26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Realised Volatility Estimation Shortcuts: An Empirical Analysis

David E. Allen[a], Chialin Chang[b], Kok Haur Ng[c] and Shelton Peiris[d]

[a]School of Mathematics and Statistics, the University of Sydney, NSW, Australia, School of Business and Law, Edith Cowan University, WA, Australia, and Department of Finance, Asia University, Taiwan.
[b] Department of Applied Economics and Department of Finance, National Chung Hsing University, Taichung, Taiwan.
[c]Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya.
[d]School of Mathematics and Statistics, the University of Sydney. NSW, Australia.

Email: profallen2007@gmail

Abstract: This paper features an empirical evaluation of Yang and Zhang's (2000) short-cut method for calculating drift-independent realised volatility (RV), based on high, low, open, and close prices using a historical time series. They suggest that this method is unbiased in the continuous limit, independent of the drift, and consistent when faced with opening price jumps. Souto and Moradi (2024) published Python script to implement this method and claimed that this was the first open-source code made available, that automates its estimation from high-frequency intraday stock data. We covert their code to R script and undertake an evaluation of the method using high-frequency hourly data on Bitcoin, originally sourced from Binance, for a period from January 1, 2018, to May 29, 2025. The method is benchmarked against the bipower variation method (BPV) of Barndorff-Nielsen and Shephard (2004a, 2004b). The ordinary least squares regression analysis reveals a close relationship between the methods. We further explore the persistence of volatility estimates by regressing the absolute values of daily Bitcoin returns on lags of themselves. The results suggest significant persistence out to seven daily lags. This finding supports the range enhanced GARCH model for modelling cryptocurrency of Fiszeder et al. (2024). (The R code developed can be obtained directly from the authors).

Keywords: Short-cut volatility estimates, Yang and Zhang (2000), Realised BiPower Covariance, R code.

1. CONTENT OF PAPER

The paper features an empirical assessment of the Yang and Zhang (2000) short-cut method for calculating drift-independent realised volatility, based on high, low, open, and close prices using a historical time series. Souto and Moradi (2024) provided open-source Python script to implement the method, which we converted into R code, and then we applied the analysis in RStudio.

The Yang and Zhang method utilizes more information: unlike traditional methods that primarily rely on closing prices (e.g., close-to-close volatility), this estimator incorporates daily high, low, open, and close prices. It is drift-independent, and thus it is suggested that the estimator provides a more robust measure of volatility. It is designed to account for "opening price jumps" and provides an efficient and precise volatility estimates given the available data.

An alternative approach is to use higher frequency intraday data to estimate volatility, so called 'realised volatility' measures. Some of the benefits in contrast to daily, or lower frequency data, include the capturing of intraday dynamics. In an intraday sample you have more observations which should improve accuracy, and the estimates should be less noisy. The method should improve forecasting accuracy and pick up the impact of news events leading to changes in volatility more readily than end-of-day measures.

On the other hand, realised measures may be more susceptible to market microstructure noise such as bid-ask bounce, discrete price quotes, trading pauses etc. So, simply summing squared returns at the highest possible frequency can lead to biased estimates due to these frictions.

We required a benchmark to assess the effectiveness of the Yang and Zhang (2000) method, so we choose the Realised BiPower Covariance (rBPCov) method, introduced by Barndorff-Nielsen and Shephard in (2004a). This is regarded as being the preferred method for calculating realised volatility. This procedure has several advantages: it is robust to "jumps" in asset prices. Standard realised covariance methods estimate the *total* quadratic covariation of asset prices (which includes both the continuous diffusion component and any jump components), rBPCov specifically targets and consistently estimates the quadratic covariation of the continuous component of the price process. The core idea behind bipower variation (BPV) is to utilize products of absolute returns (or sums of absolute returns) over adjacent high-frequency intervals.

The paper is divided into four sections, following this introduction, the data and research method are introduced in section two, section three presents the results and section four concludes.

2. DATA AND RESEARCH METHOD

2.1 Data Set

The data used was Bitcoin historical data (2018-2025), in hourly observations, in US dollar terms (BTC/USDT), from January 1, 2018, to May 29, 2025, providing 64,823 datapoints (A feature of Bitcoin is that it trades 24 hours per day). The data was sourced from the Binance API and made available on Kaggle at the following webaddress:https://www.kaggle.com/datasets/novandraanugrah/bitcoin-historical-datasets-2018-2024?resource=download&select=btc_15m_data_2018_to_2025.csv.

