Realized Cryptocurrency Volatility Forecasting

Junjie Hu*
WeiYu Kuo°
Wolfgang Karl Härdle*

*Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
°Department of International Business
National Chengchi University
RCVJ_Forecasting

Crytocurrency Market

- Unregulated and unlikely to be regulated
- Free market or chaotic market?
- An extremly volatile market for sure. Unpredictable volatility?

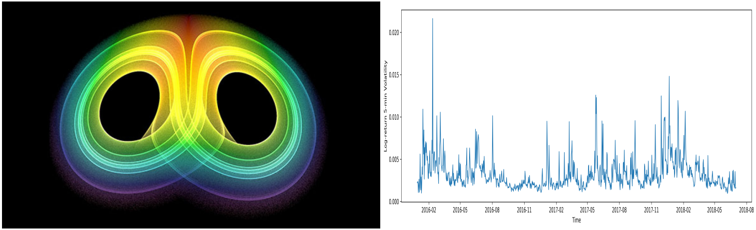

Figure 1: Chaos and Randomness

Distribution of Logarithmic Returns

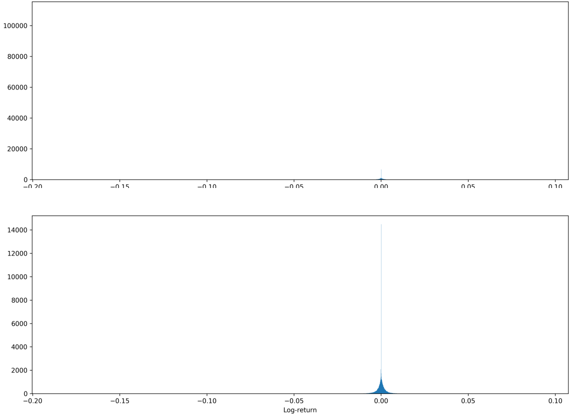

Figure 2: Histogram of 5-min freq. BTC log-return. Top panel: With zero returns; Bottom panel: without zero returns

Annualized Realized Variance of Major Indices

	AEX	DJI	FTSE	HSI	SPX	SSEC	BTC-G
count	4842	4704	4769	4645	4709	4508	965
mean	0.16	0.12	0.13	0.14	0.12	0.22	0.81
std	0.38	0.29	0.32	0.40	0.32	0.46	1.68
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.02	0.01	0.02	0.02	0.01	0.03	0.01
50%	0.05	0.04	0.05	0.05	0.04	0.08	0.29
75%	0.14	0.11	0.12	0.13	0.11	0.22	2.99
max	7.03	5.55	7.74	16.45	7.18	7.70	18.59

Table 1: Realized variance comparison between market indices (Oxford-man Realized Library) and Bitcoin

- Much larger mean, standard deviation and higher extreme values
- Overnight bias correction of RV for market indices (Bollerslev et al (2018))
- Cause of high RV. Volatility? Jump?

Program Trading

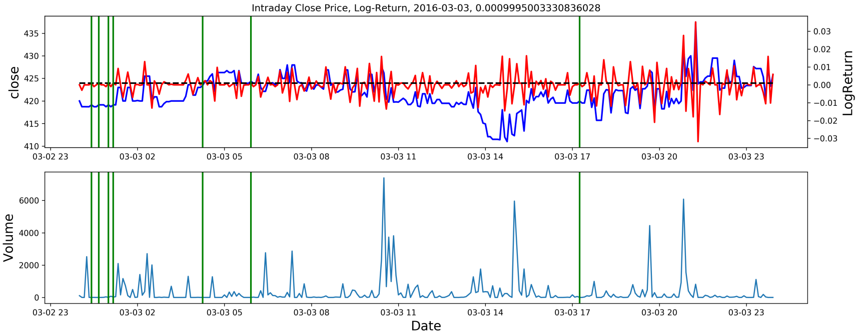

Figure 3: One day sample of 5-min freq BTC. Price and Log-return on top panel, Trading Volume on bottom panel

- Same log-return, repeating more than 100 times during 2016
- 24-hours trading capability, might cause more volatile
- Algorithm trading crash (1987 Stock Market Crash)

Realized Voaltilty and Jumps

- Volatility: A central role in finance, financial derivatives pricing (options, volatility swap), risk management
- ABDL(2001, 2003): "Realized Variance" (RV) from high-frequency data with good dynamics (lognormal, long-memory)
- Fleming et al.(2001, 2003): Economic value of realized volatilty timing
- Bollerslev et al.(2018): Similary risk (realized volatility) across different asset classes (commodities, currencies,equity, bonds)
- Forecasting on RV with continous component and jump component (BNS (2004), Corsi et al(2010))

Realized Volatility of Cryptocurrency

- Importance of RV on emerging market
- Rigorous research contribution to industries
- Limited researches on Realized Cryptocurrency Volatility

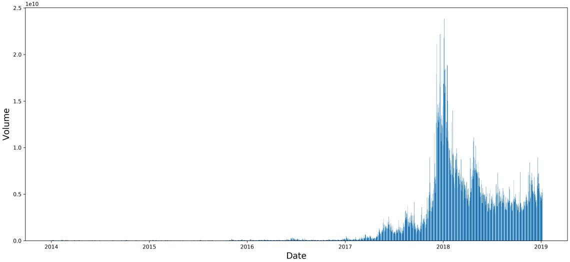

Figure 4: Daily Trading Volume of Bitcoin

Research Questions

- Investigation on realized cryptocurrency variance process
- Predictability of realized cryptocurrency volatility
- Analysis on realized jumps
- Economic value from a better *RV* forecast for investors?

