УДК 330.46

O. CHERNYAK, I. VASYLCHENKO

Taras Shevchenko National University of Kyiv

APPLICATION OF LOWER ORDER STATISTICS IN ROBUST NEIGHBORHOOD TRUNCATION QUARTICITY ESTIMATORS

Authors analyze robust neighborhood truncation estimators which operate on lower order statistics log-returns. Performed simulations proved additional efficiency and jump robustness of these estimators of integrated quarticity.

Keywords: *asset price, integrated volatility, integrated quarticity, high frequency data, market microstructure noise.*

Last decade financial markets were highlighted with emergence and rapid development of the new industry sector - high frequency trading. Some years ago it took transactions more then ten seconds in order to execute, while nowadays hundreds of them can squeeze in one second. Such a change was mainly driven by decimalization of trading prices and advances in technologies: computational powers and data transfer speeds have grown exponentially. While such operating speeds are unreachable for human trading, more and more market participants started building up computational centers and developing quantitative algorithms with a goal to outperform competitors.

Eventually, these market transformations have led to generation of enormous amounts of high frequency data sets, which due to their structure sometimes require review of statistical approaches or creation of radically new ones. Estimation of integrated volatility and integrated quarticity is one of those questions, which have gained a lot of attention in recent years. Irregularity of the intraday returns of the asset price within high frequency data sets coupled with microstructure noise required new robust approaches to estimating these values, thus, extensive work in this direction was conducted by solid number of authors.

In the paper [1] authors introduced for the first time complementary volatility measure, termed realized volatility, which is coupled together with realized quarticity measure.

Bipower variation, as an initial term in multipower variation estimator theory, was proposed by [2]. This paper shows that introduced realized bipower variation dispose some robustness to jumps in price processes. It was demonstrated that realized bipower variation can estimate integrated power volatility in stochastic volatility models and moreover, under some conditions, it can be a good measure to integrated variance in the presence of jumps.

Authors [3] came up with two new jump robust estimators of integrated variance based on high frequency return observations, namely MinRV and MedRV. Their findings prove that these estimators can be good alternative to the multipower variation estimators.

Article [4] presented the family of efficient robust neighborhood truncation (RNT) estimators for the integrated power variation based on the order statistics of a set of unbiased local power variation estimators on a block of adjacent returns. Efficient RNT estimators represent extension of neighborhood truncation estimator's theory.

One of the recent works [5], proposes new methodology based on Fourier analysis to estimate spot and integrated quarticity. Authors explain that Fourier methodology allows reconstructing the latent instantaneous volatility as a series expansion with coefficients gathered from the Fourier coefficients of the observable price variation and can be extended to higher even powers of volatility and to the multivariate case. They prove that the Fourier estimator of integrated quarticity is consistent in the absence of noise, then test this new methodology with the use of Monte Carlo experiments and apply it to S&P 500 index futures.

In [6] authors analyze in detail different volatility estimators under the presence of market microstructure noise. They also discuss influence of sampling frequency on efficiency of estimators and propose a way of achieving the optimal one under condition of asymptotically small noise.

In current paper we would like to focus attention on examination and comparison of different combinations of RNT quarticity estimators (RNTQ) that use lower order statistics of log-returns (LOS RNTQ) and higher order statistics (HOS RNTQ). Authors [7] made an assumption that LOS RNTQ estimators are more affected by market microstructure noise and did not include them to overall simulation analysis. This fact seemed to us being worth of further investigation, while based on the simulations performed in [8], estimators RNTQ6 1(123), RNTQ6 2(123) under the jump presence were one of the best ones in terms of bias and RMSE error, and in general demonstrated decent performance in simulations with stochastic volatility and sparse sampling of stock returns.

During computations we used pre-averaging technique published by [9]. In an elegant way it allows to lower the impact of market microstructure noise on the resulting quarticity estimations.

Integrated quarticity concept and some theories of its estimators can be looked through in [3], [4], [8] or any other related article, while now we would like to move directly to robust neighborhood truncation estimators and its application.