Table 1 provides a descriptive summary of the Bitcoin hourly dataset converted into 2701 daily continuously compounded returns. The mean and median returns are positive at around 0.07 percent a day, there is very pronounced volatility, given that the standard deviation is 0.036 and the coefficient of variation is a massive 46.705. There is negative skewness of -1.1456 and very pronounced excess kurtosis of 17.089. A plot of the daily Bitcoin returns is provided in Figure 1.

Table 1: Descriptive statistics daily continuously compounded Bitcoin returns

Mean	Median	Minimum	Maximum
0.00076720	0.00070766	-0.50261	0.17845
Standard Deviation	Coefficient of Variation	Skewness	Excess Kurtosis
0.035832	46.705	-1.1456	17.089

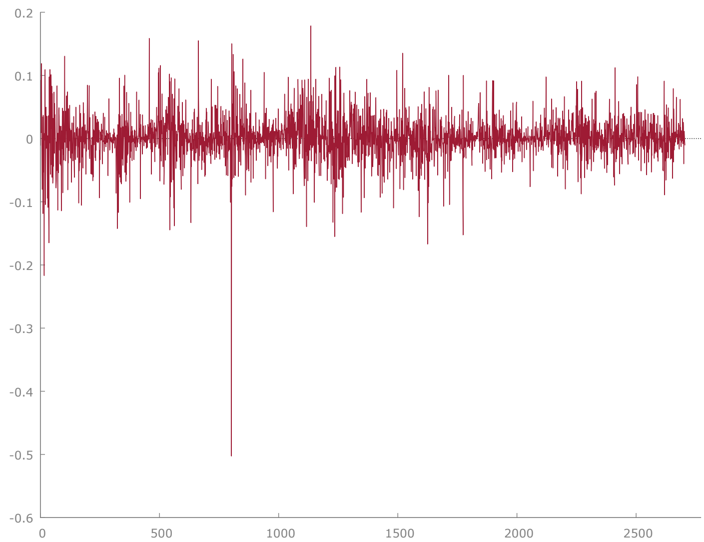

Figure 1: Plot of daily Bitcoin returns.

2.2 Research Method

2.1.1. Yang-Zhang Estimator

The formula for the Yang-Zhang volatility (variance) is given as:

$$\sigma_{YZ}^2 = \sigma_{OC}^2 + k\sigma_{CO}^2 + (1-k)\sigma_{RS}^2. \tag{1}$$

where σ_{YZ}^2 is the Yang-Zhang daily volatility (variance), σ_{OC}^2 is the variance of open-to-close returns, σ_{CO}^2 is the variance of close-to-open returns (overnight returns), σ_{RS}^2 is the Rogers-Satchell volatility (variance) and k is a weighting factor. We take the logs of all the series: $o_t = \ln(O_t)$ (log of open price on day t), $h_t = \ln(H_t)$ (log of high price on day t), $l_t = \ln(L_t)$ (log of low price on day t), and $c_t = \ln(C_t)$ (log of close price on day t).

Open-to-Close volatility (σ_{OC}^2) is a component captures the volatility within the trading day. For N days of data, assuming zero drift, is given by:

$$\sigma_{OC}^2 = \frac{1}{N}\sum_{i=1}^{N}(c_i - o_i)^2. \tag{2}$$

Close-to-Open volatility (σ_{CO}^2) captures the over-night volatility, again assuming no drift, we have:

$$\sigma_{CO}^2 = \frac{1}{N}\sum_{i=1}^{N}(o_i - c_{i-1})^2. \tag{3}$$

Rogers-Satchell Volatility (σ_{RS}^2) estimator is drift-independent and robust to opening price jumps, the average estimator is given by: $\sigma_{RS}^2 = \frac{1}{N}\sum_{i=1}^{N}[(h_i - c_i)(h_i - o_i) + (l_i - c_i)(l_i - o_i)]$. (4)

The weighting factor k was derived by Yang and Zhang (2000) as the optimal factor that minimizes the variance of their estimator. This is given by:

$$k = \frac{0.34}{1.34 + \frac{N+1}{N-1}},$$ where N is the number of days in the estimation period.

