Journal of Economic Theory and Econometrics, Vol. 28, No. 1, Mar. 2017, 21–48

Data-Based Ranking of Integrated Variance Estimators Across Size Deciles

Benjamin Carlston*

Abstract In recent years, there has been an explosion of research on the volatility of stock returns. As high frequency stock price data became more readily available, there have been many proposed estimators of integrated variance which attempt to take advantage of the informational gains of high-frequency data while minimizing any potential biases that arise from sampling at such a fine scale. These estimators rely on various assumptions about the price process which can make them difficult to compare theoretically. Relying on the methods of Patton (2011a), this paper analyzes the performance of five different classes of integrated variance estimators when applied to various stocks of differing market capitalization in an attempt to discover the circumstances under which one estimator should be chosen over another.

Keywords Integrated Volatility, Forecast Evaluation

JEL Classification C52, G14, G17

*Eberhardt School of Business, University of the Pacific, 3601 Pacific Avenue, Stockton CA 95211, USA. bcarlston@pacific.edu

Received January 22, 2017, Revised March 24, 2017, Accepted March 25, 2017

1. INTRODUCTION

While early work in realized measures such as Merton (1980) and Zhou (1996) recognized the benefits of utilizing higher frequency data to measure variability over a longer period, only recently has high-frequency intraday price data become available. Subsequently, over the past several years with the increased availability of high-frequency stock data, there has been a strong research focus on the best ways to exploit this increase in information. One specific use of high-frequency data on which there has been a strong focus is that of measuring price volatility, or quadratic variance. Realized variance (RV), or the sum of squared intraday returns (see Andersen, Bollerslev, Diebold, and Labys, 2001; Barndorff-Nielsen and Shephard, 2002), has been the launching pad for many other estimators of quadratic variance that utilize high-frequency returns data. Many of these have sought to reduce microstructure noise, isolate the continuous component of volatility, reduce finite sample bias, or otherwise improve upon our ability to measure the variation of asset prices[1]. These different estimators are often based on different assumptions about the price process. Additionally, one may be based upon sampling in calendar time while another utilizes tick time sampling. These, as well as other tractability issues, often prohibit the theoretical asymptotic comparison of the various estimators.

One shortcoming of the realized variance estimator is that it is only a measure of the *total* variation of the price process, or quadratic variance. For years, research in finance has been based on a continuous price process, however, recently it has become clear that price processes are better represented as continuous brownian motion with jumps[2]. In many fields, e.g. risk management, options pricing and volatility forecasting, it can be useful to obtain an estimate

[1]See Barndorff-Nielsen and Shephard (2004), Mancini (2009), Lee and Mykland (2008), Andsersen, Dobrev, and Schaumburg (2008, 2009). Additional estimators that use subsampling techniques include Zhang (2006), Zhang, Mykland, and Ait-Sahalia (2005). For pre-averaging methods see Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Podolskij and Vetter (2009). Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) develop kernel-based autocovariance adjustments to reduce the effects of noise.

[2]See, for example, Andersen, Benzoni, and Lund (2002), Bates (2000), Chan and Maheu (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Eraker (2004), and Eraker, Johannes and Polson (2003).

of the continuous component of quadratic variance, or integrated variance (IV), without including the variation caused by jumps. The need to isolate and estimate the integrated variance in the presence of possible jumps has led to the development of various estimators which attempt to exclude jump variation[3] as well as tests to determine if and when jumps may have occurred in the price data[4]. With a variety of possible IV estimators available it is useful to obtain a better understanding of which measure to use when investigating a specific empirical question. Since it is often difficult to theoretically compare the different estimators, we must rely on empirical methods to determine the most appropriate estimator for any given asset.

Recently, Patton (2011a) developed an empirical method that allows for the ranking of various estimators when the true underlying process is unobserved, as is the case in the volatility of asset prices. In his implementation of the empirical ranking using high-frequency returns for IBM, Patton (2011a) finds that for the simple RV estimator it is optimal to use a sampling frequency between 15 seconds and 5 minutes. This technique has also been used by Patton and Sheppard (2009) who rank various estimators of quadratic variance and find it is often optimal to use a combination of estimators as opposed to simply choosing a single estimator. This paper utilizes similar techniques to examine the performance of IV estimators for stocks across different size deciles. We chose to examine stocks of differing market capitalizations because many studies focus solely on large-cap stocks which may behave quite differently than stocks of smaller market capitalizations. Sorting the stocks based on market capitalization is a simple way of separating stocks into groups with different trading frequencies.

Studies that estimate the integrated variance for specific assets often focus their attention on stocks with the largest market capitalization[5]. We chose to look

[3]See Barndorff-Nielsen and Shephard (2004), Mancini (2009), Andersen, Dobrev, and Schaumburg (2008, 2012), Christensen, Oomen, and Podolskij (2010), and Ait-Sahalia and Jacod (2010)

[4]See Barndorff-Nielsen and Shephard (2004, 2006), Huang and Tauchen (2005), Andersen, Bollerslev, and Diebold (2007), and Ait-Sahalia and Jacod (2010) among others.

[5]For example, Andersen, Dobrev, and Schaumburg (2012) and Christensen, Oomen, and Podolskij (2010). Additionally, in the ranking procedures of Patton (2011a) and Patton and Sheppard (2009) only data for IBM was examined.

over a wide range of stocks from different size deciles in order to explore how the various estimators perform on stocks with varying liquidity. Specifically, this study will be conducted using 30 stocks from the NYSE. Once all of the stocks were sorted into their size decile, the top ten stocks from the tenth, sixth, and second deciles (where the tenth decile contained the largest stocks on the NYSE) were selected for the study. Such a wide array of stocks will allow for the examination of the relative performance of the estimators on assets of varying size and liquidity. This paper looks to expand upon the findings of Patton (2011a) and Patton and Sheppard (2009) by comparing various IV estimators over a wide range of sampling frequencies and stocks.