Let $r_i = S_i - S_{i-1}$, $i = 1,\ldots,n$ be n equally spaced logarithmic returns of the asset price. Then, we denote i-th block of absolute returns as $\underline{r}_{i,m} = \left(|r_i|,\ldots,|r_{i+m-1}|\right)$, $i = 1,\ldots,n-m+1$ and j-th order statistic of the i-th absolute return block as $q_j\left(|r_1|,\ldots,|r_m|\right) = q_j\left(\underline{r}_{i,m}\right)$. Naturally $q_1\left(\underline{r}_{i,m}\right) \le \ldots \le q_m\left(\underline{r}_{i,m}\right)$.

Following these notations, baseline Neighborhood Truncation estimator (NT) is given by

$$NT_n^{(j,m)}(p) = d_{(j,m)}(p)\left(\frac{n^{p/2}}{n-m+1}\right)\sum_{i=1}^{n-m+1}\left[q_j\left(\underline{r}_{i,m}\right)\right]^p, \qquad j = 1,\ldots,m, \tag{1}$$

where $d_{(j,m)}(p) = \left(E\left[q_j\left(|Z_1|^p,\ldots,|Z_m|^p\right)\right]\right)^{-1}$, $Z_i \sim i.i.d.\ N(0,1)$, $i = 1,\ldots,m$.

Placing scaling factor $d_{(j,m)}(p)$ in front of p-th power of the j-th absolute order statistic gives as an unbiased estimator for σ^p.

Neighborhood truncation estimator is a family of estimators, which also incorporates such estimators as $MinPV(p)$ and $MedPV(p)$. In fact, $MinPV(p)$ is $NT_n^{(1,2)}(p)$ with a scaling factor $d_{(1,2)}(p)$ and $MedPV(p)$ is $NT_n^{(2,3)}(p)$ with $d_{(2,3)}(p)$.

So far, we have been speaking about straightforward picking j-th statistic from each of return blocks, applying respective power and scaling coefficient. Finally, summarizing all these values provides us with the NT estimation of power variance.

Robust neighborhood truncation estimator represents further extension of this approach. General algorithm, proposed in [7], is defined by:

$$RNT_n^{(j,I)}(p) = d_{(j,I)}(p)\frac{1}{n-m+1}\sum_{i=1}^{n-m+1} q_j\left(\varepsilon_{k_1}(\underline{r}_{i,m}),\ldots,\varepsilon_{k_H}(\underline{r}_{i,m})\right) \tag{2}$$

where

$$\varepsilon_{k_H} = d_{(k_H,m)}(p)n^{p/2}\left[q_{k_H}(\underline{r}_{i,m})\right]^p. \tag{3}$$

Roughly speaking such a setup provides a linear combination of NT estimators, which secures better robustness and efficiency comparing to the baseline NT estimation.

Firstly, within the given i-th return block we calculate properly scaled functional of needed order statistics $\varepsilon_{k_1}(\underline{r}_{i,m}),\ldots,\varepsilon_{k_H}(\underline{r}_{i,m})$. Vector $I=(k_1,\ldots,k_H),\ 1\le H\le m$ in this case defines which vector of order statistics we would like in each concrete return block. To received set of H unbiased estimators for σ^p $\{\varepsilon_{k_1},\ldots,\varepsilon_{k_H}\}$ we apply j-th order statistics, which is scaled by respective factor $d_{(j,I)}(p)$. This gives us final value of return functional for the i-th return block.

Naturally, the $d_{(j,I)}(p)$ scaling factor, which converts j-th order statistics, applied to the set of unbiased σ^p estimators $\varepsilon_{k_1}(\underline{r}_{i,m}),\ldots,\varepsilon_{k_H}(\underline{r}_{i,m})$, into a robust unbiased estimator of the given i-th return block, depends on the initial RNTQ estimator configuration:

$$d_{(j,I)}(p) = \left(E\left[q_j\left(d_{(k_1,m)}(p)Z^p{}_{(k_1,m)},\ldots,d_{(k_H,m)}(p)Z^p{}_{(k_H,m)}\right)\right]\right)^{-1} \tag{4}$$

$$d_{(k_h,m)}(p) = \left(E\left[q_{k_h}\left(|Z_1|^p,\ldots,|Z_m|^p\right)\right]\right)^{-1},\ Z_i \sim i.i.d.\ N(0,1), \tag{5}$$

$$I=(k_1,\ldots,k_H),\ 1\le H\le m. \tag{6}$$

For the further examination we have picked a group of RNTQ estimators which covered various order statistics configurations:

- RNTQ5 1(123) RNTQ5 2(123);
- RNTQ5 1(345) RNTQ5 2(345);
- RNTQ6 1(123) RNTQ6 2(123);
- RNTQ6 1(456) RNTQ6 2(456);
- RNTQ7 1(1234) RNTQ7 2(1234);
- RNTQ7 1(4567).