Outline

1. Motivation ✓
2. Realized Variance
3. Jump Detection
4. HAR Forecasting Model
5. Realized Utility Evaluation
6. Summary

Continuous-Time Framework

- A general continous-time jump diffusion process:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), 0 \leq t \leq T \quad (1)$$

p: Logarithmic price
μ: Continous and locally bounded variation process
σ: Strictly positive stochastic volatility process with right continuous sample path
W: Brownian motion
κ: Size of a jumps
q: Counting measure for jumps

Realized Variance

- Logarithmic return at time $t + j\Delta$ can be noted as $r_{t+j\Delta}$, where Δ is the sampling frequency

$$r_{t+j\Delta} \stackrel{\text{def}}{=} p\,(t + j\Delta) - p\,(t + (j-1)\Delta) \tag{2}$$

- Daily Realized Variance RV_{t+1} for period $[t : t+1]$

$$RV_{t+1}(\Delta) \stackrel{\text{def}}{=} \sum_{j=1}^{1/\Delta} r^2_{t+j\Delta} \tag{3}$$

- RV converges to Quadratic Variation QV as Δ goes to 0

$$RV_{t+1}(\Delta) \xrightarrow{p} QV_{t+1} \tag{4}$$

Realized Variance

- Variance of $p(t)$ measured by QV

$$QV_{t+1} = \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leq t+1} \kappa^2(s) \tag{5}$$

- Hence, RV has

$$RV_{t+1}(\Delta) \xrightarrow{p} \underbrace{\int_t^{t+1} \sigma^2(s)ds}_{IV_{t+1}} + \underbrace{\sum_{t<s\leq t+1} \kappa^2(s)}_{J_{t+1}} \tag{6}$$

- Two dynamic components: Intergated Variance (hereafter IV_{t+1}) and jump component (hereafter J_{t+1})

Realized Bipower Variation (BNS(2004))

- Realized BiPower Variance (hereafter BPV_{t+1}) for period $[t, t+1]$

$$BPV_{t+1}(\Delta) \stackrel{\text{def}}{=} \frac{\pi}{2} \sum_{j=2}^{h/\Delta} |r_{t+j\Delta}||r_{t+(j-1)\Delta}| \tag{7}$$

- BPV converges to IV as Δ goes to 0

$$BPV_{t+1}(\Delta) \rightarrow \int_t^{t+1} \sigma^2(s)ds \tag{8}$$

- Finally, J_{t+1} for period $[t, t+1]$ estimated

$$\begin{aligned} RV_{t+1}(\Delta) - BPV_{t+1}(\Delta) &\xrightarrow{\text{p}} \sum_{t<s\leq t+1} \kappa^2(s) \\ J_{t+1} &\stackrel{\text{def}}{=} max\left\{RV_{t+1}(\Delta) - BPV_{t+1}(\Delta), 0\right\} \end{aligned} \tag{9}$$

Data Source

- DYOS &
  A free data source: https://www.cryptodatadownload.com/
- 5-min freq from 3 different exchanges, 913 trading days (01.01.2016-01.07.2018)&
  1-min freq from 1 exchange, 966 trading days (01.01.2016-23.08.2018)
- Realized Labrary, 31 indices realized variance and bipower variance

RV and *BPV* for Market Indices: FTSE

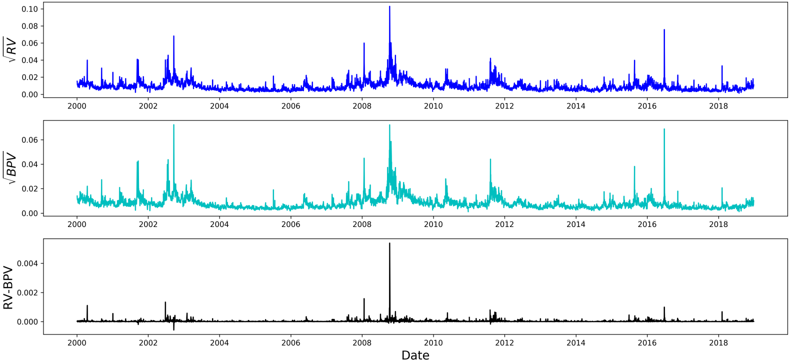

Figure 5: The Financial Times Stock Exchange 100 Index, 2000-2018

RV and *BPV* for Market Indices: SSEC

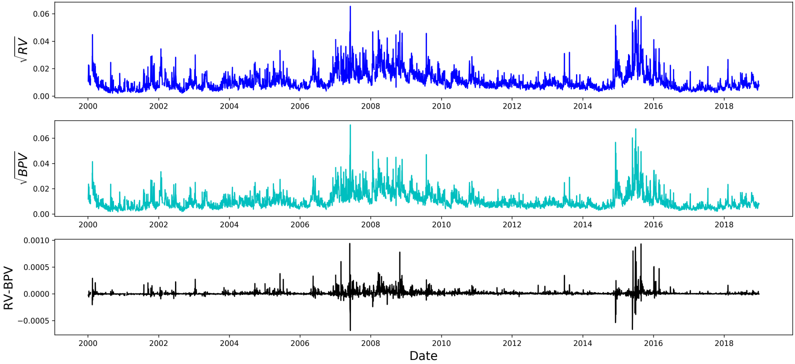

Figure 6: The Shanghai Stock Exchange 50 Index, 2000-2018

- Trading hour bias corrected by squared over-night prices (Bollerslev(2018))