The rest of the paper is organized as follows: Section 2 describes the empirical ranking method of Patton (2011a) in further detail. Section 3 discusses the various IV estimators in more detail. A detailed description of the data, including stock selection, cleaning methods, and summary statistics is included in Section 4. Section 5 presents the results of both pairwise comparisons as well as tests for the best estimator among a large set of possibilities. Section 6 concludes.

2. RANKING METHOD

One main difficulty in comparing measures of integrated variance is that the true IV is unobservable. This difficulty, coupled with the abundance of possible sampling frequencies and estimators, has led to the need to better understand which choice of sampling frequency and tuning parameters should be used in any given empirical analysis. The empirical techniques presented in Patton (2011a) allow for the comparison of different realized measures even when the true underlying process is unobservable. By proving various moment and distributional conditions, he is able to appeal to existing volatility forecasting literature in order to compare the various RV estimators.

In addition to being able to circumvent problems arising from measuring the accuracy of estimators to an unobservable target, these data-based techniques also allow for the correlation between microstructure noise and the price process and are straightforward to implement (in part because there is no need to calculate integrated quarticity or the variance of the noise process). However, it is necessary to have a proxy of the true process that is conditionally unbiased in finite samples. Simulation results will be presented in Section 4 to justify the choice of proxy.

Let θ_t denote the latent IV process which we are interested in measuring as accurately as possible by choosing any one of a possible k realized measures, $X_{i,t}$ for $i = 1,2,...,k$. In order to determine whether one estimator is expected to be more accurate than another, we are interested in measuring

$$E[\Delta L(\theta_t, \mathbf{X}_t) \equiv E[L(\theta_t, X_{i,t})] - E[L(\theta_t, X_{j,t})] \quad \text{for } i \neq j \tag{1}$$

where $L(\cdot,\cdot)$ is any predetermined distance measure. The work of Patton (2011a) provides a way in which this value can be accurately estimated even though the process θ_t is unobservable.

In order to begin the ranking process, one needs to choose the distance measure to be used when determining the most accurate IV estimator. Any robust pseudo-distance measure (see Patton, 2011b) of the form

$$L(\theta, X) = \tilde{C}(X) - \tilde{C}(\theta) + C(X)(\theta - X) \tag{2}$$

where $\tilde{C}(X)$ is the anti-derivative of the decreasing, twice differentiable function C(X) can be chosen when using this data ranking technique. While Patton (2011b) and Patton and Sheppard (2009), consider both the QLIKE and MSE loss functions, MSE was chosen as the loss function in this paper's ranking of IV estimators. The motivation for using MSE is that the QLIKE function doesn't readily accommodates estimates of zero and it is frequently the case, especially with the small stocks in the early half of the sample, that estimates of integrated variance will be 0 in our sample. The MSE loss function can easily manage zero values and there is no need to manipulate the data and replace the zero estimates (with an average/minimum value), potentially complicating inference. Additionally, MSE is a generally accepted and widely used distance measure.

The specific functional forms of MSE is

$$\text{MSE} \qquad L(\theta, X) = (\theta - X)^2 \tag{3}$$

Since θ_t is unobservable, we must use a proxy, $\tilde{\theta}_t$, to consistently estimate the difference in average accuracy of the different estimators. In order to obtain accurate results, this proxy must be conditionally unbiased for a finite number of observations N. Once a finitely unbiased proxy has been selected, Patton (2011a) shows that using an average of leads of the proxy, $Y_t = \sum_{j=1}^{J} \tilde{\theta}_t$, it is possible to obtain estimates of $E[\Delta L(\theta_t, \mathbf{X}_t)]$. These results do require one of two assumptions about the true process θ_t; it must either be a random walk or an AR(p) process. In this study, we hold to the assumption that the true integrated variance process follows a random walk. The random walk assumption is supported empirically in work by Andersen, Bollerslev, and Diebold (2007) and Hansen and Lunde (2014). Further support comes from the results of Patton (2011a) and Patton and Sheppard (2009) where there was little difference between the random walk case and the AR(p) case. If the random walk assumption holds, then

$$\mathrm{E}[\Delta L(\theta_t, X_t)] = \mathrm{E}[\Delta L(Y_t, X_t)] \tag{4}$$

With a couple of additional technical assumptions, then it can be shown that

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \Delta L(Y_t, X_t) - E[\Delta L(\theta_t, X_t)]\right) \xrightarrow{d} N(0, \Omega_1) \tag{5}$$

as $T \to \infty$. Additionally, the conditions that allow for the use of the stationary bootstrap will be met (see Propostion 2 of Patton (2011b) for the proof of (4) and (5) and the proof that the stationary bootstrap may be employed). These results allow for testing multiple estimators using the methods of Hansen, Lunde, and Nason (2011) and Giacomini and White (2006) which will be the primary testing methods we consider. White (2000) and Romano and Wolf (2005) also developed techniques to test multiple models simultaneously.

3. IV ESTIMATORS

Before we continue with a presentation of the results, further discussion about each estimator under consideration is necessary. For notational purposes, let us consider a jump-diffusive price process where the logarithmic price, P_t, is observed at $N+1$ discrete points throughout the trading day.

The first estimator of daily integrated variance is the Bipower Variation (BV) estimator of Barndorff-Nielsen and Shephard (2004). One benefit to this estimator is that it is a consistent estimator of IV under the assumption of no market microstructure noise but otherwise general conditions. The specific form that it takes is similar to that of the RV estimator, but instead of using squared returns it uses the product of neighboring absolute returns. It has the following functional form

$$BV_N = \frac{\pi}{2}\frac{N}{N-1}\sum_{i=1}^{N-1}|\Delta P_t||\Delta P_{t+1}| \qquad (6)$$

where $\Delta P_t = P_t - P_{t-1}$. The intuition behind this estimator is that because there are finitely many jumps during a trading day, as the sampling frequency goes to zero (or as $N \to \infty$), there will not be two jumps in any subsequent returns. Additionally, the diffusive return will go to zero. For illustrative purposes, consider two returns ΔP_t and ΔP_{t+1} where the first return contains a jump and the second does not. The returns when there is no jump, or ΔP_{t+1} in this case, will go to zero as the sampling frequency gets small. This will cause the product $|\Delta P_t||\Delta P_{t+1}|$ to go to zero and thus eliminate any variation in returns due to the jump component of the price process. One shortcoming of this estimator, however, is that it is biased in finite samples. This bias arises from the fact that in finite samples the diffusive return $|\Delta P_{t+1}|$ does not equal zero and thus the jump return $|\Delta P_t|$ is not completely cancelled out. This drives up the estimated value of IV and creates an upward bias.