Numbers 1 and 2 before parentheses are values of j coefficient and combinations $(123)\ldots(4567)$ are combinations of vector $I=(k_1,\ldots,k_H),\ 1\le H\le m$. In proposed setup estimator RNTQ7 2(4567) was omitted due to pure efficiency caused by low jump robustness.

Due to the fact that usually there is no closed form solutions for $d_{(j,I)}(p)$ and $d_{(k_h,m)}(p)$ values, they were obtained using equations 4-6 via simulations (Table 1).

Table 1. **Scaling factors of RNTQ5 and RNTQ7 estimators for different order statistics**

	$d_{(1,3)}(4)$	$d_{(2,3)}(4)$	$d_{(3,3)}(4)$	$d_{(1,I)}(4)$	$d_{(2,I)}(4)$
$RNTQ5\,(123)$	35,14029	5,75253	1,44264	3,67611	1,31886
$RNTQ5\,(345)$	1,44314	0,39879	0,08642	2,60658	1,21894
$RNTQ6\,(123)$	62,75698	10,88057	2,96839	3,52776	1,29788
$RNTQ6\,(456)$	0,95240	0,30849	0,07552	2,32949	1,17506

	$d_{(1,4)}(4)$	$d_{(2,4)}(4)$	$d_{(3,4)}(4)$	$d_{(4,4)}(4)$	$d_{(1,I)}(4)$
$RNTQ7\,(1234)$	104,37888	18,57741	5,31021	1,85594	4,56927
$RNTQ7\,(4567)$	1,85712	0,69869	0,25216	0,06739	2,70389

One can observe that together with the rise of the returns quantity, coefficients grow even more, with a sharp distinction between the groups of lower order and higher order returns.

The asymptotic distribution of RNTQ estimator for pure BM process without jumps [4]:

$$\sqrt{n}\left(RNTQ_N^{(j,I)} - \int_0^1 \sigma_s^{\,4} ds \right) \xrightarrow{L} N\left(0, \eta(j,I;4) \int_0^1 \sigma_s^{\,8} ds \right), \quad j = 1, \ldots, H. \tag{7}$$

While trying to approximate to some extent the efficiency factors $\eta(j,I;4)$ of estimators from chosen target group, we have received values postulated in the table 2.

Table 2. **Approximate values of $\eta(j,I;4)$ for some RNTQ5, RNTQ6 and RNTQ7 estimators applied to the pure Brownian motion process**

RNTQ5 1(123)	**30,59377**	RNTQ6 1(456)	11,0131
RNTQ5 2(123)	**22,59096**	RNTQ6 2(456)	10,30176
RNTQ5 1(345)	11,51716	RNTQ7 1(1234)	**40,48261**
RNTQ5 2(345)	10,59576	RNTQ7 4(1234)	**29,89296**
RNTQ6 1(123)	**39,75629**	RNTQ7 1(4567)	11,29826
RNTQ6 2(123)	**28,65126**		

Analogously to the MPV estimator's property mentioned in [7], scrutinized RNTQ estimators, under the no-jump null hypothesis, have a tendency to improve efficiency when block size of returns gets smaller. Another important result is, that under pure Brownian motion process (BM), HOS RNTQ perform definitely better then LOS RNTQ. Estimators RNTQ5 1(345), RNTQ 2(345), RNTQ6 1(456), RNT6 2(456) and even RNT7 1(4567) have asymptotic variances settled around values 10–11. Meanwhile, LOS RNTQ estimators starting from RNTQ5 1(123) constantly grow in variance measure, hitting values up to 30–40.

In order to examine proposed RNTQ estimators applied to some market patterns, we have used following models:

– Brownian motion process (BM) with and without jumps;

– Stochastic volatility model with intraday U-shape volatility pattern (SV-U model);

– Sparse sampling model (irregular trade intervals).

Within all the models (except sparse sampling), we simulate data between 9:30 and 16:00 with a 1 second interval, which results in 23400 observations per day.