ACF of $\log(RV)$: cryptocurrencies

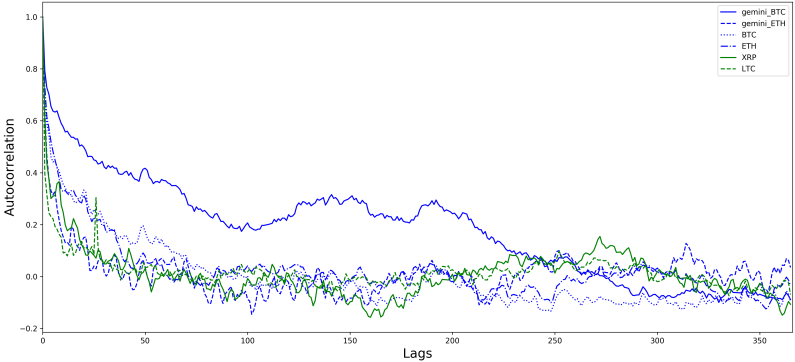

Figure 7: ACF of $\log(RV)$ decay, BTC-G, BTC-D, ETH-G, ETH-D, XRP, LTC

ACF of $\log(RV)$: Global Indices V.S BTC-G

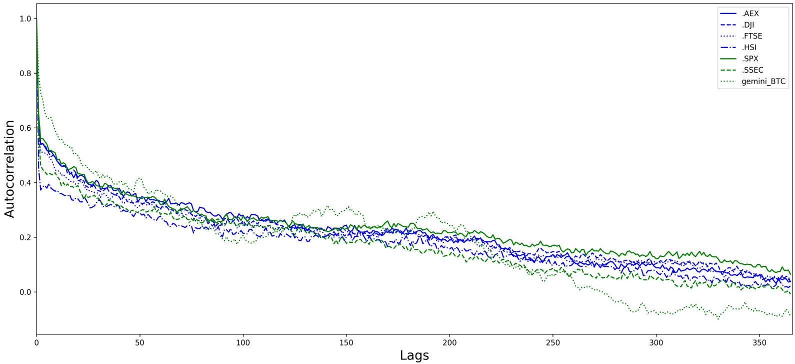

Figure 8: ACF of $\log(RV)$ decay: Comparison between 6 global market indices and BTC-G

Realized Variance Separation

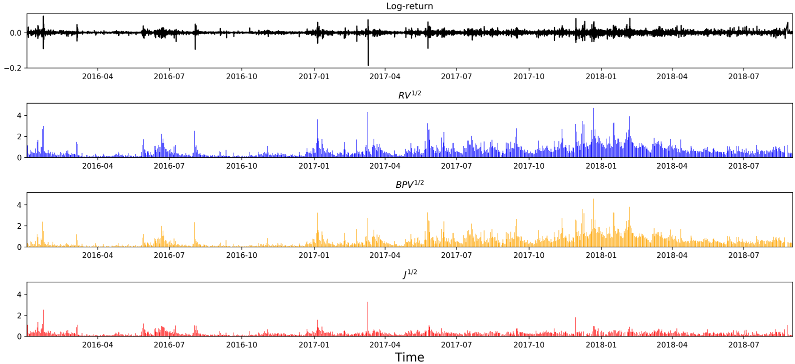

Figure 9: 5-min freq Log-Return of Bitcoin, Realized Volatility, Bipower Volatility and Jump Process

Why BPV Is Biased To Large Jumps

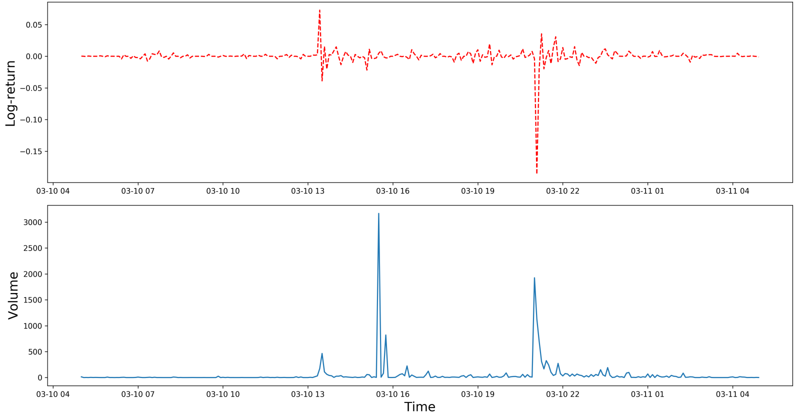

Figure 10: 5-min freq BTC on 10th, Mar 2017, Log-return; Trading Volume

Threshold Bipower Variation

- Threshold Bipower Variance, hereafter $TBPV_{t+1}$, for period $[t, t+1]$ (Mancini(2009))

$$
\begin{aligned}
TBPV_{t+1}(\Delta) \quad = \quad & \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta}||r_{t+(j-1)\Delta}| \\
& \cdot \mathbf{I}\left\{|r_{t+j\Delta}|^2 \leq \theta_{t+j\Delta}\right\} \\
& \cdot \mathbf{I}\left\{|r_{t+(j-1)\Delta}|^2 \leq \theta_{t+(j-1)\Delta}\right\}
\end{aligned}
$$

- Where $\theta_{t+j\Delta} = C_\theta^2 \cdot \widehat{V}_{t+j\Delta}$, and $\mu_1 = \sqrt{2/\pi}$
- BPV: Biased by big jumps, TBPV: Problematic with small jumps

Corrected Threshold Bipower Variation

- A modified TBPV(Corsi et al. (2010))

$$TBPV_{t+1}(\Delta) = \mu_1^{-2} \sum_{j=2}^{1/\Delta} Z_1(r_{t+j\Delta}, \theta_{t+j\Delta}) \cdot Z_1(r_{t+(j-1)\Delta}, \theta_{t+(j-1)\Delta}) \tag{10}$$