A recent attempt to extend upon the thought underlying the BV estimator while minimizing any potential finite sample bias are the MinRV and MedRV estimators of Andersen, Dobrev, and Schaumburg (2009). These estimators are shown to be consistent IV estimators and are more robust to finite jumps in finite samples. The functional forms of the two estimators are as follows:

$$\text{MinRV}_N = \frac{\pi}{\pi-2}\frac{N}{N-1}\sum_{i=1}^{N-1}\min(|\Delta P_t|,|\Delta P_{t+1}|)^2 \quad (7)$$

$$\text{MedRV}_N = \frac{\pi}{6-4\sqrt{3}+\pi}\frac{N}{N-2}\sum_{i=1}^{N-1}\text{med}(|\Delta P_{t-1}|,|\Delta P_t|,|\Delta P_{t+1}|)^2 \quad (8)$$

The intuition behind these estimators is similar to that of bipower variation; these estimators seek to eliminate the variation in returns due to jumps by taking either the minimum or the median return over a small block size of two or three returns. The jump robustness of the MedRV and MinRV estimators of Andersen, Dobrev, and Schaumburg (2009) in relation to bipower variation is that the variation due to the jump return will be completely eliminated by the minimum or median operator. These estimators do rely on the assumption of a constant variance over each block of returns, and by using block sizes of only two or three returns, the MinRV and MedRV estimators are less vulnerable to bias due to intraday volatility patterns.

The fourth IV estimator we will consider in this analysis is the truncation-type estimator, see Mancini (2009), Jacod (2008), and Ait-Sahalia and Jacod (2009).

$$TRV_c = \sum_{i=1}^{N}(\Delta P_t)^2 1_{\{|\Delta P_t|<cN^{-\bar{\omega}}\}} \quad (9)$$

where $\bar{\omega} \in (0, 0.5)$ and c is the truncation parameter or threshold for truncating jumps. In the specific implementation of the estimator, we follow Christensen, Oomen, and Podolskij (2010) and set $\bar{\omega} = 0.47$.

This form of estimator relies on filtering out returns that exceed a threshold chosen by the researcher. Through these means, the large returns that are the result of jumps in the price process will be eliminated and ideally the only returns that will be considered are diffusive returns. One potential difficulty with truncation estimators is in the choice of threshold. By selecting a different truncation thresholds for the comparison, we hope to gain further insight into the extent to which the choice of the truncation parameter can effect the results.

4. IMPLEMENTATION

4.1. DATA

For the purposes of a more thorough ranking, as well as to discover if any estimators may generally perform better for one class of stock over another, the empirical rankings are done over a collection of 30 different stocks. The stocks were chosen based on their market capitalization for the year 2007. All NYSE stocks were sorted into ten size deciles. The ten largest stocks from the tenth (large cap), sixth (mid cap), and second (small cap) size deciles (with the tenth being the largest stocks and the first being the smallest) were then chosen as our sample. Stocks from a range of size deciles were chosen to examine how liquidity and market capitalization may affect the relative performance of the estimators. It may well be the case that a specific estimator outperforms in highly liquid stocks while another shines when used on less liquid ones. Two additional requirements were that the available data for the stock dates to at least 2002, and that the ticker symbol corresponded to an actual company (e.g. mutual funds were ignored). The thirty stocks that were chosen, the total sample size for each stock, as well as some descriptive statistics are presented in Table 1.

The TAQ trades data was then cleaned in a method very similar to that outlined in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). Observations were filtered based on time stamp so that only those occurring between 9:30am and 4:00pm were included. Any zero price was removed and only trades that occurred on the NYSE were included in the final dataset. All entries that had been corrected (or had the variable CORR $\neq 0$) as well as any observation with an abnormal *sale condition* were removed (specifically trades where COND has a letter code, except for “E” and “F” – please refer to the Daily TAQ Quick Reference Card for additional details about the sale condition classifications). If there were multiple prices for a single timestamp, then the median price was used for that second. The last recorded price was used for any second on which there was no recorded trade. From this dataset, any desired sampling frequency is readily obtained. Additionally, returns due to stock splits and overnight returns are set to zero.

4.2. SAMPLING FREQUENCIES AND TUNING PARAMETERS

For a few of the estimators, the only parameter that one must choose is the sampling frequency. When choosing the sampling frequency, there is always a trade-off between added information and possible noise contamination. The sampling frequencies under consideration are 10 and 30 seconds and 1, 2, 5, 10, 15, 20, and 30 minutes. When varying the sampling frequency, we set the truncation parameter as $c = 6\sqrt{IV}$, where IV is the daily integrated variance estimated using BV (see Christensen, Oomen, and Podolskij (2010) for a similar choice of truncation level). When changing the truncation parameter, c, we fix the sampling to be at 2 minutes.

4.3. CHOICE OF PROXY

As was previously mentioned, when implementing the data-based ranking technique of Patton (2011a), it is necessary to have a proxy of the latent variable that is unbiased in finite samples. If an unbiased proxy were unavailable, then the equality in Equation (4) would not hold. In the case of quadratic variance, 5-minute RV is widely accepted as an unbiased proxy in finite samples of quadratic variance. In the case of integrated variance, the choice of proxy is not so obvious. The final decision on the choice of proxy was based on the simulation results of both Andersen, Dobrev, and Schaumburg (2012) and Christensen, Oomen, and Podolskij (2009).

