.813 ***	0.851 ***	0.853 ***	0.859 ***	0.819 ***	0.809 ***	0.814 ***
DAX	0.846 ***	0.868 ***	0.869 ***	0.874 ***	0.850 ***	0.849 ***	0.846 ***
CAC	0.826 ***	0.831 ***	0.825 ***	0.829 ***	0.829 ***	0.856 ***	0.826 ***
FTSE	0.824 ***	0.858 ***	0.858 ***	0.864 ***	0.828 ***	0.828 ***	0.824 ***

*** Indicates the significance at 1% level.

4. Discussion

Each competing volatility estimator was tested against benchmark $RTSRV_t^{\Delta,k,\theta}$ using Mincer–Zarnowitz regression, PIT test, and upper tail dependence measurements within the Gumbel copula. The results of the first two methods for comparison of the realized volatility estimators indicate that RV_t^Δ, $ARV_t^{\Delta,k}$, $TSRV_t^{\Delta,k}$, and $HYRV_t^{\Delta,\theta}$ underestimate the performance of estimates during severe stress and price jumps. Therefore, the results do not give a clear answer regarding which estimator fits the best to the benchmark. In that case, the upper tail dependence is a favorable method to use because it takes into account extreme values.

Among all the competing estimators, only jump robust ones have produced almost similar volatility estimates as $RTSRV_t^{\Delta,k,\theta}$. In the case of France, $RTSRV_t^{\Delta,k,\theta}$ is best fitted

with $RTSRV_t^{\Delta,k}$, BPV_t^{Δ}, $ARV_t^{\Delta,k}$, and $MedRV_t^{\Delta}$ volatility estimators. As $TSRV_t^{\Delta,k}$ is robust to microstructure noise, it is of no surprise that estimates are as good as $RTSRV_t^{\Delta,k,\theta}$.

While estimating the integrated variance in a simulation study, looking at the asymptotic properties, the two times scaled estimator was shown to be consistent and unbiased to microstructure noise [33].

The results are important to financial analysts and investors because they offer a recommendation which realized the volatility estimator to use for the observed stock indices. This research contributes to the previous studies with an empirical dataset consisting of high-frequency price observations comprising four main European market indices (DAX, CAC, FTSE and MIB) because there are very few studies that take into account the calculations of realized volatility estimators on developed European markets.

5. Conclusions

The popularity of high-frequency data is enlarging, even if it is still hard to obtain such data. However, the scientific focus on research of volatility estimators has increased and expanded the knowledge from the already existing literature. This increase is mainly due to the availability of high-frequency data. Even though realized variance RV_t^{Δ} is the most used high-frequency estimator, it is biased due to microstructure noise. The benchmark robust two times scaled realized variance $RTSRV_t^{\Delta,k,\beta}$ is microstructure noise and jump robust. In this research, the data observed are intraday 1 s observations. It is determined that the optimal sampling frequency for the robust two times scaled realized variance $RTSRV_t^{\Delta,k,\beta}$ is from 10 to 30 s. There is no consensus in the previous studies regarding the "best" realized volatility estimator. The objective of this research paper is to determine whether the robust two times scaled realized variance $RTSRV_t^{\Delta,k,\beta}$ is a superior volatility estimator for each of the four considered European markets by performance comparison of two groups of estimators, i.e., estimators which are robust to microstructure noise as well as jump-robust estimators. Due to inconclusive results from Mincer–Zarnowitz and PIT test, the upper tail dependence was introduced because it examines the events in tails i.e., extreme values. The results indicated that the medianized block of three returns $MedRV_t^{\Delta}$ performed most similar to $RTSRV_t^{\Delta,k,\beta}$ for Italy, Germany and UK. For France, the two times scaled realized variance $TSRV_t^{\Delta,k}$ realized volatility estimator was the most similar (approximately equal) to the benchmark. Since the medianized block of three returns $MedRV_t^{\Delta}$ is robust only to price jumps, we conclude that the Italian, German and UK financial markets are more contaminated by price jumps than by microstructure noise. The French financial market is more contaminated by microstructure noise than by price jumps. This contributes to the existing literature in several ways. The main finding considers the selection of optimal slow time scale frequency in favor of two times scaled estimator in each market individually, at the same time ensuring robustness to price jumps. Another novelty is the usage of a combination of three tests for benchmarking: Mincer–Zarnowitz regression, PIT test and upper tail dependence test within the Gumbel copula where the results are given for each of the observed developed European markets. The direction of further research would be to investigate the realized covariance of different assets [34,35] using high–frequency data, which is a great challenge due to non–synchronization issue, presence of co–jumps and sampling frequency selection.