For the sparse sampling model another approach is used: initially, for each trading day we generated BM process with 23400 values, and at the next step values out of resulted time series were picked using Poisson distribution with $\lambda = 2$, in order to get non-homogeneous data time-arrivals.

This approach was providing us with a sample, whose size varied on average between 10850 and 11050 time points.

Unconditional daily volatility is set to 0.000159, which is equivalent to around 20% per annum. In each of the cases 2400 days were simulated, which covers almost 10 years of stock market activity.

BM model with one random jump clearly showed significantly greater biases of estimators RNTQ5 1(345), RNTQ5 2(345), RNTQ6 1(456), RNTQ6 2(456) and RNTQ7 4(4567) (especially with a sampling window greater then 120 seconds). All the other estimators, while grouped quite tightly, together show relatively small bias (Fig. 1). With RMSE errors situation looks quite similar, with a breaking point again at 120 second sampling window size.

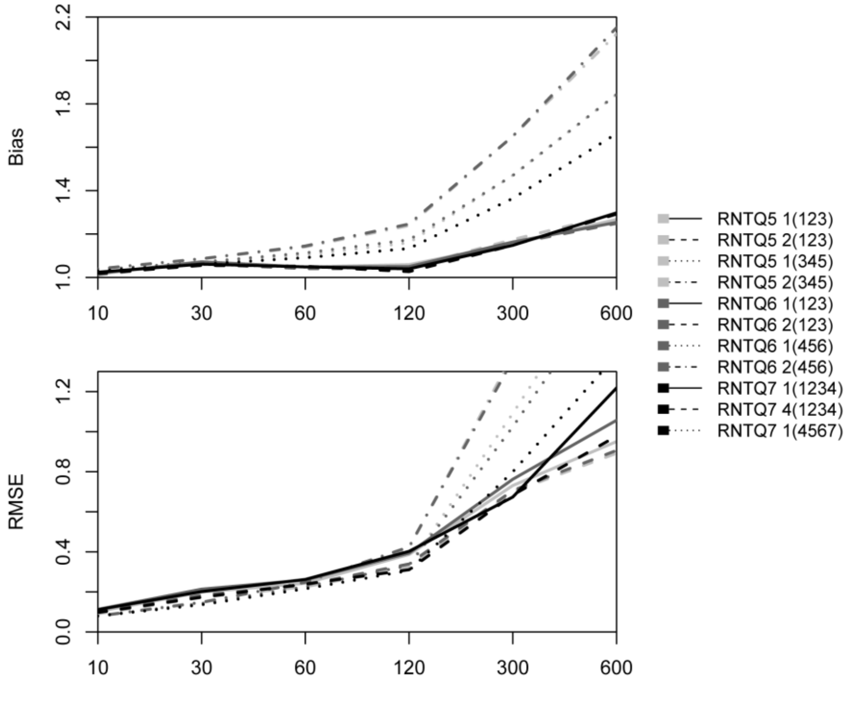

Fig.1. **RNTQ estimators applied to BM stochastic process with 1 jump of a randomly distributed 2–5% size**

We can definitely say that LOS RNTQ are more robust to the presence of a random jump within trading interval. This seems reasonable, while picking values out of the group of lower order returns, most surely will let us omit the jump component, in case such is present within observable interval. This simulation does not demonstrate difference between, say, estimators RNTQ5 1(345) and RNTQ7 1(4567), but we suppose it will be more evident under presence of greater quantity of jumps, which can be verified separately.

Under the simulation of SV-U model all estimators tend to have downward bias, and it is hard to single out some particular one significantly better then the others (Fig. 2). Estimators like RNTQ5 2(345) or RNTQ6 2(456) are slightly more efficient, both in terms of bias and RMSE error. Overall, applied to SV-U model, HOS RNTQ estimators are a bit more efficient then LOS RNT