In Table 2, we reproduce the simulation results of Andersen, Dobrev, and Schaumburg (2012). They consider six different models for the price process and they compare the finite sample properites of the RV, BV, MinRV, MedRV, Tripower variation (TV) and QRV estimators. In the interest of space, we will refer the interested reader to Section 2.5 of their paper for the detailed description of their six models for stock prices. Examining Table 2, we see that at the 1 minute sampling frequency, the relative bias for MedRV is between 1.026 and 0.990. This tells us that across all of the six models under consideration, the MedRV estimate of the true integrated variance is no more than 2.6% larger than and 1% smaller than the true value on average. Furthermore, Christensen, Oomen, and Podolskij (2009) find that for 100,000 simulations from 10 different models, the expected relative bias for MedRV is 1 for 6 of the models under consideration. It is no more than 1.03 for 3 of the remaining 4 models. The only model for which MedRV is largely biased is when the price process follows a brownian motion with an outlier (represented by 2 consecutive jumps of opposite signs – a violation of an underlying assumption of MedRV). From these simulation results, the proxy used for the ranking technique was chosen to be the MedRV estimator. It was sampled at the 2 minute frequency following the advice of Andersen, Dobrev, and Schaumburg (2012).

5. RESULTS

5.1. TRUNCATED REALIZED VARIANCE

Before examining the formal statistical tests, it is worthwhile to visually examine the performance of the truncation estimator as the truncation parameter changes. By plotting the average loss as a function of the truncation parameter we are able to identify the threshold that will provided the lowest average loss for a specific stock. Similar to volatility signature plots, see Andersen, Bollerslev, Diebold and Labys (2000), Figures 1 - 3 provide a visual means of selecting the optimal threshold parameter. In each of these plots, the sampling frequency is fixed at 2 minutes and the truncation parameter is allowed to vary. The graphs in Figures 1 - 3 plot the average MSE loss as a function of the truncation parameter (number of standard deviations above which a return is considered a jump). The plot with the crosses is using the 2 minute MedRV as the proxy while the dotted line is the average MSE loss using 2 minute BV as the loss function. The value at which the plots obtain a minimum is the value of the threshold that will yield the lowest average MSE loss for any given stock. In almost all of the individual plots for the 30 stocks we see the minimum occurring at either 2 or 3. This result is robust across each of the size deciles under consideration, and indicates that the result is quite robust. While not included in the plots in order to keep them as clean as possible, this result is robust to also using 2 minute MinRV as the proxy.

5.2. GIACOMINI AND WHITE (2006) TESTS

In this section, we implement Giacomini and White (2006) tests on whether two competing IV estimators have equal average accuracy conditional on the same information set, $\mathscr{F}_{t-1}$. The specific null hypothesis in question is

$$H_0: \qquad E[L(\theta_t, X_{t,i})|\mathscr{F}_{t-1}] - E[L(\theta_t, X_{t,j})|\mathscr{F}_{t-1}] = 0 \tag{10}$$

The results of Patton (2011a) allow this test to be run using a simple regression of the following form

$$L(\theta_t, X_{t,i}) - L(\theta_t, X_{t,j}) = \beta_0 + e_t \tag{11}$$

These regressions were run for each pairing of estimators based on the 2 minute sampling frequency. As discussed previously, the proxy used for the regressions is the MedRV estimator based on data sampled at the five minute frequency. Table 3 and Table 4 report the percentage of times an estimator significantly outperforms or underperforms in pairwise tests.

One interesting trend that emerges for the 10 stocks in the largest decile is that the 2 minute RV estimator is rarely out-performed by the other estimators. However, this trend does not hold for the stocks in the middle and lower deciles. This suggests that for highly liquid securities the impact of jumps is so limited that RV and BV provide statistically similar estimates of integrated variance. For a large portion of the stocks in the middle size decile, 2 minute RV is significantly worse at estimating the integrated variance than each of the estimators specifically developed for IV estimation. The poor performance of RV in pairwise comparisons continues into the lower decile as well. This discrepancy between stocks in the largest size decile and those in the lower deciles is most likely explained by the discrepancy in the average number of trades between large and small stocks. As the summary statistics in Table 1 indicate, the largest stocks trade at a much higher frequency than the smaller stocks. Increased trading activity will likely result in the presence small, frequent jumps in large stocks. The presence of small, frequent jumps would lead to a smaller difference between the RV estimator and any of the IV estimators since it would be more difficult for the later to distinguish jump returns from non-jump returns. Small, frequent jumps in large cap stocks would explain why RV is significantly outperformed at a much higher frequency for the Mid and Small cap stocks in Table 4.

In Table 3, the TRV_3 estimator significantly outperforms the other estimators for the 2 minute sampling frequency. Furthermore, as shown in Table 4, TRV_3 is never significantly worse than the other estimators. This suggests that the TRV_3 estimator is a better choice than the others as it is never significantly worse than the other estimators and it often significantly outperforms the other estimators. The data also suggests that the easily implemented BV estimator performs quite well across a variety of stocks and the additional refinements of the more sophisticated estimators may not be as helpful when applied to actual stock data.

5.3. SET OF BEST ESTIMATORS

While pairwise comparisons can be informative, it can be difficult to test every possible specification of the estimators in a pairwise manner. The model confidence set (MCS) procedure of Hansen, Lunde, and Nason (2011) is implemented in order to test which estimators, sampling frequency, and tuning parameters are significantly better than the others. While the MCS may not yield a single estimator as being the best, it will result in a group of estimators which will contain the best estimator with a specific level of confidence. This procedure will allow the data to determine which estimator(s) most accurately estimate IV. If an estimator appears in the MCS for a stock it implies that the estimator is in the group of "best" estimators for the stock.

There are a total of 62 different estimators under consideration. For each estimator (BV, MedRV, MinRV, and TRV_6) we have 9 possible sampling frequencies ranging from 10 seconds to 30 minutes. Additionally, we fix the sampling frequency of TRV at 2 minutes and then vary the truncation parameter to take on one of the following values 1, 2, 3, 4, or 5 standard deviations.