- Where $Z_\gamma(x,y) =$

$$\begin{cases} |x|^\gamma & , x^2 \leq y \\ \frac{1}{2\Phi(-c_\theta)\sqrt{\pi}}\left(\frac{2}{c_\theta^2}y\right)^{\frac{\gamma}{2}} \cdot \Gamma\left(\frac{\gamma+1}{2}, \frac{c_\theta^2}{2}\right) & , x^2 > y \end{cases} \tag{11}$$

- Where Φ and Γ

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{s^2}{2}} ds, \Gamma(\alpha, x) = \int_{x}^{+\infty} s^{\alpha-1} e^{-s} ds \tag{12}$$

Local Variance Estimation $\widehat{V}_t$

- a non-parametric filter of length $2L+1$ (Fan and Yao, 2003)

$$\widehat{V}_t^Z = \frac{\sum_{i=-L, i\neq -1,0,1}^{L} K(\frac{i}{L}) \cdot r_{t+i}^2 \cdot \mathbf{I}\{r_{t+i}^2 \leq c_\theta^2 \cdot \widehat{V}_{t+i}^{Z-1}\}}{\sum_{i=-L, i\neq -1,0,1}^{L} K(\frac{i}{L}) \cdot \mathbf{I}\{r_{t+i}^2 \leq c_\theta^2 \cdot \widehat{V}_{t+i}^{Z-1}\}}, Z = 1, 2, 3... \tag{13}$$

- $K(x)$: Gaussian Kernel. C_θ: immaterial coefficient, trade-off between effciency and bias
- Initial value $\widehat{V}_t^0 = \inf$; Iteration until converged
- Evaluation within each day, avoid using furture information

Threshold Jump

- Test statistics c-z for TJ_t (BNS(2004), Corsi et al.(2010))

$$c\text{-}z = \Delta^{-\frac{1}{2}} \frac{\{RV_t - TBPV_t\} RV_t^{-1}}{\sqrt{(\frac{\pi^2}{4} + \pi - 5) \max\{1, \frac{TTriPV_t}{TBPV_t^2}\}}} \tag{14}$$

- Significant threshold-Jump, hereafter TJ_{t+1}, for period [t,t+1]

$$TJ_{t+1} \stackrel{\text{def}}{=} \mathbf{I}(c\text{-}z > \Phi_\alpha) \cdot \max(RV_{t+1} - TBPV_{t+1}, 0) \tag{15}$$

- Threshold continuous process, hereafter TC_{t+1} for period $[t, t+1]$

$$TC_{t+1} \stackrel{\text{def}}{=} RV_{t+1} - TJ_{t+1} \tag{16}$$

Volatility Separation

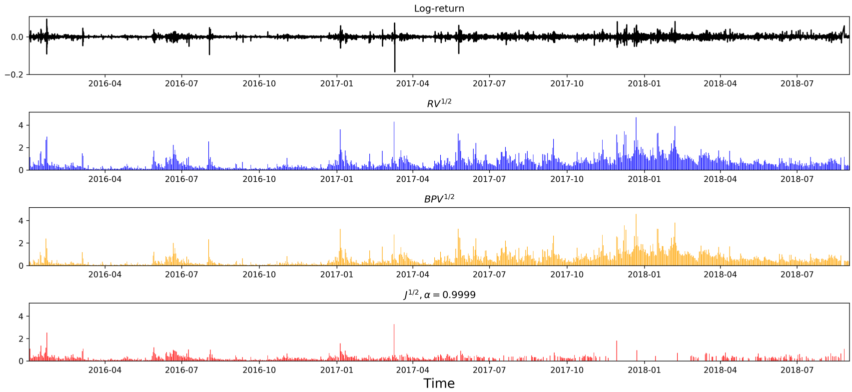

Figure 11: $J^{1/2}$ separation. Significant jumps separation using BPV. Confidence level $\alpha = 0.9999$

Threshold Volatility Separation

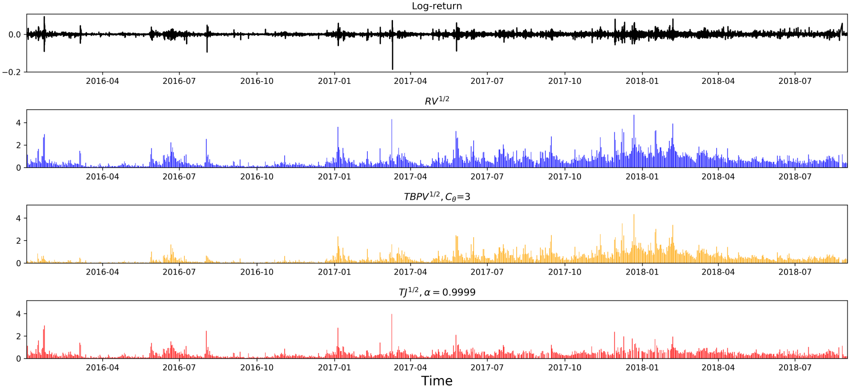

Figure 12: $TJ^{1/2}$ separation. Significant threshold-jumps separation using TBPV. Confidence level $\alpha = 0.9999$ and $c_\theta = 3$

Unconditional *RV* Distribution

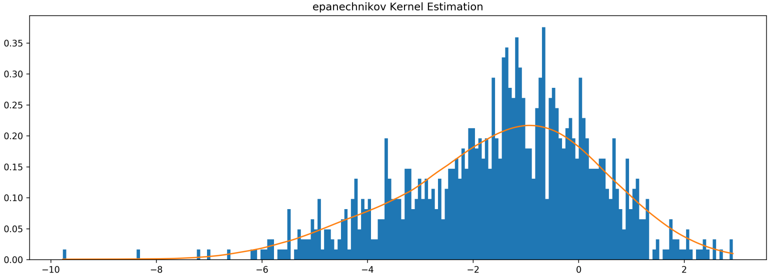

Figure 13: Histogram and Epanechinikov KDE of daily log(RV) (annualized) for 5-min freq BTC , bandwidth=1.8

