

Dong, Yingjie; Tse, Yiu-Kuen

Article

Business time sampling scheme with applications to testing semi-martingale hypothesis and estimating integrated volatility

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Dong, Yingjie; Tse, Yiu-Kuen (2017) : Business time sampling scheme with applications to testing semi-martingale hypothesis and estimating integrated volatility, Econometrics, ISSN 2225-1146, MDPI, Basel, Vol. 5, Iss. 4, pp. 1-19, <https://doi.org/10.3390/econometrics5040051>

This Version is available at:

<https://hdl.handle.net/10419/195435>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

Business Time Sampling Scheme with Applications to Testing Semi-Martingale Hypothesis and Estimating Integrated Volatility

Yingjie Dong ^{1,*} and Yiu-Kuen Tse ²

¹ Business School, University of International Business and Economics, 10 Huixin Dongjie, Beijing 100029, China

² School of Economics, Singapore Management University, Singapore 178903, Singapore; yktse@smu.edu.sg

* Correspondence: 02541@uibe.edu.cn; Tel.: +86-152-1004-9697

Academic Editors: Deniz Erdemlioglu, Olivier Scaillet and Kamil Yilmaz

Received: 3 August 2017; Accepted: 17 October 2017; Published: 13 November 2017

Abstract: We propose a new method to implement the Business Time Sampling (BTS) scheme for high-frequency financial data. We compute a time-transformation (TT) function using the intraday integrated volatility estimated by a jump-robust method. The BTS transactions are obtained using the inverse of the TT function. Using our sampled BTS transactions, we test the semi-martingale hypothesis of the stock log-price process and estimate the daily realized volatility. Our method improves the normality approximation of the standardized business-time return distribution. Our Monte Carlo results show that the integrated volatility estimates using our proposed sampling strategy provide smaller root mean-squared error.

Keywords: autoregressive conditional duration model; high-frequency data; integrated volatility; time-transformation function

JEL Codes: C410; G120

1. Introduction

In high-frequency financial data analysis, researchers usually do not use all available data but would select a subgrid of transactions. To choose the subgrid, two issues have to be considered: selecting the sampling scheme and choosing the target average sampling frequency. Three sampling schemes are commonly used in the literature: Calendar Time Sampling (CTS), Tick Time Sampling (TTS) and Business Time Sampling (BTS). Under the CTS scheme, transactions are selected by regularly spaced calendar time, such as every 5 s/min. The TTS scheme selects transactions with regularly spaced number of ticks, e.g., every 5 or 10 ticks. The BTS transactions are often selected to ensure approximately equal volatility for the returns over each interval. Thus, the CTS and TTS schemes are implemented based on explicit criteria (i.e., regular calendar-time length or number of ticks, respectively). In contrast, the BTS scheme depends on the unobserved volatility. As a result, the CTS and TTS schemes have been used widely in the literature, while the BTS scheme is used less frequently.

The BTS scheme possesses some desirable properties for high-frequency financial data analysis. It dates back to [Dacorogna et al. \(1993\)](#), and see also [Zhou \(1998\)](#), [Peters and De Vilder \(2006\)](#), and [Mykland \(2012\)](#). In particular, the BTS scheme yields independently and identically distributed (iid) normal returns for a semimartingale price process even when there is leverage or feedback effect.¹

¹ [Dambis \(1965\)](#) and [Dubins and Schwarz \(1965\)](#) show that a process compiled from a continuous local martingale with equal quadratic-variation increments is a Brownian motion. Leverage effect refers to the asymmetry between equity returns and

In contrast, due to the leverage effect or varying volatility, the calendar-time returns may not be iid normal even if the price process is a continuous local martingale. The assumption of iid Gaussian returns is required for the Gaussian-likelihood approach [Nowman \(1997\)](#) and several widely used integrated volatility estimates, including the multipower variation (MPV) estimate of [Barndorff-Nielsen et al. \(2006\)](#), the quantile realized volatility (QRV) method of [Christensen et al. \(2010\)](#) and the nearest neighbor truncation method of [Andersen et al. \(2012, 2014\)](#). However, in the implementation, researchers often sample transactions under the CTS or TTS scheme. [Andersen et al. \(2007, 2010\)](#) find that the normalized daily and weekly returns sampled in business time accord well with the standard normal distribution. To sample the time points, they use a sequential method to include intraday returns until the cumulative squared returns exceeds the average daily or weekly realized volatility.

One drawback of the method in [Andersen et al. \(2007, 2010\)](#) is that its performance in obtaining returns with approximately equal volatility over each interval deteriorates as the sampling frequency increases. In addition, researchers need to choose a threshold to obtain sampled returns at the target frequency. In this paper, we propose a new method to implement the BTS scheme, which has better performance as the sampling frequency increases and needs no threshold. We estimate the intraday integrated volatility using the jump-robust Tripower Realized Volatility (TRV) method of [Barndorff-Nielsen et al. \(2006\)](#) and calculate a time-transformation (TT) function over the investigated time period. The value of the TT function corresponds to the cumulative increments of the estimated intraday integrated volatility over time and the sampled BTS transactions are obtained using the inverse of the TT function.

We test the semi-martingale hypothesis on the BTS returns of 40 stocks selected from the New York Stock Exchange (NYSE). Our proposed BTS method improves the normality approximation of the empirical standardized return distribution. We further explore the use of the BTS scheme in estimating integrated volatility. First, we consider the Realized Volatility (RV) estimates for daily integrated volatility when the returns are sampled using the BTS, CTS and TTS schemes. Our Monte Carlo simulation shows that the TRV estimates (with and without subsampling) using the BTS returns provide the smallest root mean-squared error (RMSE). Second, we modify the ACD-ICV method of [Tse and Yang \(2012\)](#), making use of the BTS scheme. Our modified ACD-ICV estimator performs better than the Realized Kernel (RK) estimates [Barndorff-Nielsen et al. \(2008\)](#) and the method of [Tse and Yang \(2012\)](#), which uses price-event sampling.

The rest of this paper is as follows. Section 2 outlines our proposed implementation of the BTS scheme. Section 3 reports some empirical results on testing the semi-martingale hypothesis. In Section 4, we outline the estimation methods of daily volatility examined in this paper. We report the results of our Monte Carlo study in Section 5 and draw conclusions in Section 6. The Appendix A provides further details of the jump detection procedure and computation of the BT time-transformation function. Some additional results can be found in the accompanying online supplementary material.

2. Intraday Periodicity and the BTS Scheme

The stylized fact of intraday periodicity is a well known phenomenon in high-frequency financial data analysis. Trading activities are usually high at the beginning and close of the trading day and low around lunch time. Modifications are needed to get rid of this pattern before data are fitted into high-frequency models. To adjust for the intraday periodicity of transaction activity, [Tse and Dong \(2014\)](#) use a time-transformation function computed by pooling all transactions over all trading days in the sample. The TT function transforms all observed transactions from the calendar time to a transformed time for which transactions are evenly observed. However, the time-transformation

volatility. That is, large negative returns tend to be associated with higher future volatility than positive returns of the same magnitude. Feedback effect refers to the case when the volatility function is correlated over time.

function proposed by Tse and Dong (2014) will not be appropriate for BT sampling if volatility and transaction activity exhibit different intraday periodicity patterns.

Figure 1 presents the intraday periodicity in volatility and trading activity of the stock JP Morgan (JPM) from January 2010 to April 2013. Figure 1A plots the means of the 1 min intraday realized volatilities over all trading days in the sample period, expressed in annualized standard deviation in percent, while Figure 1B plots the total number of transactions at each second from 9:30 a.m. to 4:00 p.m. over all trading days in the sample. We observe that the realized volatility at the beginning of the trading day is approximately two to three times larger than that near the end of the trading day. In contrast, the number of transactions at the end of the trading day is two to three times larger than that in the morning. This finding is quite regular across other stocks in our sample, which shows that intraday periodicity patterns in volatility and trading activity are quite different.² Thus, TTS returns in the morning will have larger volatility than those in the afternoon. In contrast, by construction, BTS returns will have nearly constant volatility over the whole trading day.