The results are summarized in Tables 5 - 7. The tables are divided into the three size groups that we are considering. The model confidence set is determined for the IV estimators for each stock, and the tables indicate the number of times that a specific estimator was in the model confidence set, or set of best estimators. For each group of stocks, there were two or three stocks which included all of the estimators in the model confidence set. Another noticeable trend was that the TRV with a truncation of 2 or 3 standard deviations was almost always selected as one of the best estimators for integrated variance regardless of the size decile.

In section 5.1, we found that the truncation estimator had the lowest average MSE loss when the truncation parameter was set to either 2 or 3 standard deviations. In Table 3, we found that the TRV estimator using a truncation of 3 standard deviations was significantly outperformed the other estimators in pairwise tests roughly half of the time. Table 4 reports that TRB with a truncation of 3 standard deviations is never significantly outperformed in pairwise tests. The presence of this same estimator in the model confidence set of virtually every stock (28 out of 30) further confirms the performance of TRV using a truncation of 2 or 3 standard deviations. Also, the pairwise Giacomini and White (2006) tests indicated that only TRV was able to outperform 2 minute BV across all of the stocks.

6. CONCLUSION

The recent availability of high frequency asset returns has led to the development of a numerous variety of volatility measures. With the current number of estimators and no feasible way to compare them theoretically, researchers are always confronted with the question of which estimator will provide the most accurate IV measure for a given asset. In addition to choosing the estimator and any subsequent tuning parameters, one must also select a sampling frequency for the data. This paper applies the data-based ranking technique of Patton (2011a) to estimators of integrated variance. It focuses on stocks over a wide range of sizes and, as a result, also liquidities. Stock prices of small companies often behave differently than large companies. Since empirical examples are often estimated using only a few large companies, we wanted to determine if the performance of various estimators would extend to stocks with different liquidity characteristics. The results indicate that if one is interested in measuring the integrated variance, then TRV with a threshold of 2 or 3 standard deviations not only has the lowest MSE loss for the truncation parameters but is almost always included in the set of "best" estimators of integrated variance for a stock regardless of market size and trading volume. In pairwise Giacomini and White (2006) tests, TRV_3 significantly outperforms the other estimators (sampled at a 2 minute frequency) for the majority of the 30 stocks included in the analysis. Also, TRV_3 is found to never significantly underperform in the pairwise tests. When implementing the model confidence sets, we find that the TRV estimator with a truncation parameter of 2-3 standard deviations is almost always in the set of "best" estimators for the stocks (it appears for either 9 or 10 out of 10 stocks in each category). While there doesnt appear to be a one-size-fits-all solution for estimating integrated variance, the truncated estimator with a threshold of 2-3 standard deviations does well in a variety of situations. Additionally, this paper provides a quick, visual method for determining the optimal truncation threshold for any given asset, similar to volatility signature plots.

REFERENCES

Ait-Sahalia, Y., and J. Jacod (2011). Testing Whether Jumps Have Finite or Infinite Activity, The Annals of Statistics 39, 1689-1719.

Ait-Sahalia, Y., and J. Jacod (2012). Analyzing the Spectrum of Asset Returns: Jumps and Volatility Components in High Frequency Data, Journal of Economic Literature 50, 1007-1050.

Andersen, T.G., L. Benzoni, and J. Lund (2002). Estimating jump-diffusions for equity returns, Journal of Finance 57, 1239-1284.

Andersen, T.G., T. Bollerslev, and F. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, Review of Economics and Statistics 89, 701-720.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2000). Great Realizations, Risk 13, 105-108.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility, Journal of the American Statistical Association 96, 42-55.

Andersen, T., D. Dobrev, and E. Schaumburg (2008). Duration-Based Volatility Estimation, Working Paper.

Andersen, T., D. Dobrev, and E. Scaumburg (2012). Jump-Robust Volatility Estimation using Nearest Neighbor Truncation, Journal of Econometrics 169, 75-93.

Barndorff-Nielsen, O.E., and N. Sheppard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society, Series B 64, 253-280.

Barndorff-Nielsen, O.E., and N. Sheppard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps, Journal of Financial Econometrics 2, 1-37.

Barndorff-Nielsen, O.E., and N. Sheppard (2006). Econometrics of testing for jumps in financial economics using bipower variation, Journal of Financial Econometrics 4, 1-30.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Sheppard (2009). Realised Kernels in Practice: Trades and Quotes, The Econometrics Journal 12, C1-C32.

Bates, D.S. (2000). Post-‘87 Crash fears in the S&P500 futures option markets, Journal of Econometrics 94, 181-238.

Chan, W.H. and J.M. Maheu (2002). Conditional jump dynamics in stock market returns, Journal of Business and Economic Statistics 20, 377-389.

Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics, Journal of Econometrics 116, 225-257.

Christensen, K., R. Oomen, and M. Podolskij (2010). Realised Quantile-Based Estimation of the Integrated Variance, Journal of Econometrics 159,74-96.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices, Journal of Finance 59, 1367-1403.

Eraker, B., M.S. Johannes, and N.G. Polson (2003). The impact of jumps in volatility, Journal of Finance 58, 1269-1300.

Giacomini, R. and H. White (2006). Tests of Conditional Predictive Ability, Econometrica 74, 1545-1578.

Hansen, P.R. and A. Lunde (2014). Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error, Econometric Theory 30, 60-93.

Hansen, P.R., Lunde, A., and J.M. Nason (2011). The Model Confidence Set, Econometrica 79, 453-497.

Huang, X. and G. Tauchen (2005). The Relative Contribution of Jumps to Total Price Variance, Journal of Financial Econometrics 3, 456-499.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales, Stochastic Processes and their Applications 118, 517-559.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: The pre-averaging approach, Stochastic Processes and their Applications 119, 2249-2276.

Lee, S.S. and P.A. Mykland (2008). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics, The Review of Financial Studies 21, 2534-2563.