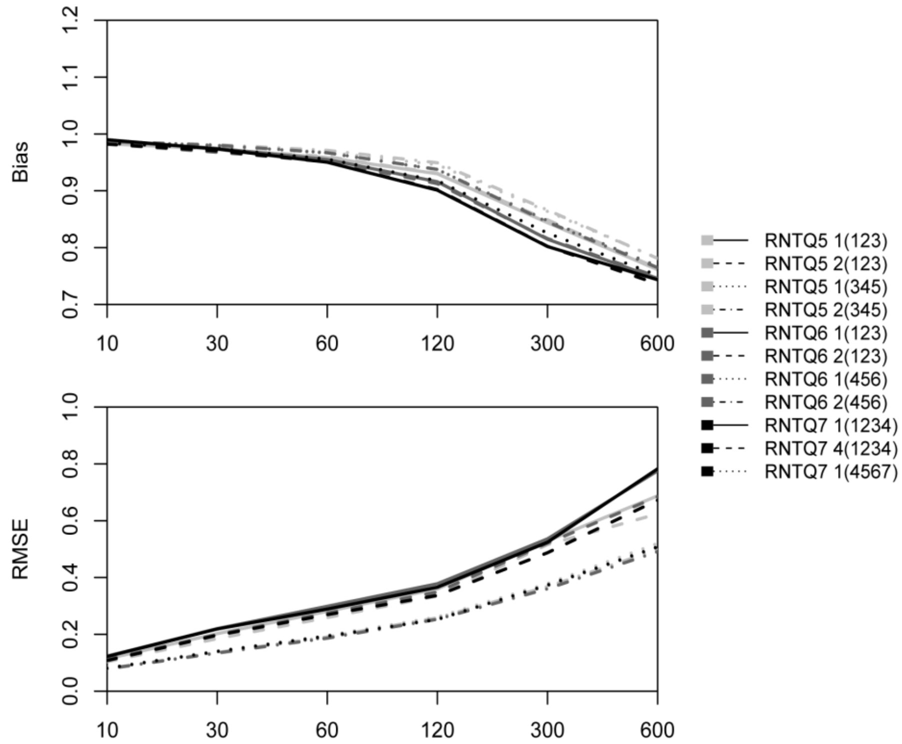

Fig. 2. **RNTQ estimators applied to stochastic volatility model with intraday U-shape**

Last simulation showed instability of LOS RNTQ estimators against sampling window size. Fig. 3 reveals that choice of sampling window is quite important when data is sparsely sampled - picking appropriate one can let us reach lower levels of bias. Based on Fig. 3, choosing pre-averaging sampling windows of 10-30 seconds and less (as well as greater then 300 seconds in our case), can lead to rise in bias.

On the contrary to that, HOS RNTQ estimators revealed constantly good performance, all the time stably demonstrating low bias. Thus, in case we speak about non-equidistant returns, in terms of lower RMSE errors and bias, HOS RNTQ seem to be more attractive then LOS RNTQ.

Eventually, LOS RNTQ estimators were much more jump robust then HOS RNTQ, and they also showed decent performance in stochastic volatility model and Brownian motion with sparse sampling simulations. Constructed models and simulation results are in line with respective literature, thus derived efficiency of LOS RNTQ estimators appears to be reliable enough and should not be rejected.