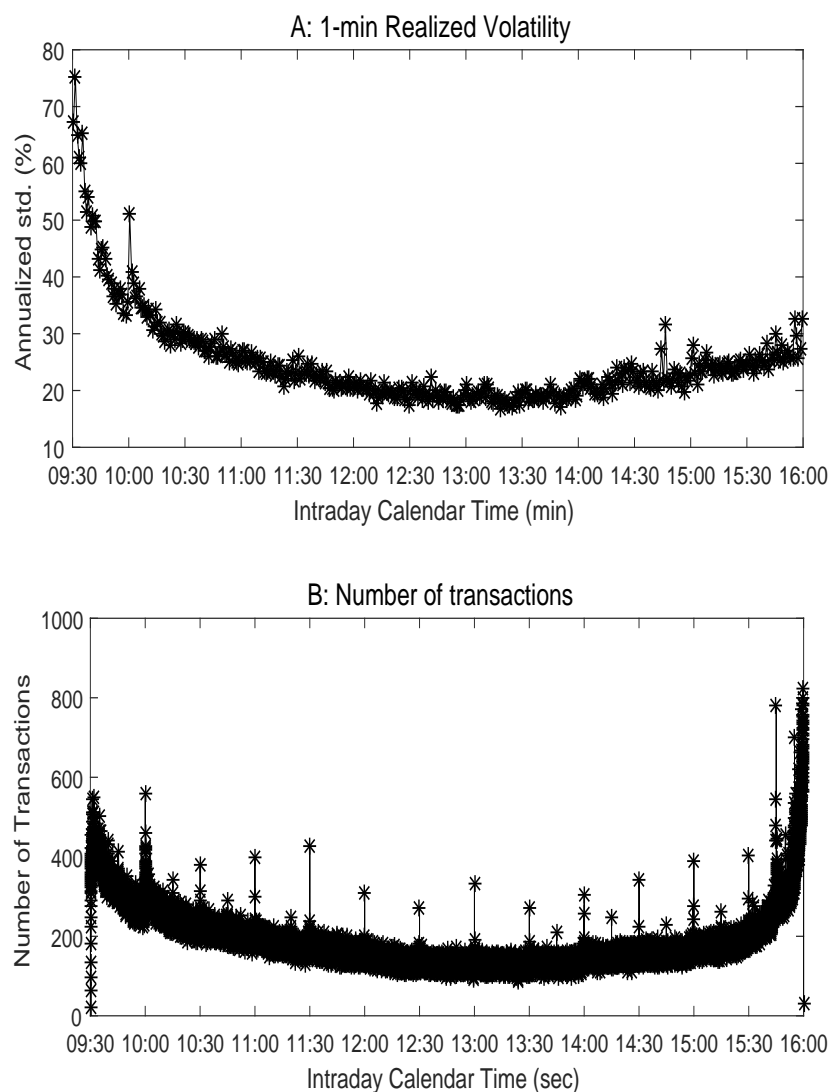


Figure 1. Intraday volatility and trading activity of JPMorgan Chase (JPM), January 2010 to April 2013.

² Results for other stocks can be found in the online supplementary material (Figure S1).

To implement the BTS scheme, Peters and De Vilder (2006) select transactions based on increments of estimated quadratic volatility. To alleviate the microstructure noise effect, transactions are selected sparsely, such as at 2 min frequency, to calculate the volatility. One drawback of this method is that it assumes away the price jumps, although this assumption is often rejected in the literature (see, e.g., Barndorff-Nielsen and Shephard (2004, 2006), Huang and Tauchen (2005), Lee and Mykland (2008) and Boudt et al. (2011)).³ Another drawback of this method (also in Andersen et al. 2010) is that researchers need to select a threshold value to obtain transactions at the target sampling frequency. The procedure of choosing the threshold has to be iterated, especially when the target frequency is high.

In this paper, we adopt the time-transformation approach to implement the BTS scheme. Let T denote the calendar-time length in seconds aggregated over all trading days in the sample. The time-transformation function $Q(t)$ at calendar time t (in sec), for $0 \leq t \leq T$, is computed as the empirical proportion of the intraday integrated volatility up to time t .⁴ The diurnally transformed time corresponding to calendar time t is denoted by \tilde{t} , with $\tilde{t} = T \times Q(t)$. Thus, returns over equal intervals of diurnally transformed time will have approximately equal integrated volatility. To obtain BTS transactions at a given frequency, we can select equally spaced transactions at the diurnally transformed time and choose the corresponding BTS transactions using the inverse function $Q^{-1}(t)$. In contrast to the methods in the literature, we do not need to choose a threshold value when implementing the BTS scheme.

We outline our method to obtain the BTS transactions as follows. Consider a sequence of estimated intraday integrated volatility V_k for K consecutive time intervals over the period $(0, T]$, with the end point in each time interval represented by t_k , for $k = 1, \dots, K$.⁵ Denote the collection of points t_k by \mathcal{H}_V ($t_0 = 0$) and define $N_{t_0} = 0$ and $N_{t_k} = \sum_{i=1}^k V_i$ for $k = 1, \dots, K$. The time-transformation function is calculated as $Q(t_k) = N_{t_k} / N_{t_K}$, for $t_k \in \mathcal{H}_V$. $Q(t)$ at any calendar-time point t can then be computed using a cubic interpolation that preserves monotonicity in t .⁶ To sample a sequence of calendar-time points with BTS duration h , we take equally spaced diurnally transformed BT points $\tilde{t}_j, j = 0, \dots, L$, with $\tilde{t}_j - \tilde{t}_{j-1} = h$ and $L = \lceil T/h \rceil$. Then, $t_j = Q^{-1}(\tilde{t}_j/T)$ are the required calendar-time points for the BTS scheme.

Figure 2A,B present the time-transformation functions for stock JPM based on intraday volatility and trading activity, respectively. These functions are computed by merging the data over the complete sample period, resulting in representative one-day time-transformation functions.⁷ Note that the two transformation functions exhibit different intraday patterns, with the compression of diurnally transformed time during market open more prominent for volatility than for trading activity.

³ One drawback of the jump detection methods is the presence of the spurious detections due to multiple testing issues. See Bajgrowicz et al. (2016) for a discussion.

⁴ As there are 6.5 h of trades in a trading day for the NYSE, for m trading days we have $T = 23400m$ s. $Q(t)$ at calendar time t (in s) is an increasing function of t , with $t = 0, 1, \dots, T$, $Q(0) = 0$ and $Q(T) = 1$.

⁵ Here, t_k are calendar-time points which need not to be regularly spaced. We outline the detailed steps in calculating V_k and t_k , for $k = 1, \dots, K$, in the Appendix A. V_k can be any suitable estimates of intraday integrated volatility. In this paper we use the TRV method of Barndorff-Nielsen et al. (2006) to calculate V_k for its robustness to jumps.

⁶ We use the Matlab (2015a, Mathworks, Natick, MA, United States) command *pchip* in this paper. Given $Q(t)$ and the calendar-time point t , the diurnally transformed time \tilde{t} is $\tilde{t} = T \times Q(t)$. Conversely, given $Q(t)$ and a diurnally transformed time \tilde{t} , the corresponding calendar time is $t = Q^{-1}(\tilde{t}/T)$.

⁷ In the empirical applications in this paper, the time-transformation function for BTS is extended over the whole sample period, which takes account of varying volatility over different trading days.

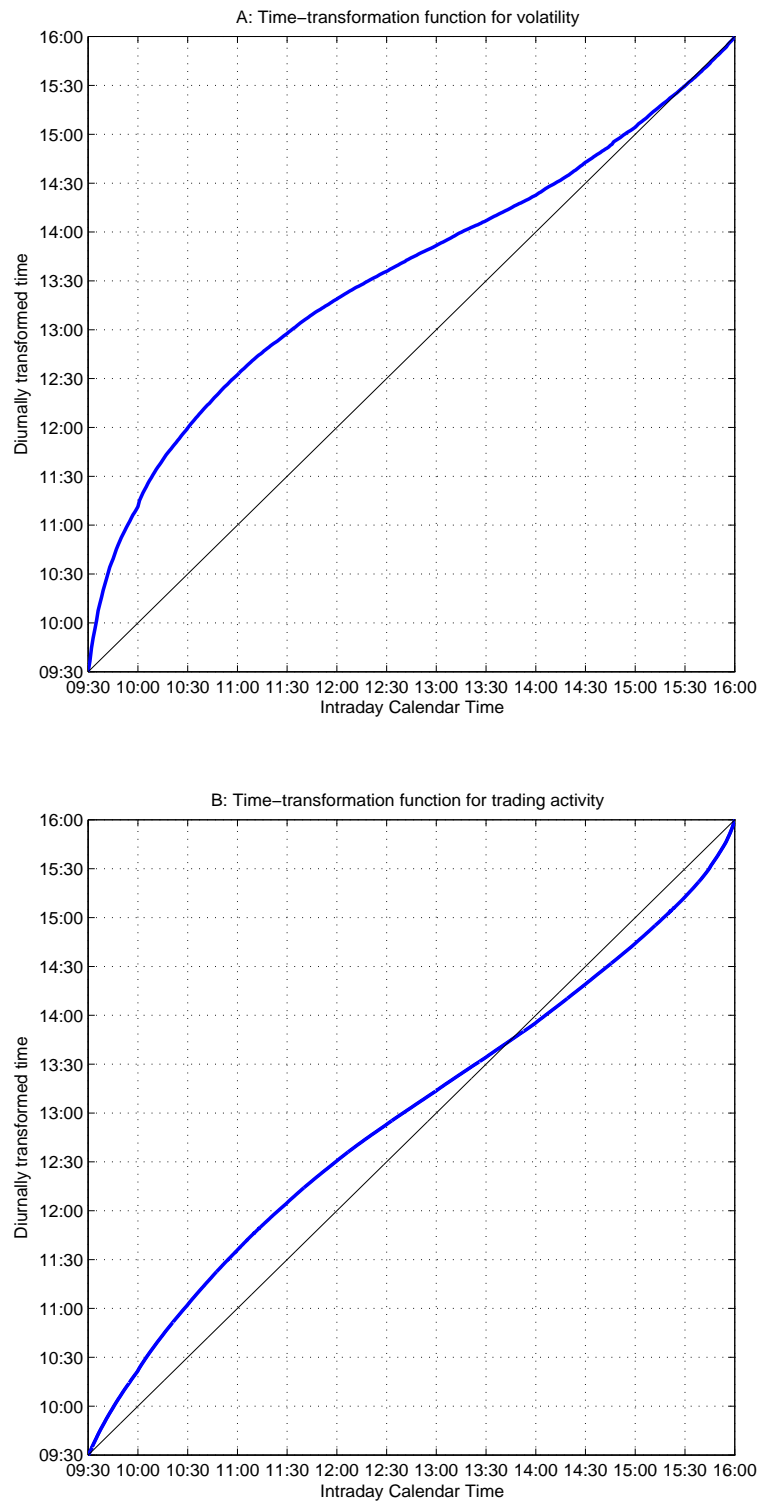


Figure 2. Time-transformation functions for JPM.