Mancini, C. (2009). Non-Parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps, Scandinavian Journal of Statistics 36, 270-296.

Merton, R.C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation, Journal of Financial Economics 8, 323-361.

Patton, A.J. (2011a). Volatility Forecast Comparison using Imperfect Volatility Proxies, Journal of Econometrics 160, 246-256.

Patton, A.J. (2011b). Data-Based Ranking of Realised Volatility Estimators, Journal of Econometrics 161, 284-303.

Patton, A.J. and K. Sheppard (2009). Optimal combinations of realised volatility estimators, International Journal of Forecasting 25, 218-238.

Podolskij, M. and M. Vetter (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15, 634-658.

Romano, J.P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping, Econometrica 73, 1237-1282.

White, H. (2000). A Reality Check for Data Snooping, Econometrica 68, 1097-1126.

Zhang, L. (2006). Efficient Estimation of Stochastic Volatility using Noisy Observations: A Multi-Scale Approach, Bernoulli 12, 1019-1043.

Zhang, L., P.A. Mykland, and Y. Ait-Sahalia (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, Journal of the American Statistical Association 100, 1394-1411.

Zhou, B. (1996). High-Frequency Data and Volatility in Foreign-Exchange Rates, Journal of Business and Economic Statistics 14, 45-52.

APPENDIX

Table 1: Summary Statistics for the 30 selected stocks

Ticker	Year Range	Average # trades per day	Average # trades per day 1993-2002	Average # of trades per day 2003-2012
AIG	1993-2012	3,942.3	927	7,092
BAC	1993-2012	6,524.7	1,119	12,171
GE	1993-2012	6,341.8	2,071	10,803
IBM	1993-2012	4,689.8	1,737	7,774
JNJ	1993-2012	4,441.3	1,113	7,918
MO	1993-2012	3,544.4	1,323	5,864
PFE	1993-2012	4,889.3	1,672	8,250
PG	1993-2012	4,264.5	1,058	7,614
T	1993-2012	4,463.0	1,394	7,660
WMT	1993-2012	4,937.6	1,225	8,816
AKS	1995-2012	1,936.7	146	3,384
ATW	1997-2012	944.3	132	1,402
BPL	1993-2012	131.37	22	246
CTV	1997-2011	1,176.1	209	1,567
CW	1993-2012	299.7	21	571
DRQ	1997-2012	622.6	59	927
HXL	1993-2012	633.5	33	1,259
KEX	1996-2012	478.9	49	758
RBA	1998-2012	312.12	15	459
VMI	2002-2012	487.8	95	501
CCC	1993-2012	380.9	39	738
CIA	2002-2012	144	44	153
CMO	1993-2012	327.8	85	581
CRN	1998-2010	156.2	18	186
CV	1993-2012	82.23	22	144
CYD	1994-2012	208.4	8	360
DCO	1996-2012	96.4	35	136
HGR	1998-2012	263.7	83	347
MIG	1995-2012	189.2	12	317
RHB	1998-2012	316.4	96	383

Table 2: Simulation Results from Andersen, Dobrev, and Schaumburg (2012)

Relative Bias for 12 sec. sampling frequency

	RV	BV	TV	QRV	MinRV	MedRV
Model 1:BM	1.000	1.000	1.000	0.999	1.000	1.000
Model 2: SV-U	0.999	0.999	0.998	0.971	0.999	0.998
Model 3: BM + Sparcity	0.999	0.974	0.965	0.969	0.955	0.962
Model 4: BM + 1 Jump	1.244	1.018	1.009	1.001	1.002	1.002
Model 5: BM + 4 Jumps	1.250	1.035	1.020	1.006	1.006	1.006
Model 6: BM + Noise	1.078	1.079	1.079	1.078	1.079	1.079

Relative Bias for 1 min. sampling frequency

	RV	BV	TV	QRV	MinRV	MedRV
Model 1:BM	1.001	1.000	1.000	1.000	1.000	1.000
Model 2: SV-U	0.995	0.993	0.990	0.969	0.993	0.991
Model 3: BM + Sparcity	1.001	0.993	0.991	0.988	0.988	0.990
Model 4: BM + 1 Jump	1.242	1.038	1.023	1.006	1.007	1.007
Model 5: BM + 4 Jumps	1.250	1.073	1.051	1.026	1.023	1.026
Model 6: BM + Noise	1.003	1.004	1.003	1.003	1.004	1.004

Relative Bias for 5 min. sampling frequency

	RV	BV	TV	QRV	MinRV	MedRV
Model 1:BM	1.001	1.001	1.002	1.001	1.002	1.002
Model 2: SV-U	0.990	0.979	0.968	0.967	0.979	0.969
Model 3: BM + Sparcity	1.002	1.001	1.003	0.998	1.000	1.002
Model 4: BM + 1 Jump	1.241	1.075	1.053	1.031	1.024	1.027
Model 5: BM + 4 Jumps	1.251	1.131	1.107	1.086	1.073	1.082
Model 6: BM + Noise	1.002	1.004	1.004	1.001	1.005	1.002

Table 3: Summary of results from pairwise GW tests using the MSE distance measure at a 2 minute frequency. This table shows the percentage of times a measure of integrated variance significantly outperformed another measure in a pairwise test using 2 minute MedRV as the proxy.

Market Cap.	RV	BV	MinRV	MedRV	TRV_6	TRV_3
Large	0.02	0.16	0.04	0.02	0.24	0.44
Mid	0.00	0.20	0.18	0.16	0.12	0.62
Small	0.00	0.18	0.22	0.28	0.16	0.40

Table 4: Summary of results from pairwise GW tests using the MSE distance measure at a 2 minute frequency. This table shows the percentage of times a measure of integrated variance was significantly worse than another measure in a pairwise test using 2 minute MedRV as the proxy.