For N trading days, the Yang-Zhang estimator for the daily variance is:

$$\sigma_{YZ}^2 = \frac{1}{N}\sum_{i=1}^{N}(c_i - o_i)^2 + \frac{0.34}{1.34+\frac{N+1}{N-1}}\frac{1}{N}\sum_{i=1}^{N}(o_i - c_{i-1})^2 + \left\{\left(1 - \frac{0.34}{1.34+\frac{N+1}{N-1}}\right)\left(\frac{1}{N}\right)\sum_{i=1}^{N}[(h_i - c_i)(h_i - o_i) + (l_i - c_i)(l_i - o_i)]\right\}.$$
(6)

The final Yang-Zhang volatility is the square root of this variance: $\sigma_{YZ} = \sqrt{\sigma_{YZ}^2}$.

2.1.2. Barndorff-Nielsen and Shephard Realised Bipower Variance/Covariance

Barndorff-Nielsen and Shephard (2004a, 2004b), provide a recognised way of estimating the realised volatility of a set of high frequency (within the day) observations for a financial time-series that is robust to jumps. Their method provides a useful benchmark for testing the effectiveness of Yang-Zhang (2000) measure. We used the R library package 'highfrequency' by Boudt et al. (2022) to undertake the analysis on the Bitcoin data set and then uses the daily volatility estimates as a benchmark for testing the veracity of Yang-Zhang (2000) method.

Barndorff-Nielsen and Shephard (2004a, 2004b) proceed by letting $(\boldsymbol{Y}_t = Y_t^1, \dots, Y_t^d)'$ be a d-dimensional vector of logarithmic asset prices. They assume that $\boldsymbol{Y}_t$ follows a general jump-diffusion semi-martingale process:

$$Y_t = \int_0^t a_u du + \int_0^t \sigma_u dW_u + \sum_{j=1}^{N_t} J_j. \quad (7)$$

In the above-expression, a_u is the d-dimensional drift process, σ_u is a $d \times d$ matrix-valued spot volatility process, representing the instantaneous covariance of the continuous component. $\boldsymbol{W}_u$ is a d-dimensional standard Brownian motion, and N_t is a counting process representing the number of jumps up to time $t. J_j$ are the jump sizes at the jump times.

The **realised quadratic covariation** over a period $[0, T]$, based on M intraday observations for returns $\Delta_i \boldsymbol{Y} = \boldsymbol{Y}_{i\Delta t} - \boldsymbol{Y}_{(i-1)\Delta t}$ (where $\Delta t = T/M$), is given by:

$$RV_T = \sum_{i=1}^{M} \Delta_i Y (\nabla_i Y)^t \xrightarrow{p} \int_0^T \Sigma_u du + \sum_{0<u<T} \Delta Y_u (\Delta Y_u)', \quad (8)$$

where $\Sigma_u = \sigma_u \sigma_u'$ is the instantaneous covariance matrix of the continuous component, and the second term is the sum of squared jump sizes (the jump component of quadratic variation). This shows that RV_T estimates the total quadratic covariation, including jumps.

Barndorff-Nielsen and Shephard (2004a and 2004b), further demonstrate how to isolate the continuous component and how it is robust to jumps, via the realised bipower covariation. Further, under regularity conditions, this realised bipower covariation converges in probability to the integrated covariance of the continuous component. We estimate this benchmark measure, applied to the Bitcoin data, using the 'highfrequency' R package.

3. RESULTS

Descriptive summaries of the Yang-Zhang estimates of RV are shown in Table 2 together with the Barndorff-Nielsen and Shephard RV benchmark.