3. Testing the Semi-Martingale Hypothesis Using BTS Returns

As discussed above, BTS returns are iid normal for a semi-martingale price process even when there is leverage and/or feedback effect, whereas the CTS returns may not be iid normal even if the price process is a continuous local martingale. We now examine empirically the behavior of the BTS returns following the study of Andersen et al. (2010).

3.1. The Semi-Martingale Hypothesis

Let $r_k = Y_{t_k} - Y_{t_{k-1}}$ be the jump-adjusted returns over the time interval (t_{k-1}, t_k) .⁸ If the log-price follows a jump-diffusion process⁹ with no leverage and volatility feedback, r_k standardized by the integrated volatility will follow a standard normal distribution, i.e.,

$$\frac{r_k}{\left(\int_{t_{k-1}}^{t_k} \sigma^2(\tau) d\tau\right)} \sim N(0, 1), \quad k = 1, 2, \dots, \quad (1)$$

where $\sigma(\cdot)$ is the instantaneous volatility function. The above result, however, will not hold if $\sigma(\cdot)$ exhibits correlation over time (feedback effect) or with the log-price innovation (leverage effect). On the other hand, if the jump-adjusted returns are sampled over business time so that, over each business-time interval $(\tilde{t}_{k-1}, \tilde{t}_k)$, we have

$$\tilde{t}_k = \inf_{s > \tilde{t}_{k-1}} \left\{ \int_{\tilde{t}_{k-1}}^s \sigma^2(\tau) d\tau > \bar{\sigma}^2 \right\} \quad (2)$$

for a given volatility threshold $\bar{\sigma}^2$,¹⁰ Then, the jump-adjusted returns \tilde{r}_k over the business-time intervals $(\tilde{r}_{k-1}, \tilde{r}_k)$ satisfy

$$\frac{\tilde{r}_k}{\bar{\sigma}} \sim N(0, 1), \quad k = 1, 2, \dots. \quad (3)$$

To sample a sequence of BTS returns, Andersen et al. (2010) include intraday returns until the cumulative squared 5-min returns exceed a threshold, defined as the average daily or weekly realized volatility in calendar time (ABFN method hereafter). They report improved accuracy of the normal approximation under this sampling scheme.

3.2. Empirical Results of the Tests

To examine empirically the performance of our proposed BTS method versus the ABFN method, we use data of the top 40 market-capitalization stocks from the NYSE in 2010. We extract the tick-by-tick transaction data of these stocks from the TAQ database from January 2010 to April 2013. To clean the raw data, we follow the steps described in Tse and Dong (2014). Using the sequential jump-detection procedure of Andersen et al. (2010), we investigate the proportion of detected jumps (number of the detected jumps over the total number of sampled returns) when different sampling intervals and sampling schemes are used.¹¹

We test the normality assumption of the drift-corrected and jump-adjusted BTS returns (returns with jumps removed). For each trading day, all BTS returns with jumps are deleted and the

⁸ The jump-adjustment procedure can be found in the Appendix A.

⁹ The Brownian semimartingale process can be defined as $dY_t = \mu_t dt + \sigma_t dW_t$, where $\mu_t dt$ is the drift term, the instantaneous volatility process σ_t is càdlàg, and W_t denotes a standard Brownian motion independent of the drift. In this paper, we further add jumps to the Brownian semimartingale and assume the price process to be a generic jump-diffusion process. That is, $dY_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t$, where $dq_t = 1$ when there is a jump at time t , and $dq_t = 0$ otherwise, and κ_t denotes the jump size if a jump occurs at time t . We assume the jump component to be a finite activity jump process. Note that when there are infinite number of jumps in the data, our BTS method will work if we select BTS transactions based on estimated integrated volatility that are robust to the presence of Lévy-type jumps. See Lee and Hannig (2010) for the evidence of the presence of the Lévy-type jumps and see Barndorff-Nielsen et al. (2006) for an analysis of the multipower variation estimates when there are infinite number of jumps.

¹⁰ Note that, given \tilde{t}_{k-1} and $\bar{\sigma}^2$, \tilde{t}_k is the minimum business time so that the integrated volatility over the interval $(\tilde{t}_{k-1}, \tilde{t}_k)$ reaches $\bar{\sigma}^2$.

¹¹ As our focus here is the testing of the semi-martingale hypothesis, the results of the jump detection are not presented. Details of the selected stocks and results of the jump tests can be found in the supplementary material (Tables S1–S3) for which sampling frequencies of 1 min, 5 min and 10 min are used. When the sampling frequency is equal to 1 min, more than 12 stocks report jump proportions with values exceeding 10% under all sampling schemes. This suggests that sampling frequency that is too high (such as 1 min) may render misleading results when they are used for jump detection using the method of Andersen et al. (2010). See Oomen (2006) for an analysis of the performance of the realized variance estimator among alternative sampling schemes.

jump-adjusted 30-min, daily or weekly BTS returns are then computed by summing the remaining consecutive jump-adjusted BTS returns. We apply the Lilliefors (1967) test for normality to the jump-adjusted CTS returns, ABFN returns, and BTS returns. The results are reported in Table 1. It can be seen that the BTS method and the ABFN method substantially improve the normality approximation of the standardized return distribution over the Calendar-Time method. While the results for the weekly data are similar for the BTS method and the ABFN method, the BTS sampling scheme at higher frequencies restores normality for several stocks.

Table 1. Test of normality hypothesis for no-jump returns of 40 New York Stock Exchange (NYSE) stocks.

Frequency	CTS		ABFN		BTS	
	5%	1%	5%	1%	5%	1%
Weekly	15	9	3	1	4	0
Daily	40	40	12	5	6	4
30 min	40	40	40	40	30	26

Notes: The figures are the numbers of stocks (out of 40) for which the normality hypothesis at the 5% and 1% levels of significance are rejected based on the Lilliefors test implemented for 30 min, daily and weekly jump-adjusted returns, 2010–2013. The returns are constructed over 5 min sampling frequency using the Calendar Time Sampling (CTS) method, the method of Andersen et al. (2010) (ABFN) and the method proposed in this paper (BTS). The total number of trading days of each stock ranges from 830 to 853.

We further compare the performance of various methods by computing the moments of 30 min returns with jump-adjustment. In addition to using the Lilliefors (1967) test for normality, we also examine the skewness and kurtosis of the sampled returns of the 40 stocks and compute the average of the absolute difference of the calculated skewness and kurtosis versus 0 and 3, respectively. The results are reported in Table 2. We observe that, for high-frequency returns, the BTS method performs the best in restoring normality, while the TTS method performs the worst. This observation confirms our finding in Section 2 that we cannot use the TTS scheme to approximate the BTS scheme to select intraday returns. For illustration, Figure 3 presents the QQ (Quantile-Quantile) plots of the 30 min jump-adjusted ABFN returns and BTS returns for the JPM stock data. Our proposed BTS method performs better than the ABFN method in restoring normality for returns sampled at high frequency.¹²

Table 2. Mean of the absolute difference for no-jump 30 min returns of 40 New York Stock Exchange (NYSE) stocks.

Measures of shape	CTS	TTS	ABFN	BTS
Skewness (diff.)	0.1272	0.2091	0.1105	0.0384
Kurtosis (diff.)	5.3745	10.7893	2.3749	0.2792

Notes: The figures are the average of the absolute difference of the calculated skewness and kurtosis versus 0 and 3, respectively. To calculate the skewness and kurtosis, we use 30 min jump-adjusted returns (2010–2013) of 40 stocks. The 30 min returns are constructed over 5 min sampling frequency using the Calendar Time Sampling (CTS) method, the Tick Time Sampling (TTS) method, the method of Andersen et al. (2010) (ABFN) and the method proposed in this paper (BTS). The total number of trading days of each stock ranges from 830 to 853.

¹² QQ plots of the jump-adjusted ABFN returns and BTS returns at daily and weekly frequencies are very similar. We also calculate the ACF values of the sampled 30 min returns up to lag 150. All returns sampled using various methods exhibit no periodicity. The correlation between the ABFN returns and BTS returns increases as the sampling frequency decreases, and the value is around 13.5% at weekly frequency.