Author Contributions: Conceptualization, J.A. and M.Č.; methodology, J.A. and M.Č.; software, M.Č.; validation, A.Ž.; formal analysis, J.A.; investigation, M.Č.; resources, J. A.; data curation, M.Č.; writing—original draft preparation, M.Č.; writing—review and editing, J.A. and A.Ž.; visualization, J.A. and M.Č.; supervision, A.Ž.; project administration, A.Ž. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Abbreviations

The following abbreviations are used in this manuscript:

ARV	Average subsampled realized variance
BPV	Bipower variation
CAC	*Cotation Assiste en Continu*—a benchmark French stock market index
DAX	*Deutscher Aktien IndeX*—a benchmark German stock market index
FTSE	*Financial Times Stock Exchange*—a benchmark British stock market index
GLS	Generalized least squares
HYRV	Hayashi–Yoshida realized variance
IV	Integrated variance
MedRV	Medianized block of three returns
MIB	*Milano Indice di Borsa*—a benchmark Italian stock market index
MinRV	Minimized block of two returns
OLS	Ordinary least squares
PLM	Pseudo likelihood method
PIT	Probability integral transformation
RMSE	Root mean squared error
RTSRV	Jump robust two times scaled realized variance
RV	Realized variance
TRTH	Thomson Reuters Tick History data base
TSRV	Two times scaled realized variance
UK	United Kingdom
US	United States

References

1. Andersen, T.G.; Bollerslev, T.; Diebold, F.X.; Labys, P. The Distribution of Realized Exchange Rate Volatility. *J. Am. Stat. Assoc.* **2001**, *96*, 42–55. [CrossRef]
2. Arnerić, J.; Matković, M. Challenges of integrated variance estimation in emerging stock markets. *Zbornik Radova Ekonomski Fakultet u Rijeka* **2019**, *37*, 713–739. [CrossRef]
3. Floros, C.; Gkillas, K.; Konstantatos, C.; Tsagkanos, A. Realized Measures to Explain Volatility Changes over Time. *J. Risk Financ. Manag.* **2020**, *13*, 125. [CrossRef]
4. Boudt, K.; Cornelissen, J.; Croux, C.; Laurent, S. Nonparametric Tests for Intraday Jumps: Impact of Periodicity and Microstructure Noise. In *Handbook of Volatility Models and Their Applications*; Bauwens, L., Hafner, C., Laurent, S., Eds.; John Wiley and Sons: Hoboken, NJ, USA, 2012; pp. 447–463. [CrossRef]
5. Brownlees, C.T.; Gallo, G.M. Financial econometric analysis at ultra–high frequency: Data handling concerns. *Comput. Stat. Data Anal.* **2006**, *51*, 223–2245. [CrossRef]
6. Clements, A.; Liao, Y. Forecasting the variance of stock index returns using jumps and cojumps. *Int. J. Forecast.* **2017**, *33*, 729–742. [CrossRef]
7. Barndorff-Nielsen, O.E.; Shephard, N. Estimating quadratic variation using realized volatility. *J. Appl. Econom.* **2002**, *17*, 457–477. [CrossRef]
8. Andersen, T.G.; Bollerslev, T.; Diebold, F.X.; Labys, P. Modeling and Forecasting Realized Volatility. *Econometrica* **2003**, *71*, 579–625. [CrossRef]
9. Liu, C.; Maheu, J.M. Intraday dynamics of volatility and duration: Evidence from Chinese stocks. *Pac. Basin Financ. J.* **2012**, *20*, 329–348. [CrossRef]