Table 2: Descriptive statistics Yang-Zhang RV and Barndorff-Nielson and Shephard RV benchmark

Yang-Zhang RV

Mean	Median	Minimum	Maximum
0.00015919	7.6161e-05	1.1269e-06	0.0096028
Standard Deviation	Coefficient of Variation	Skewness	Excess Kurtosis
0.00036679	2.3042	12.963	252.83

Barndorff-Nielson and Shephard RV benchmark

Mean	Median	Minimum	Maximum
0.0013513	0.00062817	9.5734e-06	0.081530
Standard Deviation	Coefficient of Variation	Skewness	Excess Kurtosis
0.0031696	2.3457	13.347	272.57

It can be seen in Table 2 that the results of the two sets of estimations are very similar. The main difference being that we programmed the Yang-Zhang estimator in R in real numbers whereas the estimates of RV in the 'highfrequency' R package are in percentage terms, so the values of the estimations differ by a factor of a 100. If we convert the Yang-Zhang RV estimations into percentage terms, the mean Yang-Zhang estimate of RV is 0.0159 whereas the mean estimate of the benchmark RV in the 'highfrequency' package is 0.00135. The coefficient of variation in the former is 2.304 and in the latter 2.346, the skewness in Yang-Zhang is 12.963 and in Barndorff-Nielsen and Shephard BPV is 13.347, and finally the two values of the excess kurtosis are 252.83 and 272.57. Plots of the two different estimates of Bitcoin daily RV are shown in Figure 2.

The two distributions are right skewed so have some extreme values in the right tail and their means are greater than their medians. The Barndorff-Nielsen and Shephard bipower variation method, by its construction, aims to isolate and estimate the continuous component of volatility, effectively "ignoring" or making itself robust to the presence of jumps. Since jumps are a major driver of fat tails in financial data, the theoretical and empirical distribution of the bipower variation estimator (as an estimate of continuous volatility) will tend to have lighter tails than the Yang-Zhang method. The latter will still incorporate the impact of intraday jumps through its reliance on the high-low range, which will make its distribution appear to have fatter tails when jumps are present. This effect is accentuated by the fact that Bitcoin trades continuously, so all jumps are 'within the day'.

We assessed the similarity between the two sets of estimates by standardising the two sets of estimates by dividing them by their standard deviation. We then ran an ordinary least squares (OLS) regression analysis of the benchmark BPV standardised estimates on the Yang and Zhang standardised estimates.

The results of the ordinary least squares regression reported in Table 3 suggest that there is a very close relationship between the two RV estimates. The slope coefficient is 0.979 which is significant at the 1% level. The R-squared is 0.956 and the F statistic is 62076, which is highly significant. This suggests that RV estimates based on high and low prices within the day, daily closing prices, and close to open prices, capture 95% of the BPV estimate of RV, at least in the case of this hourly sample of Bitcoin prices.