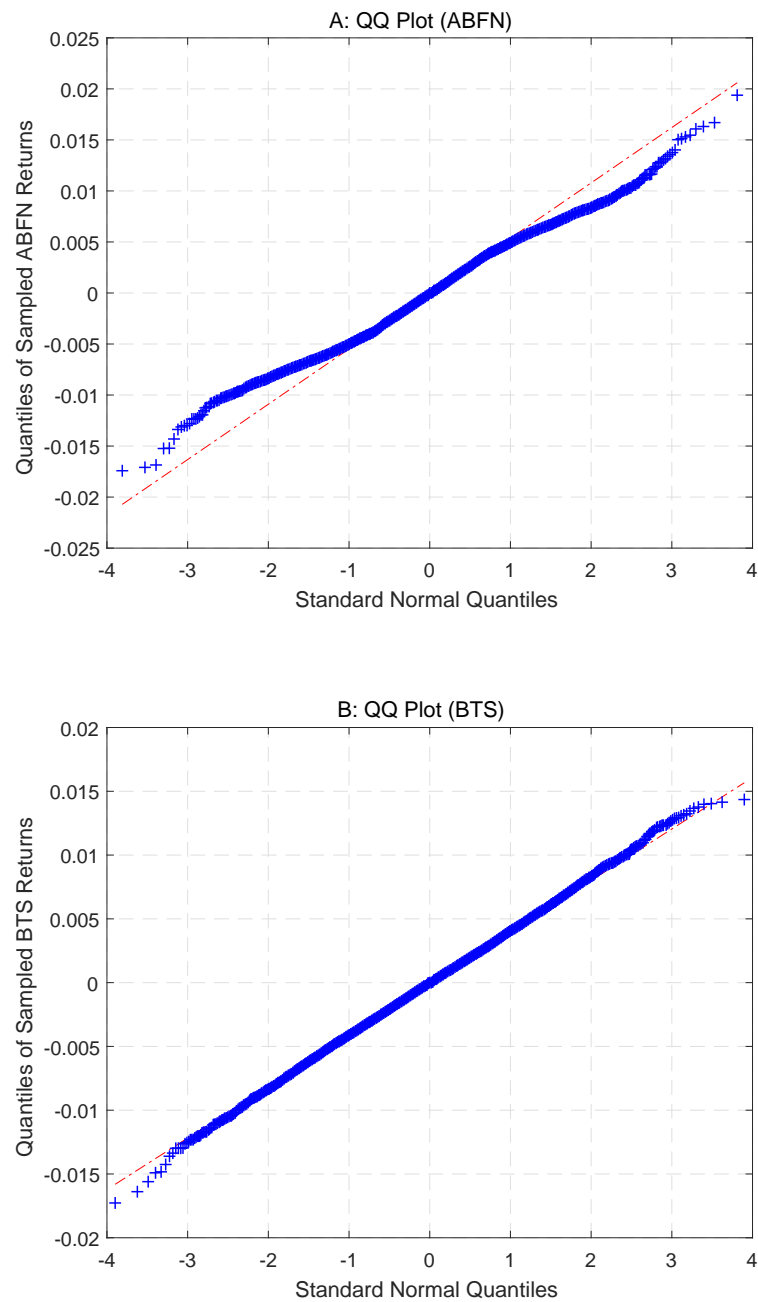


Figure 3. QQ (Quantile-Quantile) plots of sampled 30 min returns for JPMorgan Chase (JPM), January 2010–April 2013.

4. Estimation of Integrated Volatility

We now examine the use of the BTS scheme for estimating integrated volatility. We consider two methods of estimating daily volatility: Realized Volatility (RV) method and Autoregressive Conditional Duration-Integrated Conditional Volatility (ACD-ICV) method. The literature on RV estimation has grown tremendously since its inception. In this paper, we select the Tripower Realized Volatility (TRV) estimate of [Barndorff-Nielsen et al. \(2006\)](#) for its robustness to price jumps.¹³ We compare the

¹³ We do not use the bipower realized volatility method here since the TRV method is more robust to the presence of jumps, especially when the sampling frequency is high.

performance of the TRV estimates when returns are sampled by BTS, CTS and TTS schemes, with and without subsampling. With the same estimator used, the performance of the estimates is only differentiated by the sampling method. We also consider the use of the ACD-ICV approach, with some modifications based on the BTS methodology.

4.1. Integrated Volatility Estimation Using BT Returns

For a given subgrid \mathcal{H} , the TRV estimate is computed as

$$V_T = \xi_2^{-3} \left[\sum_{i=2}^{|\mathcal{H}|-1} |r_{i,-}|^{\frac{2}{3}} |r_i|^{\frac{2}{3}} |r_{i,+}|^{\frac{2}{3}} \right], \tag{4}$$

where $r_i = (Y_{i,+} - Y_i)$ for $Y_{i,+}, Y_i \in \mathcal{H}$ and $\xi_k = 2^{\frac{k}{2}} \Gamma((k+1)/2) / \Gamma(1/2)$ for $k > 0$, with $\Gamma(\cdot)$ denoting the gamma function and $Y_{i,+}$ denoting the elements following Y_i in \mathcal{H} . We define $|\mathcal{H}|$ as number of points in grid \mathcal{H} minus 1. For the CTS scheme, the previous-tick method is adopted when there is no transaction at the selected time point. For the TTS scheme, the number of subsampling grids S is selected to ensure that each subgrid has transactions at the target sampling frequency. To implement the subsampling method under the BTS scheme, we select BTS transactions with the sampling frequency being twice the average transaction duration. Subsampling is then implemented to obtain subgrids at the target sampling frequency.

4.2. Integrated Volatility Estimation Using the Modified ACD-ICV Method

Tse and Yang (2012) propose the ACD-ICV method to estimate daily and intraday volatility by modeling the price durations parametrically. They point out that, for short intraday intervals, such as an hour or 15 min, the RV methods use only local data for the period of interest. Thus, the infill sample size may not be large enough to justify the applicability of the asymptotics of the RV estimates. In contrast, the ACD-ICV method estimates the conditional volatility using data beyond the period of interest and produces better estimates of volatility over short intraday intervals. Their simulation results show that the ACD-ICV method performs better than other methods (such as the Realized Kernel method of Barndorff-Nielsen et al. (2008) in estimating the daily, 1 h and 15 min integrated volatility.

The ACD-ICV method samples observations from the observed transaction data based on a pre-specified price threshold δ . That is, transactions are selected whenever the absolute price change exceeds the threshold δ , which are the *price events*. Suppose $\mathcal{H}_{PE} = \{t_0, t_1, t_2, \dots, t_N\}$ is the selected price events and the i th *price duration* is $x_i = t_i - t_{i-1}, i = 1, \dots, N$. Let Φ_i denote the information set upon the price event at time t_i . Denote $\psi_{i+1} = E(x_{i+1} | \Phi_i)$ as the conditional expectation of the price duration and assume that the standardized durations $\epsilon_i = x_i / \psi_i, i = 1, \dots, N$ are iid positive random variables with a mean of unity. Given the information Φ_i at time t_i , the conditional instantaneous return variance per unit time at time $t > t_i$, denoted by $\sigma^2(t | \Phi_i)$, is

$$\sigma^2(t | \Phi_i) = \frac{\delta^2}{\psi_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right), \tag{5}$$

where $\lambda(\cdot)$ is the hazard function of ϵ_i . Assuming ϵ_i to be iid standard exponential distributed, the integrated conditional variance (ICV) over time period $[t_{n_1}, t_{n_2+1}]$ is calculated as

$$ICV = \delta^2 \sum_{i=n_1}^{n_2} \left[\frac{t_{i+1} - t_i}{\psi_{i+1}} \right]. \tag{6}$$

The conditional expectation of the price durations ψ_i can be estimated by various methods, such as the ACD method of Engle and Russell (1998) or the Augmented ACD method of Fernandes and Grammig (2006).

The original ACD-ICV method models the price durations obtained by the threshold δ by assuming that these durations have equal volatility, and this assumption may not be true. In this paper, we modify the ACD-ICV method in two ways. First, instead of modeling the durations of the price events, we model the BTS durations that are obtained based on volatility change (instead of the absolute price change). The BTS returns are sampled as in Section 2. Second, we replace δ^2 in Equation (6) by an estimate of the mean volatility over each sampled BTS return. Suppose there are K BTS returns over m trading days, and the estimated integrated volatility over these m trading days is equal to V_m . Then, each BTS return has an approximately constant integrated volatility of $V_D = V_m/K$. Instead of using δ^2 as an approximation of the integrated volatility of each price event as in Tse and Yang (2012), we use V_D to replace δ^2 in Equation (6) to obtain a new ACD-ICV estimate.

Thus, for the ACD-ICV approach, we consider three variations of estimates. We first select transactions using the price-event sampling method (where transactions are selected based on absolute price change) and estimate the daily integrated volatility using the ACD-ICV method as in Tse and Yang (2012). We denote this method by ME1. We then replace δ^2 in equation (6) by V_D , which is the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. We call this method ME2.¹⁴ Finally, we sample data using the BTS scheme (not by price events) and repeat the computation as in ME2, which is called ME3. We compare the daily volatility estimates using the ACD-ICV methods against the RK method.¹⁵

5. Monte Carlo Study

We conduct a Monte Carlo (MC) study to examine the performances of different integrated volatility estimates. Our MC set-up draws upon other models in the literature.

5.1. Simulation Models

We consider five simulation models, which are summarized in Table 3. Models MD1 and MD2 are the Heston models (Heston (1993) and Aït-Sahalia and Mancini (2008) with some modifications) with high and low volatility, respectively. MD3 is the two-factor affine stochastic volatility model with an intraday U-shape pattern Hasbrouck (1999) and Andersen et al. (2012). MD4 is a deterministic volatility set-up Tse and Yang (2012). Finally, MD5 is MD1 with price jumps.

For all set-ups above, we set the initial price to 60 and the initial value of σ to 30%. We introduce sparsity of trade to the data by simulating exponentially distributed calendar-time transaction durations.¹⁶ We first simulate transactions second by second and then generate exponentially distributed transaction durations with mean equal to 5 s, 10 s and 20 s, respectively. For simplicity, we only investigate iid market microstructure noise with constant noise-signal ratio (NSR). Based on the findings in Dong and Tse (2017), we consider cases with NSR = 0.005%, 0.01% and 0.02%. We introduce a 0.01 price rounding error in all simulations. The intraday duration periodicity is adjusted by the time-transformation method in Tse and Dong (2014) before we fit all price durations and BTS durations to the ACD model. Each model is simulated over 60 trading days, with the simulation repeated 1000 times.