- Close to log-normal distribution, similar to results from previous literatures

Summary Statistics

	$RV^{1/2}$	$\log(RV)$	BPV	TBPV	Jump(α)	TJump(α)
count	965	965	965	965	678	749
mean	0.68	-1.44	0.67	0.60	0.16	0.28
std	0.58	1.78	1.49	1.42	0.52	0.85
min	0.01	-9.77	0.00	0.00	0.00	0.00
5%	0.09	-4.75	0.00	0.00	0.00	0.01
50%	0.54	-1.24	0.19	0.15	0.07	0.11
95%	1.73	1.10	2.63	2.45	0.45	0.89
max	4.31	2.92	17.96	20.77	10.90	14.89
skewness	2.15	-0.54	5.39	6.54	14.83	11.72
kurtosis	7.03	0.45	38.85	62.78	275.28	173.30
ac(1)	0.69	0.79	0.55	0.51	0.12	0.09
ac(7)	0.42	0.64	0.21	0.21	0.07	0.05
ac(30)	0.30	0.42	0.16	0.17	0.01	-0.01
ac(100)	0.15	0.21	0.07	0.08	-0.01	-0.01

Table 2: Summary statistics for volatility related measures (annualized) from 5-min freq BTC, $ac(n)$: n-days autocorrelation. Confidence level $\alpha = 0.9999$

- More jumps are detected using *TJump* measure
- Long-memory and log-normal of $\log(RV)$

HAR Models

- HAR-RV

$$\widehat{RV}_{t,t+h} = \alpha + \widehat{CP}^{\top} \cdot \beta_C + \varepsilon_{t,t+h} \tag{17}$$

- HAR-CJ

$$\widehat{RV}_{t,t+h} = \alpha + \widehat{CP}^{\top} \cdot \beta_C + \widehat{JP}^{\top} \cdot \beta_J + \varepsilon_{t,t+h} \tag{18}$$

- Where $RV_{t_1,t_2} \stackrel{\text{def}}{=} \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} RV_t$
- $\widehat{CP} = (\widehat{C}_t, \widehat{C}_{t-7,t}, \widehat{C}_{t-30,t})^{\top}$, $\widehat{JP} = (\widehat{J}_t, \widehat{J}_{t-7,t}, \widehat{J}_{t-30,t})^{\top}$
- $\widehat{C}_t = \left\{C_t, TC_t, C_t^{Exp}, TC_t^{Exp}\right\}$, $\widehat{J}_t = \{J_t, TJ_t\}$
- $\widehat{RV} = \left\{RV, RV^{1/2}, \log(RV)\right\}$. Likewise for other variables.

Performance Evaluation

- R^2 of Mincer–Zarnowitz forecasting regressions
- Mean Squared Error (MSE)

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(RV_t - \widetilde{RV}_t\right)^2 \tag{19}$$

- Heteroskedasticity adjusted Root Mean Square Error (HRMSE) (Bollerslev and Ghysels(1996))

$$HRMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{RV_t - \widetilde{RV}_t}{RV_t}\right)^2} \tag{20}$$

- QLIKE loss function (Patton(2011))

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}\left(\log RV_t + \frac{\widetilde{RV}_t}{RV_t}\right) \tag{21}$$

In-Sample Fitting & Out-of-Sample Forecast

	HAR	HAR-CJ	HAR-TCJ	HAR-Exp-CJ	HAR-Exp-TCJ
α	0.122	0.266	0.221	0.237	0.203
	(1.341)	(3.040)	(2.083)	(2.590)	(1.899)
β_D	0.242	0.283	0.268	0.528	0.579
	(3.329)	(3.125)	(2.528)	(4.286)	(3.634)
β_W	0.104	0.079	-0.097	0.759	1.229
	(1.127)	(0.675)	(-0.431)	(1.971)	(1.999)
β_M	0.533	0.604	1.023	-0.302	-0.614
	(2.054)	(1.822)	(1.983)	(-1.153)	(-1.352)
β_{JD}		0.002	0.184	-0.038	0.167
		(0.010)	(1.282)	(-0.232)	(1.226)
R^2	0.389	0.394	0.368	0.404	0.390
MSE	1.597	1.644	1.739	1.623	1.671
QLIKE	1.459	1.673	1.686	1.571	1.591
HRMSE	1.288	1.502	1.468	1.474	1.459

Table 3: RV, (*t-values*)

- Standard errors correction: Bartlett/Newey-West
- Positive (insignificant) β_{JD} parameter, contradictory to previous researches
- Positive (significant) and persistent impact of C on RV

In-Sample Fitting & Out-of-Sample Forecast

	HAR	HAR-CJ	HAR-TCJ	HAR-Exp-CJ	HAR-Exp-TCJ
α	0.063	0.187	0.170	0.167	0.148
	(2.036)	(4.714)	(4.007)	(3.722)	(3.221)
β_D	0.446	0.434	0.413	0.631	0.660
	(8.262)	(7.784)	(5.980)	(7.846)	(5.935)
β_W	0.201	0.243	0.140	0.278	0.474
	(3.661)	(3.240)	(1.188)	(1.828)	(2.177)
β_M	0.199	0.174	0.306	-0.057	-0.274
	(2.517)	(1.917)	(2.280)	(-0.349)	(-1.062)
β_{JD}		0.119	0.250	0.058	0.223
		(1.020)	(2.968)	(0.532)	(2.680)
R^2	0.436	0.438	0.455	0.420	0.441
MSE	1.523	1.463	1.426	1.508	1.456
HRMSE	0.773	0.841	0.846	0.804	0.835
QLIKE	0.717	0.851	0.838	0.834	0.844