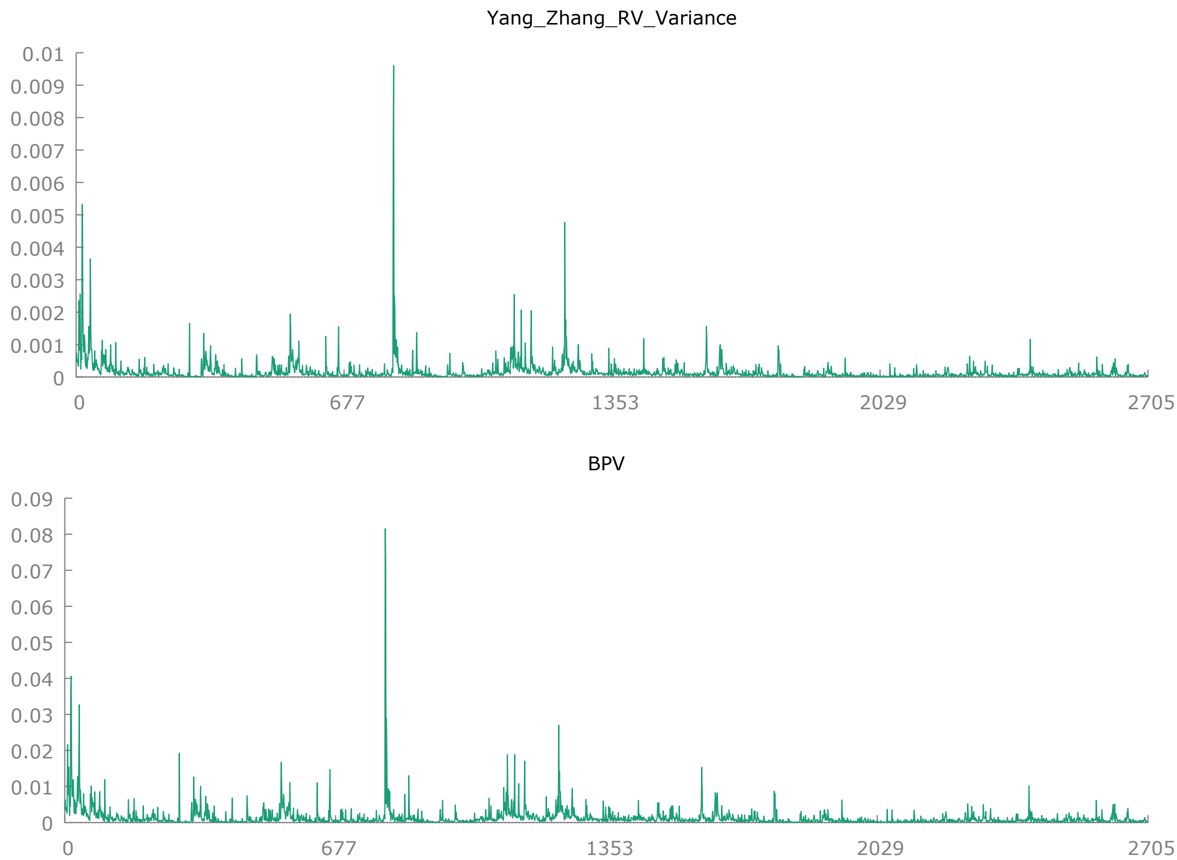

Figure 2: Plots of RV estimates for Bitcoin daily RV

Table 3: OLS regression analysis of the benchmark BPV standardised RV estimates on the Yang and Zhang standardised RV estimates

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
Constant	0.00000	0.00392827	−5.565e-14	1.0000	
Yang_Zhang_RV_Variance	0.978915	0.00392900	249.2	<0.0001	***

Mean dependent var	0.000000		S.D. dependent var	1.000000
Sum squared residual	112.8278		S.E. of regression	0.204308
R-squared	0.958274		Adjusted R-squared	0.958258
F (1, 2703)	62076.36		*P*-value (F test)	0.000000

Note: *** indicates significance at the 1% level.

There are some peculiar features of Bitcoin, in that it is traded 24 hours per day, 365 days per year. This means that aspects of traditional asset markets which feature opening/closing auctions or overnight gaps when the market is closed are not considerations. Bitcoin's continuous trading means that price discovery is constant and not subject to interruptions.

This suggests that volatility patterns might well be smoother and not display spikes at the open and the close that are a feature of traditional markets. However, the market is unregulated and Bitcoin is traded on numerous independent exchanges worldwide. This creates highly fragmented liquidity, with different order books and varying prices across platforms at any given moment. On the other hand, this means that arbitrage can be continuous, which should promote price convergence. The dominance of retail investors may mean that Bitcoin is over-exposed to news cycles, social media, pump and dump schemes, and the lack of circuit-breakers may account for Bitcoin’s relatively high volatility.

Granger and Ding (1995) mentioned the ‘Taylor effect’, in reference to Taylor (1986, Chapter 2), and the observation that persistence in volatility can be captured by the autocorrelation of the absolute return series. Takaishi and Adachi (2018) recently explored this feature of Bitcoin returns and reported that the Taylor effect is present in Bitcoin time series. They noted that: the power *d* that maximizes the autocorrelation is time lag dependent, that exchange rate returns show a daily seasonality in the Taylor effect, plus they observed no evidence of seasonality. To further explore this matter, Table 4 reports an OLS regression of the absolute value of the daily Bitcoin return regressed on ten lags of itself.

Table 4: Absolute value of daily Bitcoin return regressed on 10 lags of itself.