Market Cap.	RV	BV	MinRV	MedRV	TRV_6	TRV_3
Large	0.22	0.08	0.26	0.28	0.08	0.00
Mid	0.76	0.10	0.08	0.12	0.24	0.00
Small	0.70	0.10	0.06	0.06	0.28	0.00

Table 5: Model Confidence Set (Hansen, Lunde, and Nason (2011)) Results for 10 large cap stocks. The table reports the number of times that the estimator was found to be one of the best models (in the model confidence set). The top portion of the table uses $c = 6$ for the TRV estimator. The bottom portion of the table presents the results when the truncation parameter varies but the sampling frequency is fixed at 2 minutes.

Frequency	RV	BV	MinRV	MedRV	TRV_6
10s	2	4	3	3	3
30s	2	6	3	3	4
1m	2	8	3	3	5
2m	2	8	6	3	8
5m	3	5	6	6	9
10m	3	7	6	7	7
15m	3	6	7	6	6
20m	3	6	4	6	6
30m	3	4	4	5	5

Truncation parameter, c	2 minute TRV
1	3
2	9
3	9
4	9
5	9

Table 6: Model Confidence Set (Hansen, Lunde, and Nason (2011)) Results for 10 medium cap stocks. The table reports the number of times that the estimator was found to be one of the best models (in the model confidence set). The top portion of the table uses $c = 6$ for the TRV estimator. The bottom portion of the table presents the results when the truncation parameter varies but the sampling frequency is fixed at 2 minutes.

Frequency	RV	BV	MinRV	MedRV	TRV_6
10s	2	6	6	8	8
30s	2	9	9	9	9
1m	2	8	7	9	8
2m	2	8	8	8	5
5m	3	7	8	6	4
10m	3	6	5	6	4
15m	2	5	2	3	3
20m	2	5	2	3	3
30m	2	4	3	4	3

Truncation parameter, c	2 minute TRV
1	4
2	10
3	10
4	8
5	5

Table 7: Model Confidence Set (Hansen, Lunde, and Nason (2011)) Results for 10 small cap stocks. The tables reports the number of times that the estimator was found to be one of the best models (in the model confidence set). The top portion of the table uses $c = 6$ for the TRV estimator. The bottom portion of the table presents the results when the truncation parameter varies but the sampling frequency is fixed at 2 minutes.

Frequency	RV	BV	MinRV	MedRV	TRV_6
10s	3	4	4	5	8
30s	3	4	5	7	8
1m	3	8	7	8	6
2m	3	8	8	7	3
5m	3	6	4	5	3
10m	3	4	4	4	3
15m	3	3	3	3	3
20m	3	3	3	3	3
30m	3	3	3	3	3

Truncation parameter, c	2 minute TRV
1	5
2	10
3	9
4	5
5	3

Figure 1: Average MSE loss for TRV_{2min} where the truncation parameter varies along the x-axis. Large cap stocks. The cross is the 2 minute MedRV proxy and the dashed line is the 2 minute BV (bipower variation) proxy.

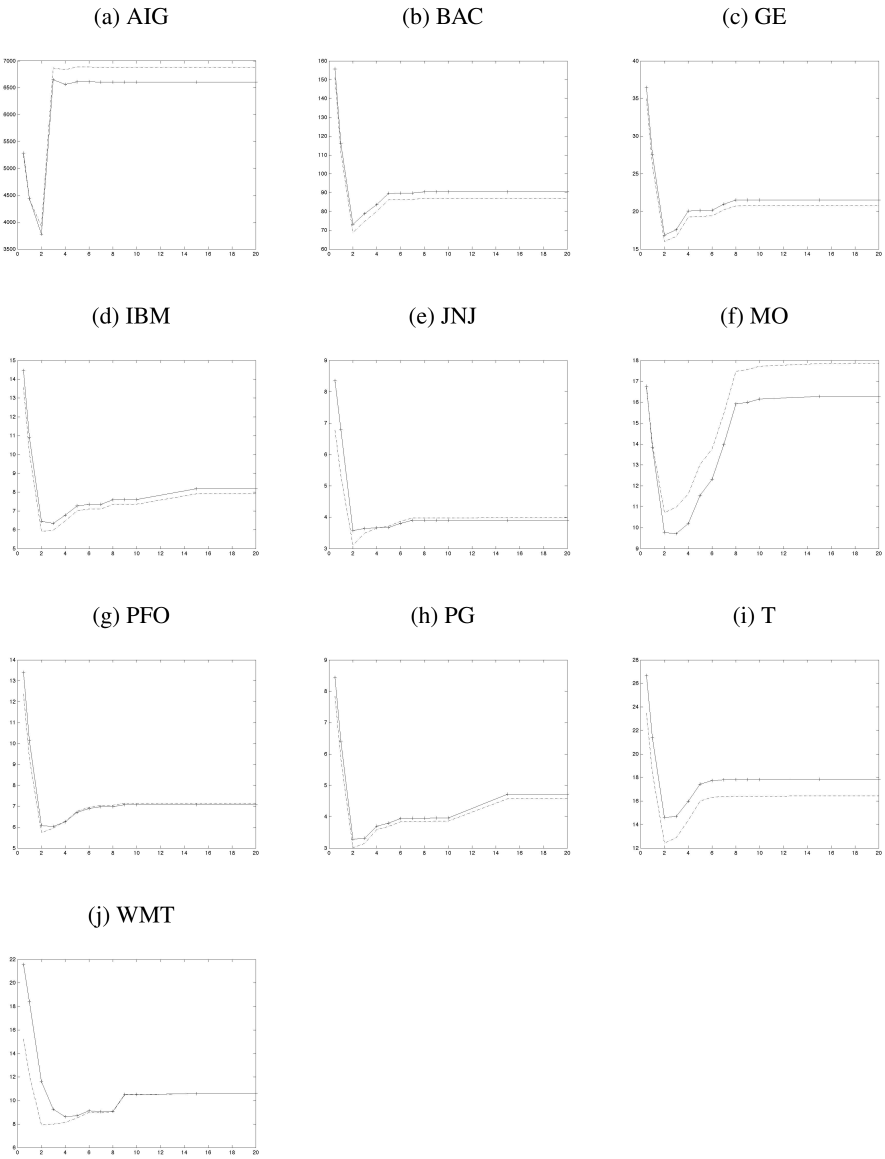


Figure 2: Average MSE loss for TRV_{2min} where the truncation parameter varies along the x-axis. Large cap stocks. The cross is the 2 minute MedRV proxy and the dashed line is the 2 minute BV (bipower variation) proxy.