¹⁴ Note that, for ME1 and ME2, the returns are sampled by price events and the ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades.

¹⁵ The RK method is selected for comparison due to its superior performance among the RV estimators (see Barndorff-Nielsen et al. (2008)). To calculate the bandwidth of the RK method, we use the subsampling realized volatility estimator and 3 min TTS returns. For the ACD-ICV methods, all results in this paper are based on conditional duration models fitted using the power ACD (PACD) model (see Fernandes and Grammig 2006).

¹⁶ Sparsity occurs as empirically transactions are not observed sec by sec. Inactive stocks typically have more sparse transactions.

Table 3. Summary of simulation models.

Model	Code	Description of Model	Description of Model Parameters
Heston Model (high volatility)	MD1	$d \log X(t) = \left(\mu - \frac{\sigma^2(t)}{2} \right) dt + \sigma(t) dW_1(t),$ $d\sigma^2(t) = \kappa(\alpha - \sigma^2(t))dt + \gamma\sigma(t)dW_2(t).$	$\mu = 0.05, \kappa = 5, \alpha = 0.25, \gamma = 0.5$ and $\text{Corr}(dW_1(t), dW_2(t)) = -0.5$.
Heston Model (low volatility)	MD2	Same as MD1.	$\alpha = 0.04$, all remaining parameters same as MD1.
Two-factor affine stochastic volatility model with U-shape intraday volatility pattern	MD3	$d \log X(t) = \sigma_u(t)\sigma_{sv}(t)dW_1(t), \sigma_{sv}^2(t) = \sigma_1^2(t) + \sigma_2^2(t),$ $d\sigma_1^2(t) = \kappa_1[\theta_1 - \sigma_1^2(t)]dt + \eta_1\sigma_1(t)dW_{21}(t),$ $d\sigma_2^2(t) = \kappa_2[\theta_2 - \sigma_2^2(t)]dt + \eta_2\sigma_2(t)dW_{22}(t),$ $\sigma_u(t) = C + Ae^{-at} + Be^{-b(1-t)}, t \in [0, 1].$	$\kappa_1 = 0.6, \kappa_2 = 0.1, \theta_1 = 0.09, \theta_2 = 0.04, \eta_1 = 0.2, \eta_2 = 0.1,$ $\rho_1 = \text{Corr}(dW_1(t), dW_{21}(t)) = 0.9,$ and $\rho_2 = \text{Corr}(dW_1(t), dW_{22}(t)) = -0.4. A = 0.75, B = 0.25,$ $C = 0.88929198, a = 10$ and $b = 10$.
Deterministic volatility model with U-shape intraday volatility pattern	MD4	$d \log X(u) = \sigma(u)dW(u), \sigma(u) = \sigma(t, \tau) = \sigma_1(t)\sigma_2(\tau),$ where t is the day of trade and τ is the intraday time. $\sigma_1(t)$ is the volatility of day $t, \sigma_2(\tau)$ is the intraday variations at time τ of each day.	$\sigma_1(t) = 20\%$ for $t = 1$ with $\sigma_1(t)$ increasing linearly in t over 20 days to reach 30%. It then remains level for the next 20 days and decreases linearly in t to 20% over 20 days. $\sigma_2(\tau)$ is computed as in Tse and Yang (2012) using the IBM tick-by-tick transaction data in 2012.
MD1 with price jumps	MD5	$d \log X(t) = \left(\mu - \frac{\sigma^2(t)}{2} \right) dt + \sigma(t)dW_1(t) + J(t)dP(t),$ $d\sigma^2(t) = \kappa(\alpha - \sigma^2(t))dt + \gamma\sigma(t)dW_2(t).$	$P(t)$ is a Poisson process with on average one price jump every two days. $J(t)$ is the size of the jumps with $J(t) \sim N(0.02, 0.004)$.

5.2. Simulation Results

We first report our results on the TRV estimates. For each model set-up, sampling frequencies of 1, 2, 3, 5 and 10 min are considered. We compute TRV based on the CTS, TTS and BTS returns, with and without subsampling. To save space, we present only the results for the BTS scheme with subsampling in Table 4. Table 5 summarizes the average difference in RMSE of the CTS and TTS schemes versus the BTS scheme over all models and parameter set-ups, with subsampling, in both absolute and relative terms. All results can be found in the supplementary material (Tables S4–S9). The BTS returns perform the best in reporting generally smaller root mean-squared error (RMSE), especially for methods with no subsampling. When the subsampling method is used, the RMSE decreases substantially. Generally, the BTS scheme still performs the best and its advantage is especially obvious for MD3 and MD4 when there is intraday volatility periodicity in the simulated price process.¹⁷ Under all sampling schemes, the TRV estimates suffer from the market microstructure noise problems when the NSR is large, resulting in high mean error (ME) at high sampling frequency. When sparsity and NSR are both low, high sampling frequency at 1 min interval produces the lowest RMSE. Overall, the BTS returns outperform the CTS and TTS returns in estimating the integrated volatility.

We now turn to the results of the ACD-ICV method. Tables 6 and 7 report the ME and RMSE of the RK and the ACD-ICV estimates for MD1 and MD5, respectively. Results for other models can be found in the online supplementary material (Tables S10–S12). Table 8 summarizes the average RMSE of Realized Kernel estimate versus the ACD-ICV estimates over all models and parameter set-ups. The RK method performs quite well for the unbiasedness property, reporting small ME for all models except MD5 (model with price jump). ME1 reports quite big absolute ME and RMSE values among all the sampling frequencies considered and it often performs worse than the RK method except for MD5.¹⁸ In contrast, the modified ACD-ICV methods, ME2 and ME3, report quite small ME and RMSE. ME3 consistently reports smaller RMSE than ME2 except for few cases in MD3 and MD5 at the 3 min and 5 min sampling frequencies. Moreover, ME2 varies more across different sampling frequencies. This demonstrates the superiority of the BTS durations over the price durations when they are fitted to the ACD model to estimate the integrated volatility. The better performance of ME3 over ME2 is mainly due to the fact that the BTS scheme performs better in yielding returns with constant volatility compared against the ACD-ICV method using the price events.

Table 4. Mean error (ME) and root mean-squared error (RMSE) of estimates of daily volatility using the Tripower Realized Volatility (TRV) method under the Business Time Sampling (BTS) scheme with subsampling.

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-s	0.005%	MD1	−0.4286	0.0470	0.0031	0.0004	−0.1777	1.3869	1.7228	2.1610	2.8165	4.0975
		MD2	−0.4321	−0.0500	−0.0513	−0.0338	−0.1309	0.9964	1.1706	1.4665	1.9242	2.7797
		MD3	−0.4435	0.0200	0.0123	−0.0019	−0.1916	1.2425	1.5657	1.9431	2.5511	3.6823
		MD4	−0.2403	−0.0913	0.0126	0.0055	−0.2069	1.0574	1.3723	1.6849	2.2278	3.2164
		MD5	−0.0687	0.6282	0.7732	1.0940	1.6463	1.4163	2.0196	2.5587	3.4485	5.2161
	0.01%	MD1	0.3382	0.4486	0.2625	0.1532	−0.1142	1.3520	1.7646	2.1562	2.8077	4.0850
		MD2	0.0786	0.2172	0.1192	0.0675	−0.0884	0.9049	1.1831	1.4608	1.9152	2.7695
		MD3	0.3291	0.4002	0.2555	0.1439	−0.1301	1.2126	1.6016	1.9521	2.5460	3.6734
		MD4	0.5904	0.3128	0.2734	0.1532	−0.1474	1.1809	1.4043	1.7059	2.2373	3.2230
		MD5	0.7280	1.0331	1.0365	1.2492	1.7128	1.5989	2.1671	2.6391	3.4936	5.2306

¹⁷ When there is intraday volatility periodicity, the BTS returns resemble more closely to normal distribution than the CTS and TTS returns.

¹⁸ This is in contrast to the findings in Tse and Yang (2012), which shows the superiority of the ACD-ICV method over the RK method via simulation using second-by-second transactions (sparsity of 1 s). The poor performance of ME1 is mainly due to the transaction sparsity, since using σ^2 as the proxy for integrated volatility over one price event becomes unreliable when transactions are sparse. Supporting evidence is provided in our simulation study that the RMSE of ME1 increases when observed transactions are more sparse.

Table 4. Cont.