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.0105661	0.00107099	9.866	<0.0001	***
absret_1	0.0999311	0.0193003	5.178	<0.0001	***
absret_2	0.0318748	0.0193945	1.643	0.1004	
absret_3	0.0271037	0.0193704	1.399	0.1619	
absret_4	0.115878	0.0193201	5.998	<0.0001	***
absret_5	0.0665995	0.0194122	3.431	0.0006	***
absret_6	0.0545868	0.0193649	2.819	0.0049	***
absret_7	0.0768649	0.0192671	3.989	<0.0001	***
absret_8	0.0465316	0.0193108	2.410	0.0160	**
absret_9	0.0164999	0.0192998	0.8549	0.3927	
absret_10	0.0154718	0.0192085	0.8055	0.4206	

Mean dependent var	0.023628		S.D. dependent var	0.026792
Sum squared resid	1.789787		S.E. of regression	0.025823
R-squared	0.074446		Adjusted R-squared	0.070997
F(10, 2684)	21.58838		P-value(F)	3.54e-39
Log-likelihood	6035.695		Akaike criterion	−12049.39

Note: *** and ** indicate significance at the 1% and 5% levels, respectively.

It can be seen in Table 4 that seven of the ten lags are significant with six significant at the 1% level, and the Adjusted R squared is 0.07. This suggests that Bitcoin volatility is very persistent.

4. CONCLUSION

We have explored the effectiveness of Yang and Zhang’s (2000) short-cut method for calculating drift-independent realised volatility, based on high, low, open, and close prices and applied the method to hourly Bitcoin prices. To benchmark the effectiveness of this approach we used the BPV method of calculating RV, of Barndorff-Nielson and Shephard (2004a, 2004b). The OLS results suggest the existence of a very close correspondence between the estimates produced by the two methods. Both the slope coefficient and the R squared of the regression are very close to 1, suggesting that investment analysts and portfolio risk modellers could use either measure, as estimates of volatility, in the case of Bitcoin, and that intraday observations do not confer a marked advantage, as daily observation from Yahoo finance that include, open, close, high and low prices, work equally well. This confirms the recent findings by Korkusuz et al. (2023) on G7 stock market indices,

We further explored the persistence in the volatility of daily Bitcoin prices using the absolute value of the return series and found significant evidence of persistence, in seven lags of daily absolute returns. Fiszeder et al. (2024) have demonstrated the superior forecasting performance of a GARCH model enhanced by range-based methods, as applied to Bitcoin, Ethereum Classic, Ethereum, and Litecoin and find that it forecasts variances, value at risk, and expected shortfall more accurately than the standard GARCH model. Our findings support this approach.

REFERENCES

Barndorff-Nielsen OE and Shepherd N (2004a). Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, (1) 1–37. https://doi.org/10.1093/jjfinec/nbh001

Barndorff-Nielsen OE and Shepherd N (2004b). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. Econometrica 72(3), 885–925.

Boudt K, Kleen O and Sjørup E (2022). Analyzing intraday financial data in R: The highfrequency package. Journal of Statistical Software 104(8), 1–36.

Fiszeder P, Małecka M and Molnár P (2024). Robust estimation of the range-based GARCH model: Forecasting volatility, value at risk and expected shortfall of cryptocurrencies. Economic Modelling 141, 106887 https://doi.org/10.1016/j.econmod.2024.106887

Granger, CWJ and Ding Z (1995). Some properties of absolute return: An alternative measure of risk. Annales d'Economie et de Statistique 40, 67–91.

Korkusuz B, Kambouroudis D and McMillan DG (2023) Do extreme range estimators improve realized volatility forecasts? Evidence from G7 Stock Markets. Finance Research Letters 55, 1-7 https://doi.org/10.1016/j.frl.2023.103992

Rogers LCG and Satchell SE (1991). Estimating variance from high, low and closing prices. The Annals of Applied Probability 1(4), 504–512.

Souto HG and Moradi A (2024). Yang & Zhang's realized volatility: Automated estimation in Python. Software Impacts 19, 100613. https://doi.org/10.1016/j.simpa.2024.100613

Takaishi T and Adachi T (2018). Taylor effect in Bitcoin time series. Economics Letters 172, 5–7.

Taylor SJ (1986) Modelling Financial Time Series. Wiley, Chichester.

Yang D and Zhang Q (2000). Drift-independent volatility estimation based on high, low, open, and close prices. The Journal of Business 73(3), 477–492.

Appendix

The R Code used to implement the Yang and Zhang (2000) RV estimation can be obtained directly from D.E. Allen email: profallen2007@gmail.com