Table 4: Squared-root model, $RV^{1/2}$, (*t-values*)

- Non-linearity modeling, consistent and more significant results
- Error are based on squared-form, i.e RV
- TJ significant impact on day-ahead volatility, i.e No mean reversion effect

In-Sample Fitting & Out-of-Sample Forecast

	HAR	HAR-CJ	HAR-TCJ	HAR-Exp-CJ	HAR-Exp-TCJ
α	-0.316	-0.122	-0.132	-0.477	-0.114
	(-6.096)	(-1.320)	(-0.778)	(-2.904)	(-0.722)
β_D	0.517	0.294	0.178	0.648	0.390
	(10.519)	(5.723)	(3.756)	(20.826)	(4.607)
β_W	0.312	0.384	0.381	0.448	0.046
	(4.186)	(5.046)	(4.567)	(2.676)	(0.396)
β_M	0.113	0.051	0.044	-0.169	0.188
	(1.992)	(0.744)	(0.597)	(-1.962)	(1.178)
β_{JD}		0.586	0.941	0.225	0.647
		(1.924)	(4.661)	(0.862)	(3.686)
R^2	0.432	0.404	0.435	0.410	0.436
MSE	1.662	1.783	1.694	1.540	1.701
HRMSE	0.605	0.698	0.655	0.689	0.611
QLIKE	0.496	0.653	0.576	0.699	0.551

Table 5: Logarithmic model, $\log(RV)$, (t-*values*)

- Positive elasticity w.r.t 1-day-lagged Jump
- Threshold separated jumps more informative
- Best performance (HRMSE, QLIKE) of log-log model

Realized Utility Framework

- Investor: Mean-variance preference with constant sharp ratio on time-varying volatility asset(Bollerslev et al (2018))
- Expected utility function approximation with assuming $W_{t+1} \sim \mathrm{N}(\mu_t, \sigma_t^2)$, $\gamma^A = -\frac{u''}{u'}$ as Pratt-Arrow absolute risk aversion function

$$\mathrm{E}[u(W_{t+1})] = \mu_t - \frac{1}{2}\gamma^A \sigma_t^2 \tag{22}$$

- Investment: ω_t on Cryptocurrencies and $1 - \omega_t$ on risk-free asset at time t

$$W_{t+1} = W_t(1 + r_f + \omega_t r_{t+1}) \tag{23}$$

Where r_{t+1}, r_f as excess return at time $t+1$ and risk free return

Volatility Timing Strategy

- Under assumption of the known constant Sharp ratio $SR = \frac{\mathrm{E}(r_{t+1})}{\sqrt{\mathrm{E}(RV_{t+1})}}$, rewriting expected utility $\mathrm{EU}(\omega_t)$ by replacing $\mathrm{V}(r_{t+1})$ with RV_{t+1}

$$\begin{aligned} \mathrm{EU}(\omega_t) &= W_t \left[\omega_t \,\mathrm{E}(r_{t+1}) + \frac{\gamma}{2}\omega_t^2 \,\mathrm{V}(r_{t+1})\right] && (24)\\ &= W_t \left[\omega_t \,\mathrm{E}(r_{t+1}) + \frac{\gamma}{2}\omega_t^2 RV_{t+1}\right] && (25)\\ &= W_t \left[\omega_t SR \cdot \sqrt{RV_{t+1}} + \frac{\gamma}{2}\omega_t^2 RV_{t+1}\right] && (26) \end{aligned}$$

Where $\gamma = \gamma^A W_t$ as relative risk aversion

- Optimal weight ω_t^* targeting SR/γ: Volatility timing strategy

$$\omega_t^* = \frac{SR/\gamma}{\sqrt{RV_{t+1}}} \tag{27}$$

Evaluating *RV* Forecasting

- Optimal expected utility function

$$\mathrm{EU}(\omega_t^*) = \frac{SR^2}{2\gamma} W_t \tag{28}$$

- For estimated $\widehat{RV}_{t+1}$ and corresponding optimal $\hat{\omega}_t$, the expected utility per wealth

$$\frac{\mathrm{EU}(\hat{\omega}_t)}{W_t} = \frac{SR^2}{\gamma}\left(\sqrt{\frac{RV_{t+1}}{\widehat{RV}_{t+1}}} - \frac{1}{2}\frac{RV_{t+1}}{\widehat{RV}_{t+1}}\right) \tag{29}$$

- Realized utility (RU): Averaging the realized expression by out-of-sample forecast $\widehat{RV}$

$$RU\left(\widehat{RV}_{t+1}\right) = \frac{SR^2}{\gamma}\frac{1}{T}\sum_{t=1}^{T}\left(\sqrt{\frac{RV_{t+1}}{\widehat{RV}_{t+1}}} - \frac{1}{2}\frac{RV_{t+1}}{\widehat{RV}_{t+1}}\right) \tag{30}$$