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-s	0.02%	MD1	1.8463	1.2523	0.7789	0.4499	0.0126	2.2762	2.1053	2.2506	2.8148	4.0590
		MD2	1.0802	0.7479	0.4603	0.2651	-0.0059	1.4503	1.3864	1.5109	1.9145	2.7529
		MD3	1.7862	1.1644	0.7397	0.4265	-0.0167	2.1440	1.9166	2.0542	2.5558	3.6603
		MD4	2.2616	1.1036	0.7889	0.4437	-0.0291	2.4966	1.7643	1.8609	2.2798	3.2357
		MD5	2.2713	1.8503	1.5649	1.5559	1.8424	2.7080	2.6518	2.8753	3.5982	5.2587
10-s	0.005%	MD1	-0.9375	-0.0435	-0.0307	-0.0333	-0.1880	1.6335	1.8357	2.1600	2.8457	4.1021
		MD2	-0.8123	-0.1034	-0.0778	-0.0560	-0.1417	1.2304	1.2500	1.4735	1.9394	2.7833
		MD3	-0.8436	-0.0958	-0.0423	-0.0340	-0.2088	1.4854	1.6584	1.9742	2.5751	3.7013
		MD4	-0.5797	-0.2436	-0.0830	-0.0584	-0.2432	1.2715	1.4404	1.7121	2.2547	3.2308
		MD5	-0.5986	0.5429	0.7290	1.0637	1.6334	1.5659	2.0991	2.5573	3.4711	5.2171
	0.01%	MD1	0.0713	0.3397	0.2356	0.1224	-0.1251	1.3501	1.8561	2.1611	2.8355	4.0856
		MD2	-0.1345	0.1520	0.1010	0.0460	-0.0992	0.9467	1.2544	1.4661	1.9323	2.7757
		MD3	0.1028	0.2815	0.2097	0.1090	-0.1529	1.2390	1.6800	1.9774	2.5686	3.6909
		MD4	0.3144	0.1905	0.1929	0.0949	-0.1830	1.1801	1.4321	1.7293	2.2613	3.2356
		MD5	0.4190	0.9362	0.9980	1.2244	1.6968	1.5275	2.2311	2.6363	3.5126	5.2537
	0.02%	MD1	2.0479	1.1137	0.7770	0.4235	-0.0051	2.5005	2.1285	2.2661	2.8358	4.0650
		MD2	1.1815	0.6675	0.4586	0.2426	-0.0175	1.5885	1.4198	1.5277	1.9301	2.7608
		MD3	1.9622	1.0222	0.7158	0.3933	-0.0340	2.3465	1.9448	2.0766	2.5855	3.6804
		MD4	2.1082	1.0387	0.7207	0.3933	-0.0644	2.4205	1.7728	1.8823	2.2997	3.2520
		MD5	2.4111	1.7259	1.5437	1.5352	1.8277	2.8718	2.6669	2.8772	3.6225	5.2623
20-s	0.005%	MD1	-1.9043	-0.4650	-0.1715	-0.1586	-0.3052	2.5154	1.9081	2.3632	2.8884	4.1176
		MD2	-1.4513	-0.3972	-0.1718	-0.1551	-0.2189	1.8465	1.3232	1.6142	1.9728	2.8010
		MD3	-1.7349	-0.4945	-0.1976	-0.1617	-0.3245	2.2880	1.7421	2.0996	2.6517	3.7326
		MD4	-1.3673	-0.5413	-0.2730	-0.2261	-0.3586	1.9372	1.6463	1.8101	2.3044	3.2759
		MD5	-1.5684	0.0539	0.6149	0.9467	1.5263	2.3452	2.0185	2.7000	3.4925	5.2178
	0.01%	MD1	-0.7727	0.0595	0.0795	-0.0012	-0.2474	1.8388	1.8658	2.3619	2.8691	4.1079
		MD2	-0.6930	-0.0470	-0.0025	-0.0515	-0.1798	1.3436	1.2747	1.6111	1.9649	2.7898
		MD3	-0.6715	0.0110	0.0681	-0.0090	-0.2678	1.6684	1.6857	2.1131	2.6393	3.7245
		MD4	-0.3601	-0.0577	0.0292	-0.0623	-0.3056	1.4354	1.5473	1.7854	2.3140	3.2870
		MD5	-0.4358	0.5872	0.8854	1.1072	1.5868	1.8245	2.1169	2.7876	3.5323	5.2285
	0.02%	MD1	1.4062	1.0834	0.5790	0.3089	-0.1229	2.2543	2.1942	2.4346	2.8722	4.0765
		MD2	0.7467	0.6389	0.3302	0.1612	-0.0953	1.4646	1.4674	1.6470	1.9576	2.7740
		MD3	1.3598	0.9923	0.5629	0.2802	-0.1573	2.1106	1.9812	2.2102	2.6445	3.7064
		MD4	1.6100	0.9172	0.6508	0.2535	-0.1811	2.1813	1.8069	1.9153	2.3599	3.2984
		MD5	1.7450	1.6299	1.4063	1.4237	1.7176	2.5623	2.6474	3.0191	3.6372	5.2592

Notes: ME and RMSE are the mean error and root mean-squared error, respectively, of the volatility estimates in annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions. NSR is the noise-signal ratio. Results for average transaction frequency of 1 min, 2 min, 3 min, 5 min and 10 min are reported. The sampling frequency of the BTS scheme equals twice the average transaction duration.

Table 5. Comparison of the Tripower Realized Volatility (TRV) estimates under different sampling schemes with subsampling.

Average RMSE Difference of the TRV Estimates	1-min	2-min	3-min	5-min	10-min
Avg. (RMSE(CTS)-RMSE(BTS))	0.3043	0.1666	0.1891	0.2215	0.1926
Avg. (RMSE(CTS)-RMSE(BTS))/RMSE(BTS) (%)	17.8262	9.8515	9.6493	8.8626	5.7456
Avg. (RMSE(TTS)-RMSE(BTS))	-0.0237	0.1261	0.1721	0.2240	0.2072
Avg. (RMSE(TTS)-RMSE(BTS))/RMSE(BTS) (%)	-0.1011	7.9977	9.0710	9.1005	6.2101

Notes: RMSE is the root mean-squared error of the daily TRV estimates in annualized standard deviation in percent. The first row is the average of the RMSE of the TRV estimates (over all simulation models) based on the CTS scheme over that based on the BTS scheme. The second row is the average relative difference in percentage. The third and fourth rows are similarly presented for the TTS scheme compared to the BTS scheme.

Table 6. Mean error (ME) and root mean-squared error (RMSE) of daily volatility estimates of the realized kernel (RK) and Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) methods for Model MD1.

Sparsity	NSR	RK	ME					RK	RMSE						
			ACD-ICV	Avg. Sampling Frequency (ACD-ICV)					ACD-ICV	Avg. Sampling Frequency (ACD-ICV)					
				1-min	3-min	5-min	10-min			15-min	1-min	3-min	5-min	10-min	15-min
5-s	0.005%	−0.1087	ME1	−5.8450	−3.4421	−2.6778	−1.8270	−1.3546	2.2978	ME1	6.2142	3.8881	3.3149	2.9738	2.9941
			ME2	−0.2043	−0.2366	−0.2205	−0.0928	0.0734		ME2	1.7527	1.7500	1.9673	2.3810	2.6985
			ME3	−0.1120	0.0540	0.2446	0.4362	0.7306		ME3	1.2772	1.4001	1.7261	1.4380	1.5706
	0.01%	−0.0876	ME1	−3.4503	−2.0139	−1.5554	−1.0491	−0.7357	2.2998	ME1	3.8811	2.6241	2.4551	2.5735	2.7766
			ME2	0.0437	0.0089	0.0301	0.1633	0.3410		ME2	1.6266	1.6906	1.9326	2.3896	2.7213
			ME3	0.1489	0.3076	0.4972	0.6950	0.9897		ME3	1.2367	1.3710	1.7078	1.4948	1.6735
	0.02%	−0.0461	ME1	1.4407	0.6924	0.4711	0.3438	0.3824	2.3040	ME1	2.0920	1.8468	2.0100	2.4290	2.7559
			ME2	0.5181	0.5045	0.5353	0.6804	0.8449		ME2	1.5778	1.7868	2.0340	2.5118	2.8656
			ME3	0.6677	0.8253	1.0079	1.2071	1.5051		ME3	1.3320	1.4939	1.8291	1.7399	1.9751
10-s	0.005%	−0.1416	ME1	−10.9555	−6.0669	−4.6302	−3.1722	−2.4450	2.6341	ME1	11.3935	6.4347	5.0926	3.9820	3.6161
			ME2	−0.3400	−0.3899	−0.3720	−0.2555	−0.0909		ME2	2.2864	1.8765	2.0363	2.4242	2.6968
			ME3	−0.2235	−0.1976	0.0212	0.2764	0.5716		ME3	1.4774	1.4302	1.5870	1.4764	1.5774
	0.01%	−0.1178	ME1	−8.5745	−4.8751	−3.7310	−2.5558	−1.9404	2.6363	ME1	8.9899	5.2580	4.2561	3.4911	3.2988
			ME2	−0.0858	−0.1337	−0.1165	0.0015	0.1719		ME2	2.0281	1.7825	1.9930	2.4033	2.7120
			ME3	0.0409	0.0713	0.2829	0.5387	0.8356		ME3	1.4179	1.3836	1.5621	1.5129	1.6586
	0.02%	−0.0710	ME1	−4.8863	−2.6923	−2.0864	−1.4154	−1.0327	2.6399	ME1	5.3222	3.2140	2.8523	2.7546	2.8704
			ME2	0.4241	0.3767	0.3976	0.5304	0.7014		ME2	1.8341	1.7717	2.0131	2.4689	2.8195
			ME3	0.5665	0.5917	0.8060	1.0634	1.3599		ME3	1.4587	1.4454	1.6547	1.7292	1.9392
20-s	0.005%	−0.1987	ME1	−17.7796	−9.5646	−7.3266	−5.0140	−3.9444	3.0600	ME1	18.3065	9.9612	7.7382	5.6130	4.7999
			ME2	−0.6770	−0.7315	−0.7231	−0.6051	−0.4518		ME2	3.1408	2.2301	2.2493	2.5140	2.7698
			ME3	−0.4760	−0.5490	−0.3795	−0.0762	0.2182		ME3	2.2789	1.6071	1.7431	1.5932	1.6135
	0.01%	−0.1723	ME1	−16.2877	−8.6764	−6.6429	−4.5336	−3.5434	3.0621	ME1	16.8362	9.0603	7.0596	5.1679	4.4519
			ME2	−0.4067	−0.4638	−0.4586	−0.3355	−0.1811		ME2	2.9539	2.0830	2.1464	2.4481	2.7132
			ME3	−0.2070	−0.2778	−0.1000	0.1926	0.4882		ME3	2.2062	1.5119	1.7034	1.5809	1.6494
	0.02%	−0.1200	ME1	−12.8662	−7.0624	−5.3456	−3.6371	−2.8233	3.0670	ME1	13.3676	7.4466	5.7892	4.3706	3.9009
			ME2	0.1191	0.0629	0.0728	0.2021	0.3562		ME2	2.6226	1.9304	2.0461	2.4303	2.7561
			ME3	0.3287	0.2614	0.4493	0.7287	1.0260		ME3	2.1632	1.4769	1.7526	1.7006	1.8539

Notes: ME and RMSE are the mean error and root mean-squared error, respectively, of the volatility estimates in annualized standard deviation in percentage. MD1 is the Heston model and the average true daily integrated volatility is around 40%. ME1 is the ACD-ICV method of [Tse and Yang \(2012\)](#). ME2 is ME1 with δ^2 replaced by V_D , the integrated volatility estimated using TRV with subsampling at 3 min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in [Tse and Dong \(2014\)](#).