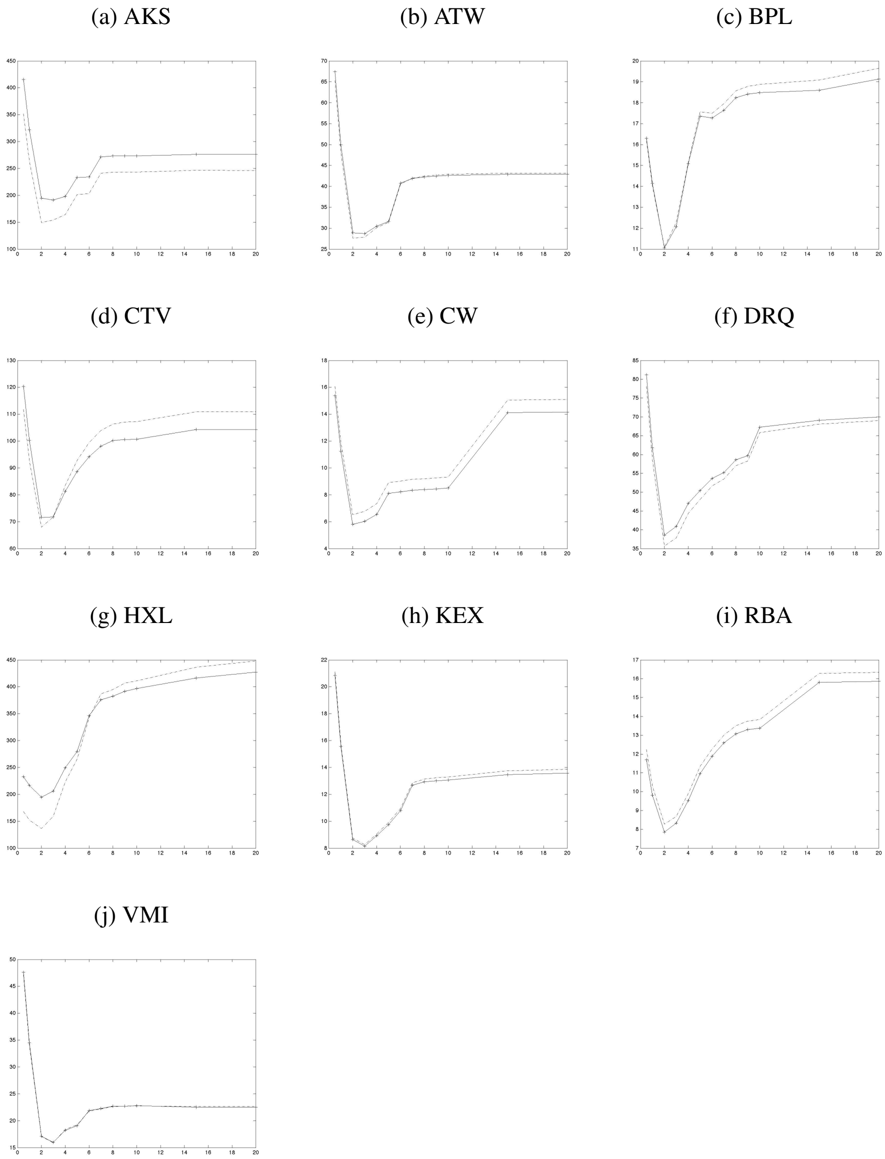
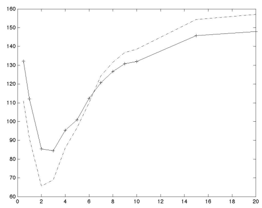
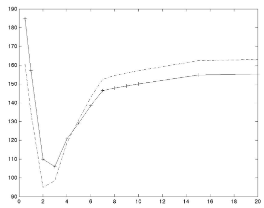


Figure 3: Average MSE loss for TRV_{2min} where the truncation parameter varies along the x-axis. Large cap stocks. The cross is the 2 minute MedRV proxy and the dashed line is the 2 minute BV (bipower variation) proxy.

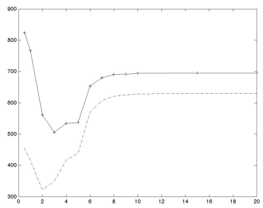
(a) CCC



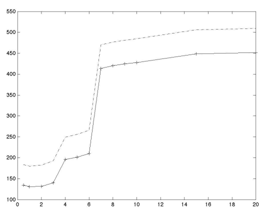
(b) CIA



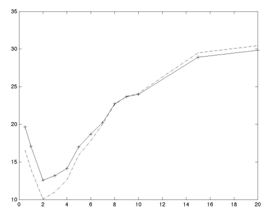
(c) CMO



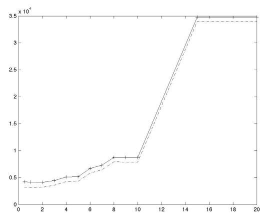
(d) CRN



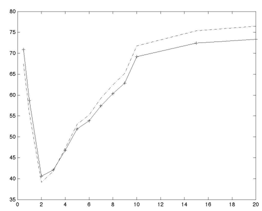
(e) CV



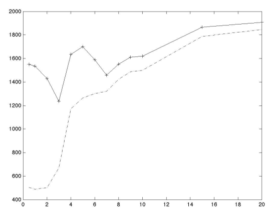
(f) CYD



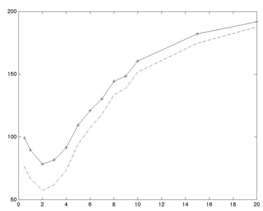
(g) DCO



(h) HGR



(i) MIG



(j) RHB