Empirical Realized Utility Results

		HAR	HAR-CJ	HAR-TCJ	HAR-Exp-CJ	HAR-Exp-TCJ
h=1	RV	3.547%	3.535%	3.534%	3.536%	3.535%
	$(RV)^{1/2}$	3.418%	3.493%	3.499%	3.485%	3.493%
	$\log(RV)$	3.137%	3.343%	3.303%	3.364%	3.265%
h =7	RV	3.758%	3.728%	3.747%	3.704%	3.715%
	$(RV)^{1/2}$	3.718%	3.776%	3.783%	3.776%	3.780%
	$\log(RV)$	3.568%	3.733%	3.640%	3.749%	3.664%
h=30	RV	3.835%	3.751%	3.792%	3.707%	3.733%
	$(RV)^{1/2}$	3.842%	3.787%	3.825%	3.765%	3.775%
	$\log(RV)$	3.780%	3.769%	3.723%	3.675%	3.732%

Table 6: BTC out-of-sample realized utility evaluated at the maximum value equals to $RU(RV_{t+h}) = \frac{1}{2}SR^2/\gamma$=4%. D-M t-test shows Better/Worse comparing with HAR model at 5% Significant level

- Jump components provide significant economic value
- Models perform better on longer forecast horizon
- Non-linear models are better from economic perspective

Summary

- Realized Volatility Processes
  Differences: Significant larger scale and frequent jumps
  Similarities: Log-noraml distributed and long-memory
  Maybe useful for cryptocurrencies options pricing
- Statistics Findings
  Threshold jump method overcomes consecutive jumps and provides more information
  Significant positive impact from TJ on RV. No mean-reversion
  Non-linear models perform better
- Economic Perspective
  Investors gain higher economic value by modeling jumps
  Longer investment horizon, higher utility
  Non-linear modeling is necessary to capture more market changes

Realized Cryptocurrency Volatility Forecasting

Junjie Hu*
WeiYu Kuo°
Wolfgang Karl Härdle*

*Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
°Department of International Business
National Chengchi University
RCVJ_Forecasting

RV and BPV for Market Indices: SPX

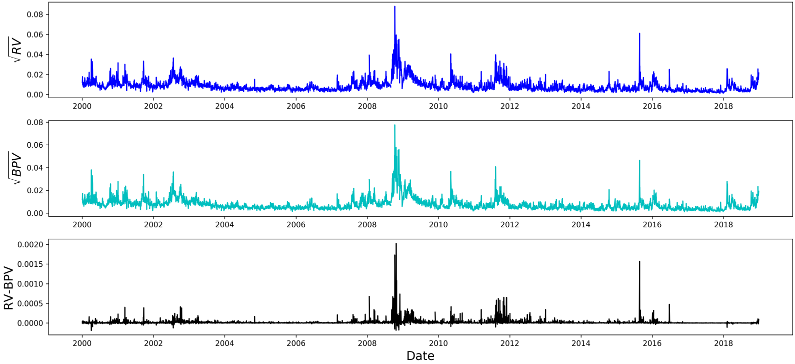

Figure 14: The Standard & Poor's 500 Index, 2000-2018

In-Sample Fitting & Out-of-Sample Forecast

	HAR	HAR-CJ	HAR-TCJs	HAR-Exp-CJs	HAR-Exp-TCJs
α	-0.316	-0.122	-0.132	-0.477	-0.114
	(-6.096)	(-1.320)	(-0.778)	(-2.904)	(-0.722)
β_D	0.517	0.294	0.178	0.648	0.390
	(10.519)	(5.723)	(3.756)	(20.826)	(4.607)
β_W	0.312	0.384	0.381	0.448	0.046
	(4.186)	(5.046)	(4.567)	(2.676)	(0.396)
β_M	0.113	0.051	0.044	-0.169	0.188
	(1.992)	(0.744)	(0.597)	(-1.962)	(1.178)
β_{JD}		0.586	0.941	0.225	0.647
		(1.924)	(4.661)	(0.862)	(3.686)
β_{JW}		-0.844	-0.406	-0.196	-0.200
		(-2.419)	(-1.343)	(-0.564)	(-0.749)
β_{JM}		-0.363	-0.202	0.009	-0.166
		(-0.753)	(-0.497)	(0.024)	(-0.473)
R^2	0.432	0.404	0.435	0.410	0.436
MSE	1.662	1.783	1.694	1.540	1.701
HRMSE	0.605	0.698	0.655	0.689	0.611
QLIKE	0.496	0.653	0.576	0.699	0.551

Table 7: Logarithmic model, $\log(RV)$, t-values in the parantheses

In-Sample Fitting & Out-of-Sample Forecast

	HAR	HAR-CJ	HAR-TCJs	HAR-Exp-CJs	HAR-Exp-TCJs
α	0.063	0.187	0.170	0.167	0.148
	(2.036)	(4.714)	(4.007)	(3.722)	(3.221)
β_D	0.446	0.434	0.413	0.631	0.660
	(8.262)	(7.784)	(5.980)	(7.846)	(5.935)
β_W	0.201	0.243	0.140	0.278	0.474
	(3.661)	(3.240)	(1.188)	(1.828)	(2.177)
β_M	0.199	0.174	0.306	-0.057	-0.274
	(2.517)	(1.917)	(2.280)	(-0.349)	(-1.062)
β_{JD}		0.119	0.250	0.058	0.223
		(1.020)	(2.968)	(0.532)	(2.680)
β_{JW}		-0.140	0.074	-0.086	0.081
		(-1.031)	(0.742)	(-0.633)	(0.824)
β_{JM}		-0.107	-0.157	-0.083	-0.113
		(-0.868)	(-1.743)	(-0.747)	(-1.488)
R^2	0.436	0.438	0.455	0.420	0.441
MSE	1.523	1.463	1.426	1.508	1.456
HRMSE	0.773	0.841	0.846	0.804	0.835
QLIKE	0.717	0.851	0.838	0.834	0.844