Table 7. Mean error (ME) and root mean-squared error (RMSE) of daily volatility estimates of the realized kernel (RK) and Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) methods for Model MD5.

Sparsity	NSR	RK	ME					RK	RMSE						
			ACD-ICV	Avg. Sampling Frequency (ACD-ICV)					ACD-ICV	Avg. Sampling Frequency (ACD-ICV)					
				1-min	3-min	5-min	10-min			15-min	1-min	3-min	5-min	10-min	15-min
5-s	0.005%	5.4227	ME1	−5.8297	−3.3882	−2.5730	−1.6045	−1.0188	9.9602	ME1	6.2026	3.8409	3.2360	2.8656	2.8789
			ME2	0.8058	0.7685	0.7892	0.9164	1.0902		ME2	1.9098	1.9199	2.1430	2.6091	2.9672
			ME3	0.8846	1.0464	1.2556	1.4577	1.7653		ME3	1.6636	1.8671	2.1671	2.0848	2.2923
	0.01%	5.4418	ME1	−3.4340	−1.9407	−1.4617	−0.8313	−0.3769	9.9686	ME1	3.8603	2.5721	2.4208	2.5277	2.7279
			ME2	1.0512	1.0187	1.0453	1.1758	1.3516		ME2	1.9215	2.0070	2.2572	2.7266	3.0797
			ME3	1.1471	1.3082	1.5160	1.7153	2.0254		ME3	1.7863	1.9921	2.2839	2.2406	2.4745
	0.02%	5.4794	ME1	1.4728	0.7713	0.5849	0.5818	0.7257	9.9860	ME1	2.0990	1.8775	2.0612	2.5042	2.8495
			ME2	1.5310	1.5181	1.5467	1.6891	1.8715		ME2	2.1539	2.3315	2.5607	3.0135	3.3723
			ME3	1.6721	1.8222	2.0263	2.2336	2.5459		ME3	2.1213	2.3145	2.5816	2.6222	2.8867
10-s	0.005%	5.3907	ME1	−10.9449	−6.0276	−4.5318	−2.9491	−2.1042	10.0436	ME1	11.3795	6.4015	5.0126	3.8136	3.4271
			ME2	0.6633	0.6106	0.6242	0.7470	0.9182		ME2	2.3104	1.9461	2.1377	2.5742	2.9195
			ME3	0.7705	0.7948	0.9985	1.2914	1.5981		ME3	1.6828	1.8435	2.0209	2.0229	2.2163
	0.01%	5.4121	ME1	−8.6446	−4.8146	−3.6363	−2.3242	−1.5842	10.0534	ME1	9.0607	5.2047	4.1744	3.3347	3.1353
			ME2	0.9153	0.8660	0.8891	1.0125	1.1800		ME2	2.1941	1.9921	2.2058	2.6559	3.0205
			ME3	1.0358	1.0604	1.2689	1.5548	1.8631		ME3	1.7876	1.9431	2.1333	2.1752	2.3929
	0.02%	5.4548	ME1	−4.8774	−2.6232	−1.9874	−1.1786	−0.6562	10.0727	ME1	5.3053	3.1592	2.7833	2.6525	2.7837
			ME2	1.4283	1.3848	1.4085	1.5407	1.7137		ME2	2.2591	2.2408	2.4554	2.9128	3.2801
			ME3	1.5657	1.5953	1.7983	2.0797	2.3929		ME3	2.0947	2.2393	2.4394	2.5464	2.8021
20-s	0.005%	5.3322	ME1	−17.8068	−9.5419	−7.2562	−4.8115	−3.5992	10.1532	ME1	18.3394	9.9398	7.6740	5.4435	4.5299
			ME2	0.3397	0.2821	0.2882	0.4019	0.5596		ME2	3.0755	2.1117	2.1666	2.5271	2.8374
			ME3	0.5340	0.4639	0.6157	0.9473	1.2592		ME3	2.2664	1.7445	1.9529	1.9189	2.0788
	0.01%	5.3562	ME1	−16.4046	−8.6377	−6.5612	−4.3376	−3.2101	10.1639	ME1	16.9502	9.0284	6.9857	5.0049	4.2219
			ME2	0.6035	0.5473	0.5600	0.6682	0.8417		ME2	2.9903	2.1008	2.1818	2.5643	2.9096
			ME3	0.8043	0.7385	0.8828	1.2168	1.5292		ME3	2.3152	1.8216	2.0152	2.0450	2.2400
	0.02%	5.4036	ME1	−12.8619	−7.0131	−5.2494	−3.4318	−2.4897	10.1854	ME1	13.3635	7.4026	5.6999	4.2129	3.6884
			ME2	1.1378	1.0778	1.0905	1.2102	1.3818		ME2	2.8324	2.2176	2.3262	2.7620	3.1219
			ME3	1.3435	1.2746	1.4297	1.7565	2.0717		ME3	2.5068	2.0651	2.2905	2.3864	2.6243

Notes: ME and RMSE are the mean error and root mean-squared error, respectively, of the volatility estimates in annualized standard deviation in percentage. MD5 is the Heston model with price jumps and the average true daily integrated volatility is around 40%. ME1 is the ACD-ICV method of [Tse and Yang \(2012\)](#). ME2 is ME1 with δ^2 replaced by V_D , the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in [Tse and Dong \(2014\)](#).

Table 8. Comparison of the root mean-squared error (RMSE) of daily volatility estimates of the realized kernel (RK) and Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) methods.

RK	ACD-ICV	Avg. Sampling Frequency (ACD-ICV)				
		1-min	3-min	5-min	10-min	15-min
3.9198	ME1	8.1955	4.7603	3.8640	3.2027	3.0690
	ME2	1.9321	1.6430	1.7542	2.1559	2.4848
	ME3	1.4483	1.3958	1.5143	1.4755	1.6280

Notes: RMSE are the root mean-squared error of the volatility estimates (over all simulation models) in annualized standard deviation in percentage. ME1 is the ACD-ICV method of Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D , the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

Our modified ACD-ICV method ME3 consistently produces lower RMSE over all models and parameter set-ups than the RK estimates. Its performance is robust over a wide range of sampling frequencies of up to 15 min. It also outperforms the TRV estimates with subsampling, except possibly for MD5, in which case their performances are comparable.¹⁹

6. Conclusions

We propose an easy-to-use time-transformation method to implement the BTS scheme at a prespecified average sampling frequency. Using 40 stocks from the NYSE, we perform normality test to the jump-adjusted daily and weekly BTS returns. Our results show that stock prices can be considered discrete observations from a continuous-time jump-diffusion process. The BTS scheme performs better than other sampling methods in yielding iid Gaussian returns, and it also performs better than the CTS and TTS schemes in estimating daily realized volatility using the TRV method. We also show the superiority of the BTS durations over the price durations in estimating the daily integrated volatility using the ACD-ICV method. Our modified ACD-ICV estimate, ME3, which models the high-frequency BTS durations using the ACD model, performs the best in reporting smaller RMSE values.

Finally, we note that there are other possible applications of the BTS scheme. For example, we can estimate the conditional instantaneous volatility or intraday integrated volatility (such as over 30 min intervals) by modeling the high-frequency BTS durations using the ACD model. Moreover, we can estimate the integrated quarticity ($IQ = \int_0^1 \sigma^4(\tau) d\tau$) by using a method similar to that implemented in this paper. That is, we first obtain transactions with approximately equal quarticity increments. The conditional instantaneous quarticity or integrated quarticity can be estimated further by modeling clustering of the corresponding durations using the ACD model. Please refer to Jacod and Rosenbaum (2013) and Andersen et al. (2014), among others, for theories and empirical applications of the asset price quarticity.