10. Hansen, P.R.; Lunde, A. A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *J. Financ. Econom.* **2005**, *3*, 525–554. [CrossRef]
11. Omen, R.C.A. Properties of Bias–Corrected Realized Variance Under Alternative Sampling Schemes. *J. Financ. Econom.* **2005**, *3*, 555–577. [CrossRef]
12. Lee, S.S.; Mykland, P.A. Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financ. Stud.* **2008**, *21*, 2535–2563. [CrossRef]
13. Bandi, F.M.; Russell, J.R. Microstructure Noise, Realized Volatility, and Optimal Sampling. *Rev. Econ. Stud.* **2008**, *7*, 339–369. [CrossRef]
14. Aït-Sahalia, Y.; Mykland, P.A.; Zhang, L. Ultra high frequency volatility estimators with dependent microstructure noise. *J. Econom.* **2011**, *160*, 160–175. [CrossRef]
15. Aït-Sahalia, Y.; Mykland, P.A.; Zhang, L. How Often to Sample a Continuous–Time Process in the Presence of Market Microstructure Noise. *Rev. Financ. Stud.* **2005**, *18*, 351–416. [CrossRef]
16. Andersen, T.G.; Bollerslev, T. Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *Int. Econ. Rev.* **1998**, *39*, 885–905. [CrossRef]
17. Zhang, L. Estimating covariation: Epps Effect, microstructure noise. *J. Econom.* **2011**, *160*, 33–47. [CrossRef]
18. Barndorff–Nielsen, O.E.; Shephard, N. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *J. Financ. Econom.* **2006**, *4*, 1–30. [CrossRef]
19. Boudt, K.; Croux, C.; Laurent, S. Outlyingness weighted covariation. *J. Econom.* **2011**, *9*, 657–684. [CrossRef]
20. Boudt, K.; Zhang, J. Jump robust two time scale covariance estimation and realized volatility budgets. *Quant. Financ.* **2013**, *15*, 1041–1054. [CrossRef]
21. Andersen, T.G.; Dobrev, D.; Schaumburg, E. Jump–robust volatility estimation using nearest neighbor truncation. *J. Econom.* **2012**, *169*, 75–93. [CrossRef]
22. Hanousek, J.; Kočenda, E.; Novotný, J. The identification of price jumps. *Monte Carlo Methods Appl.* **2012**, *18*, 53–77. [CrossRef]
23. Hanousek, J.; Kočenda, E.; Novotný, J. Intraday price behavior during information arrival in emerging markets. In *Market Microstructure in Emerging and Developed Markets*; Baker, H.K., Kiymaz, H., Eds.; John Wiley and Sons: New York, NY, USA, 2013; pp. 445–462. [CrossRef]
24. Corsi, F.; Pirino, D.; Ren ′o, R. Threshold bipower variation and the impact of jumps on volatility forecasting. *J. Econom.* **2010**, *159*, 276–288. [CrossRef]
25. Huang, X.; Tauchen, G. The relative contribution of jumps to total price variance. *J. Financ. Econom.* **2005**, *3*, 456–499. [CrossRef]
26. Andersen, T.G.; Bollerslev, T.; Diebold, F.X. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Rev. Econ. Stat.* **2007**, *89*, 710–720. [CrossRef]
27. Zhang, L.; Mykland, P.A.; Aït–Sahalia, Y. A Tale of Two Time Scales. *J. Am. Stat. Assoc.* **2005**, *100*, 1394–1411. [CrossRef]
28. Hayashi, T.; Yoshida, N. On covariance estimation of non–synchronously observed diffusion processes. *Bernoulli* **2005**, *11*, 359–379. [CrossRef]
29. Arnerić, J.; Matković, M. Sorić, P. Comparison of range–based volatility estimators against integrated volatility in European emerging markets. *Financ. Res. Lett.* **2019**, *28*, 118–124. [CrossRef]
30. Aït-Sahalia, Y.; Jacod, J. Testing for jumps in a discretely observed process. *Ann. Stat.* **2009**, *37*, 182–222. [CrossRef]
31. Patton, A.J. Data–based ranking of realised volatility estimators. *J. Econom.* **2011**, *161*, 284–303. [CrossRef]
32. Patton, A.J.; Sheppard, K. Optimal combinations of realised volatility estimators. *Int. J. Forecast.* **2009**, *25*, 218–238. [CrossRef]
33. McAleer, M.; Medeiros, M.C. Realized Volatility: A Review. *Econom. Rev.* **2008**, *27*, 10–45. [CrossRef]
34. Batrancea, L.; Rus, M.I.; Masca, E.S.; Morar, I.D. Fiscal Pressure as a Trigger of Financial Performance for the Energy Industry: An Empirical Investigation across a 16-Year Period. *Energies* **2021**, *14*, 3769. [CrossRef]
35. Batrancea, L. An Econometric Approach Regarding the Impact of Fiscal Pressure on Equilibrium: Evidence from Electricity, Gas and Oil Companies Listed on the New York Stock Exchange. *Mathematics* **2021**, *9*, 630. [CrossRef]