Table 8: Squared-root model, $RV^{1/2}$, t-values in the parantheses

Out-of-Sample Forecast, *RV* form

	HAR	HAR-CJ	HAR-TCJs	HAR-Exp-CJs	HAR-Exp-TCJs
α	0.122	0.266	0.221	0.237	0.203
	(1.341)	(3.040)	(2.083)	(2.590)	(1.899)
β_D	0.242	0.283	0.268	0.528	0.579
	(3.329)	(3.125)	(2.528)	(4.286)	(3.634)
β_W	0.104	0.079	-0.097	0.759	1.229
	(1.127)	(0.675)	(-0.43)1	(1.971)	(1.999)
β_M	0.533	0.604	1.023	-0.302	-0.614
	(2.054)	(1.822)	(1.983)	(-1.153)	(-1.352)
β_{JD}		0.002	0.184	-0.038	0.167
		(0.010)	(1.282)	(-0.232)	(1.226)
β_{JW}		-0.004	0.326	-0.052	0.212
		(-0.026)	(1.502)	(-0.276)	(1.252)
β_{JM}		-0.704	-0.372	-0.471	-0.183
		(-1.749)	(-1.788)	(-1.588)	(-1.119)
R^2	0.389	0.394	0.368	0.404	0.390
MSE	1.597	1.644	1.739	1.623	1.671
QLIKE	1.459	1.673	1.686	1.571	1.591
HRMSE	1.288	1.502	1.468	1.474	1.459

Table 9: *RV*, t-values in the parantheses

Log-return Distribution, Full Sample

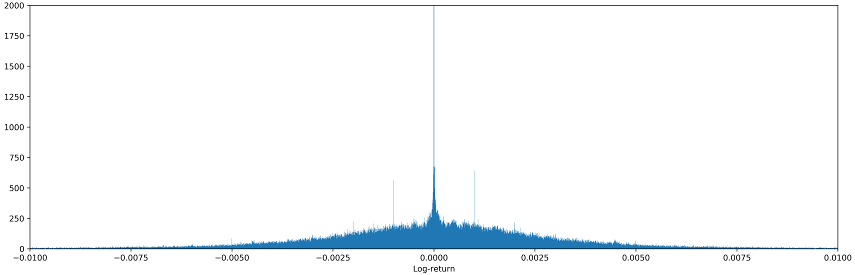

Figure 15: Histogram of BTC 5-min log-return, 1st, Jan 2016 to 1st, July 2018

- Log-returns range in $[-0.137, 0.137]$
- Zero dropped
- Suspicious values dropped

Log-return Distribution, Since 2017

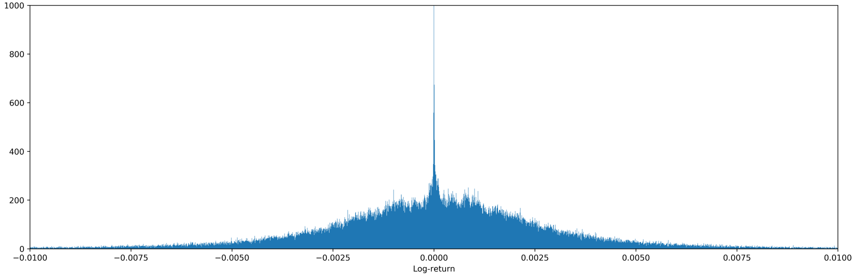

Figure 16: Histogram of BTC 5-min log-return, 1st, Jan 2017 to 1st, July 2018

- Log-returns range in $[-0.084, 0.087]$
- Zero values dropped
- Suspicious values dropped

Log-return Distribution, Full Sample

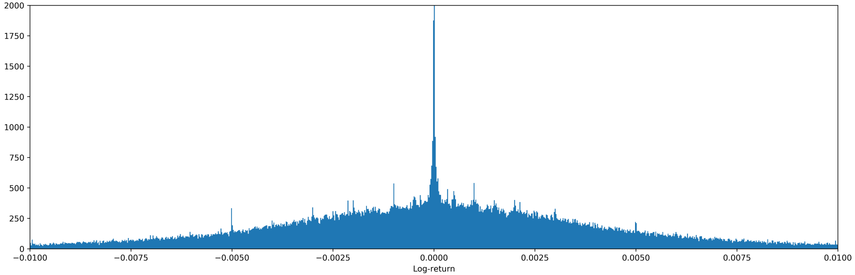

Figure 17: Histogram of ETH 5-min log-return, 1st, Jan 2016 to 1st, July 2018

- Log-returns range in $[-0.425, 0.448]$
- Zero dropped
- Suspicious values dropped

Log-return Distribution, Since 2017

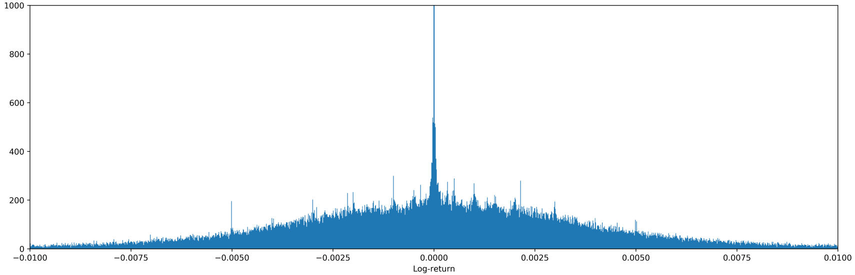

Figure 18: Hisogram of ETH 5-min log-return, 1st, Jan 2017 to 1st, July 2018

- Log-returns range in $[-0.145, 0.129]$
- Zero dropped
- Suspicious values dropped

Avoid using future information

- Rolling window within each day