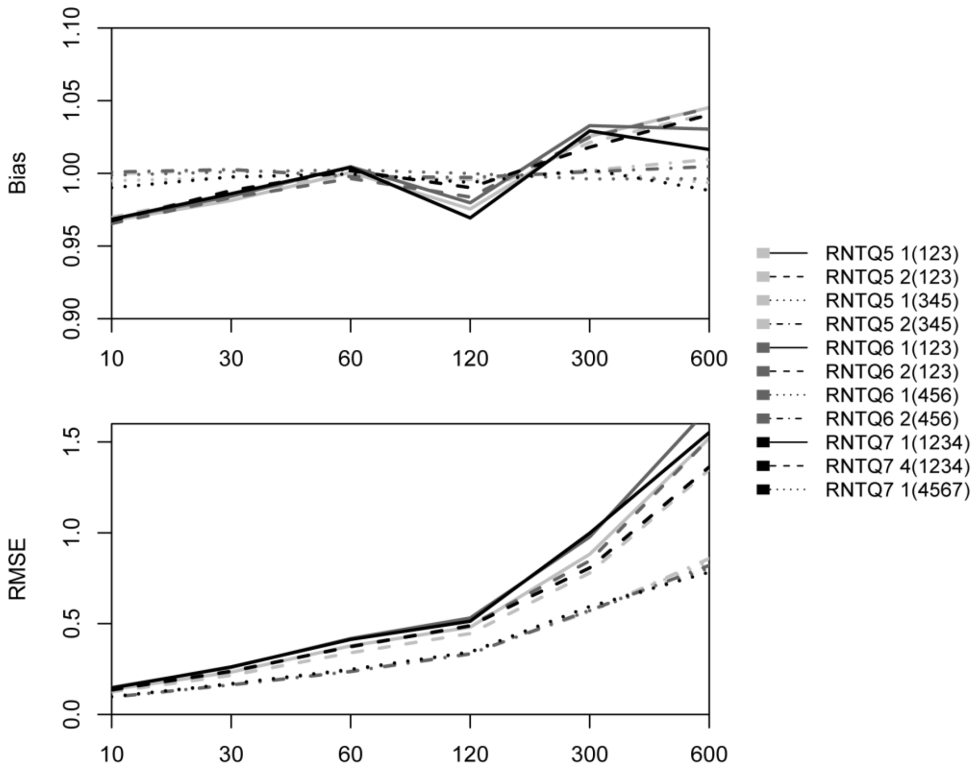

Fig. 3: **RNTQ estimators applied to BM stochastic process with sparse sampling**

Possible way to extend this research include examination of bigger set of more diverse RNTQ estimators and their assessment with simulations that would mix several price process models at one time.

References

1. Andersen T. G., Bollerslev T., Diebold F. X., Labys P. The distribution of realized exchange rate volatility // Journal of American Statistical Association.–2001– Vol. 96(453)– P. 42–55.

2. Barndorff-Nielsen O. E., Shephard N. Power and bipower variation with stochastic volatility and jumps // Journal of Financial Econometrics.–2004 – Vol.2, No.1.– P.1–37.

3. Andersen T. G., Dobrev D., Schaumburg E. Jump-robust volatility estimation using nearest neighbor truncation // NBER Working Paper No. 15533, November 2009.

4. Andersen T. G., Dobrev D., Schaumburg E. A functional filtering and neighborhood truncation approach to integrated quarticity estimation // National Bureau of Economic Research Working Papers, 2011.

5. Mancino M. E., Sanfelici S. Estimation of quarticity with high frequency data. January 2012.

6. Zhang L., Mykland P. A., Ait-Sahalia Y. A tale of two scales: Determining integrated volatility with noisy high-frequency data // Journal of the American Statistical Association.–2005– Vol.100(472).

7. Andersen T. G., Dobrev D., Schaumburg E. A functional filtering and neighborhood truncation approach to integrated quarticity estimation // National Bureau of Economic Research Working Papers, 2011.

8. Vasylchenko I.I. Quarticity estimation based on high frequency financial data // European Science and Technology: 2nd International scientific conference. – Bildungszentrum Rdk e.V. Wiesbaden 2012.

9. Jacod J., Li Y., Mykland P. A., Podolskij M., Vetter M. Microstructure noise in the continuous case: The pre-averaging approach // Stochastic Processes and their Applications.–2009.–No.119.– P.2249–2276.

Стаття надійшла до редакції 11.10.2012

Застосування нижніх порядкових статистик в робастних відтинаючих оцінках квартісіті

Черняк О.І., Васильченко І.І.
Київський національний університет імені Тараса Шевченка

Автори аналізують робастні оцінки сусіднього відтинання, що оперують з нижніми порядковими статистиками логарифмічних дохідностей акцій. Проведені симуляції демонструють додаткову ефективність цих оцінок квартісіті та їх підвищену стійкість до стрибків.

Ключові слова: ціна активу, інтегрована волатильність, інтегрована квартісіті, високочастотні данні, ринковий мікроструктурний шум.

Применение нижних порядковых статистик в робастных отсекающих оценках квартисити

Черняк А.И., Васильченко И.И.
Киевский национальный университет имени Тараса Шевченко

Авторы анализируют робастные оценки соседнего отсечения, которые оперируют с нижними порядковыми статистиками логарифмических доходностей акций. Проведенное моделирование демонстрирует дополнительную эффективность этих оценок квартисити и их повышенную устойчивость к прыжкам.

Ключевые слова: цена актива, интегрированная волатильность, интегрированная квартисити, высокочастотные данные, рыночный микроструктурный шум.