Supplementary Materials: The following are available online at www.mdpi.com/econometrics/2225-1146/5/4/51/S1, Figure S1: 1-min Realized Volatility and number of transactions, 2010–2013. Table S1: Proportion of detected jumps. Table S2: Rejection proportion of the normality hypothesis for no-jump returns under different sampling schemes. Table S3: Rejection proportion of the normality hypothesis for returns under different sampling schemes. Table S4: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the CTS scheme. Table S5: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the TTS scheme. Table S6: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the BTS scheme. Table S7: ME and RMSE of daily volatility estimates using the TRV method with subsampling under the CTS scheme. Table S8: ME and RMSE of daily volatility estimates using the TRV method with subsampling under the TTS scheme. Table S9: ME and RMSE of daily volatility using the TRV

¹⁹ MD5 is a model with price jumps, and the TRV method is constructed to be robust to price jumps.

method with subsampling under the BTS scheme. Table S10: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD2. Table S11: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD3. Table S12: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD4.

Acknowledgments: Yingjie Dong acknowledges support provided by the Fundamental Research Funds for the Central Universities in University of International Business and Economics (UIBE) (CXTD8-01).

Author Contributions: Yingjie Dong and Yiu-Kuen Tse contributed equally to this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Jump Detection Procedure

We remove price jumps using the sequential jump-adjustment procedure of Andersen et al. (2010) with some minor modifications. The test statistic Z_t for day t is computed as

$$Z_t = \sqrt{|\mathcal{H}_t|} \left[\frac{\ln V_{Rt} - \ln V_{Bt}}{((\mu_1^{-4} + 2\mu_1^{-2} - 5)Q_{Tt}BV_t^{-2})^{\frac{1}{2}}} \right],$$

where $\mu_1 = \sqrt{2/\pi}$, V_{Rt} is the realized volatility, V_{Bt} is the realized bipower volatility and Q_{Tt} is the realized tripower quarticity, and \mathcal{H}_t is the grid on day t . The realized tripower quarticity Q_{Tt} is computed as

$$Q_{Tt} = \frac{1}{|\mathcal{H}_t|} \left[\mu_{\frac{4}{3}}^{-3} \sum_{j=2}^{|\mathcal{H}_t|-1} |r_{i,-}|^{\frac{4}{3}} |r_i|^{\frac{4}{3}} |r_{i,+}|^{\frac{4}{3}} \right],$$

where $r_i = Y_{i,+} - Y_i$ for $Y_{i,+}, Y_i \in \mathcal{H}_t$ and $\mu_{\frac{4}{3}} = 2^{\frac{2}{3}}\Gamma(7/6)/\Gamma(1/2)$. Under the condition of no jumps, Z_t is approximately standard normal. At the significance level α , a jump is considered to be significant if $Z_t > z_{1-\alpha}$, where z_α is the α -quantile of the standard normal distribution. When $Z_t > z_{1-\alpha}$, we delete the return that has the maximum absolute value among all returns and repeat the jump test again for the remaining returns. This recursive procedure stops when $Z_t \leq z_{1-\alpha}$.

Appendix A.2. Computation of the BT Time-Transformation Function

To compute the BT transformation function $Q(t_k) = N_{t_k}/N_{t_k}$, where $N_{t_k} = \sum_{i=1}^k V_i$ for $k = 1, \dots, K$, we need to calculate the intraday integrated volatility V_i . To alleviate the effects of market microstructure noise and price jumps we adopt the Tripower Realized Volatility (TRV) method of Barndorff-Nielsen et al. (2006). Suppose there are S subsampling grids and the s th subgrid is denoted by \mathcal{H}_s . Under subgrid \mathcal{H}_s , for trading day d , denote the returns by $\{r_{d,1}^{(s)}, r_{d,2}^{(s)}, r_{d,3}^{(s)}, \dots, r_{d,n_s}^{(s)}\}$. We define

$$V_{Td,i}^{(s)} = \zeta_{\frac{2}{3}}^{-3} |r_{d,i-1}^{(s)}|^{\frac{2}{3}} |r_{d,i}^{(s)}|^{\frac{2}{3}} |r_{d,i+1}^{(s)}|^{\frac{2}{3}},$$

for $i = 2, \dots, n_s - 1$, where $\zeta_k = 2^{\frac{k}{2}}\Gamma((k+1)/2)/\Gamma(1/2)$ for $k > 0$ and $\Gamma(\cdot)$ is the gamma function. Moreover, $V_{Td,1}^{(s)} = V_{Td,2}^{(s)}$ and $V_{Td,n_s}^{(s)} = V_{Td,n_s-1}^{(s)}$. Since the estimate $V_{Td,i}^{(s)}$ may have large fluctuations, we group every g $V_{Td,i}^{(s)}$ together to obtain the estimated intraday integrated volatility at 10-min average sampling frequency. Thus, the p th intraday integrated volatility on day d under subgrid \mathcal{H}_s is $V_{d,p}^{(s)} = \sum_{i=p(g-1)+1}^{pg} V_{Td,i}^{(s)}$. The corresponding end point of $V_{d,p}^{(s)}$ is the end point of $r_{d,pg}^{(s)}$ and is denoted by $t_{d,p}^{(s)}$. Finally, the estimated p th intraday integrated volatility $V_{d,p}$ and its corresponding end point $t_{d,p}$ can be obtained by taking the average over all subsamples. That is, $V_{d,p} = \frac{1}{S} \sum_{s=1}^S V_{d,p}^{(s)}$ and $t_{d,p} = \frac{1}{S} \sum_{s=1}^S t_{d,p}^{(s)}$. $V_{d,p}$ and $t_{d,p}$ pooled over all m trading days form V_K for $k = 1, \dots, K$ and \mathcal{H}_V , respectively, for the computation of the time-transformation function $Q(t)$. The number of subsamples S is selected to obtain subgrids of approximately 1 min sampling frequency.

References

- Aït-Sahalia, Yacine, and Lorian Mancini. 2008. Out of sample forecasts of quadratic variation. *Journal of Econometrics* 147: 17–33.
- Andersen, Torben G., Tim Bollerslev, and Dobrislav Dobrev. 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics* 138: 125–80.
- Andersen, Torben G., Tim Bollerslev, Per Frederiksen, and Ørregaard Nielsen. 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics* 25: 233–61.
- Andersen, Torben G., Dobrislav Dobrev, and Ernst Schaumburg. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169: 75–93.
- Andersen, Torben G., Dobrislav Dobrev, and Ernst Schaumburg. 2014. A robust neighborhood truncation approach to estimation of integrated quarticity. *Econometric Theory* 30: 3–59.
- Bajgrowicz, Pierre, Olivier Scaillet, and Adrien Treccani. 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science* 62: 2198–217.
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2: 1–37.
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics* 4: 1–30.
- Barndorff-Nielsen, Ole E., Neil Shephard, and Matthias Winkel. 2006. Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and Their Applications* 116: 796–806.
- Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76: 1481–536.
- Boudt, Kris, Christophe Croux, and Sébastien Laurent. 2011. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance* 18: 353–67.
- Christensen, Kim, Roel Oomen, and Mark Podolskij. 2010. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics* 159: 74–98.
- Dacorogna, Michael M., Ulrich A. Müller, Robert J. Nagler, Richard B. Olsen, and Olivier V. Pictet. 1993. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12: 413–38.
- Dambis, Karl E. 1965. On the decomposition of continuous submartingales. *Theory of Probability & Its Applications* 10: 401–10.
- Dong, Yingjie, and Yiu-Kuen Tse. 2017. On estimating market microstructure noise variance. *Economics Letters* 150C: 59–62.
- Dubins, Lester E., and Gideon Schwarz. 1965. On continuous martingales. *Proceedings of the National Academy of Sciences of the United States of America* 53: 913.
- Engle, Robert F., and Jeffrey R. Russell. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66: 1127–62.
- Fernandes, Marcelo, and Joachim Grammig. 2006. A family of autoregressive conditional duration models. *Journal of Econometrics* 130: 1–23.
- Hasbrouck, Joel. 1999. The dynamics of discrete bid and ask quotes. *The Journal of Finance* 54: 2109–42.
- Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6: 327–43.
- Huang, Xin, and George Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3: 456–99.
- Jacod, Jean, and Mathieu Rosenbaum. 2013. Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics* 41: 1462–84.
- Lee, Suzanne S., and Jan Hannig. 2010. Detecting jumps from Lévy jump diffusion processes. *Journal of Financial Economics* 96: 271–90.
- Lee, Suzanne S., and Per A. Mykland. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies* 21: 2535–63.

- Lilliefors, Hubert W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* 62: 399–402.
- Mykland, Per A. 2012. A Gaussian calculus for inference from high frequency data. *Annals of Finance* 8: 235–58.
- Nowman, Khalid B. 1997. Gaussian Estimation of Single-Factor Continuous Time Models of The Term Structure of Interest Rates. *The Journal of Finance* 52: 1695–706.
- Oomen, Roel C. A. 2006. Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics* 24: 219–37.
- Peters, Remco T., and Robin G. De Vilder. 2006. Testing the continuous semimartingale hypothesis for the S&P 500. *Journal of Business & Economic Statistics* 24: 444–54.
- Tse, Yiu-Kuen, and Thomas Tao Yang. 2012. Estimation of high-frequency volatility: An autoregressive conditional duration approach. *Journal of Business & Economic Statistics* 30: 533–45.
- Tse, Yiu-Kuen, and Yingjie Dong. 2014. Intraday Periodicity Adjustments of Transaction Duration and Their Effects on High-Frequency Volatility Estimation. *Journal of Empirical Finance* 28: 352–61.
- Zhou, Bin. 1998. F-consistency, De-volatilization and normalization of high-frequency financial data. In *Nonlinear Modelling of High Frequency Financial Time Series*. Edited by Dunis, Christian and Bin Zhou. New York: Wiley, pp. 109–23.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).