Efficient Sampling for Realized Variance Estimation in Time-Changed Diffusion Models

Timo Dimitriadis[1], Roxana Halbleib[2], **Jasper Rennspies**[2], Jeannine Polivka[3], Sina Streicher[4], Axel Friedrich Wolter[2]

[1]University of Heidelberg
[2]University of Freiburg
[3]University of St.Gallen, Switzerland
[4]KOF Swiss Economic Institute, ETH Zürich

EEA-ESEM, Rotterdam
August 27, 2024

Idea

- **Aim**: provide more efficient daily *variance* estimates and forecasts based on high-frequency (HF) returns by sampling in *intrinsic time*

Idea

- **Aim**: provide more efficient daily *variance* estimates and forecasts based on high-frequency (HF) returns by sampling in *intrinsic time*
- Usefulness of the *variance* of financial returns: risk management, portfolio optimization and asset pricing.

Idea

- **Aim**: provide more efficient daily *variance* estimates and forecasts based on high-frequency (HF) returns by sampling in *intrinsic time*
- Usefulness of the *variance* of financial returns: risk management, portfolio optimization and asset pricing.
- *Intrinsic time* captures the real *heartbeat* of financial markets.

Idea

- **Aim**: provide more efficient daily *variance* estimates and forecasts based on high-frequency (HF) returns by sampling in *intrinsic time*
- Usefulness of the *variance* of financial returns: risk management, portfolio optimization and asset pricing.
- *Intrinsic time* captures the real *heartbeat* of financial markets.
- Why intrinsic (vs. calendar) time?
  - ◇ Get more/different insight and better (risk) measures by taking in account the market's activity when sampling data
  - ◇ Exploit better the rich information content of HF data (150GB/day from NYSE $\sim$ 100 TB HF data)

Idea

- **Aim**: provide more efficient daily *variance* estimates and forecasts based on high-frequency (HF) returns by sampling in *intrinsic time*
- Usefulness of the *variance* of financial returns: risk management, portfolio optimization and asset pricing.
- *Intrinsic time* captures the real *heartbeat* of financial markets.
- Why intrinsic (vs. calendar) time?
  - ◇ Get more/different insight and better (risk) measures by taking in account the market's activity when sampling data
  - ◇ Exploit better the rich information content of HF data (150GB/day from NYSE $\sim$ 100 TB HF data)
- **Contribution**: First paper to show theoretically and empirically the advantages of intrinsic time for the daily variance estimation from HF data

Our focus in this paper

Realized Variance - RV (Andersen and Bollerslev, 1998; Andersen et al., 2001, Barndorff-Nielsen and Shephard, 2002)

- Extensively studied in financial econometrics literature
- Unbiased and consistent estimator of daily variance (integrated variance) by using HF returns
- Underlying prices follow a **diffusion process** and exhibit no Market Microstructure Noise (MMN)
- Find optimal frequency when equidistantly sampling in clock time: 5-minute
- Under MMN, e.g., pre-averaged RV of Podolskij and Vetter (2009)

Our focus in this paper

Realized Variance - RV (Andersen and Bollerslev, 1998; Andersen et al., 2001, Barndorff-Nielsen and Shephard, 2002)

- Extensively studied in financial econometrics literature
- Unbiased and consistent estimator of daily variance (integrated variance) by using HF returns
- Underlying prices follow a **diffusion process** and exhibit no Market Microstructure Noise (MMN)
- Find optimal frequency when equidistantly sampling in clock time: 5-minute
- Under MMN, e.g., pre-averaged RV of Podolskij and Vetter (2009)

Problems:

- **Diffusion process** might not reflect reality
- Observed prices are discrete; only change after new trades materialize
- Trading intensity varies heavily over the day
- The ticks (prices) evolve over the day with time-changing variance and the arrival times are stochastic
- The diffusion model can neither model the arrivals of the transactions nor can it accommodate the discontinuity of the many little price jumps
- Discretized diffusion models (Jacod et al. 2017, ...): stochastic arrival times

Contribution of our paper

Tick-time stochastic volatility (TTSV) model:

- The price process is doubly stochastic: pure jump tick process with time-varying **tick volatility** and time-varying **trading intensity** that exhibit different (opposite) intraday patterns
- Is a time-changed/subordinated diffusion process for log-prices: joint stochastic model for asset prices and their arrival times
- It allows to study theoretically (pre-averaged) RV based on calendar time - CTS, **business time** - BTS and **transaction time** - TTS sampling
- Distinguish between (estimated) intensity and (observed) realized/jump-based sampling
- Theoretical analysis in finite sample and asymptotically
- Comprehensive simulation and empirical evidence on real data

Main findings of our paper

- Spot variance = Time-varying tick variance × trading intensity
- BTS is favourable against CTS and TTS (see literature)
- Find a new optimal sampling scheme: realised BTS (rBTS)
- rBTS samples according to observed ticks adjusted by their variation
- rBTS provides a theoretical lower bound for the efficiency of the (pre-averaged) RV estimator in finite samples and asymptotically.
- In simulation as well as in-sample and out-of-sample outperformance of rBTS

General Sampling and Realized Variance

- We define the general sampling (GS) scheme (not necessarily equidistant in calendar time) by

$$\tau := (\tau_0, \ldots, \tau_M)$$

  with $0 = \tau_0 < \tau_1 < \ldots < \tau_M = T$
- $M \in \mathbb{N}$ is the number of subintervals $(\tau_{j-1}, \tau_j]$ for $j = 1, \ldots, M$.

General Sampling and Realized Variance

- We define the general sampling (GS) scheme (not necessarily equidistant in calendar time) by

$$\tau := (\tau_0, \ldots, \tau_M)$$

with $0 = \tau_0 < \tau_1 < \ldots < \tau_M = T$

- $M \in \mathbb{N}$ is the number of subintervals $(\tau_{j-1}, \tau_j]$ for $j = 1, \ldots, M$.
- The intra-daily log-returns over a trading day are given by

$$r_j := r(\tau_{j-1}, \tau_j) = P(\tau_j) - P(\tau_{j-1}), \quad j = 1, \ldots, M. \tag{1}$$

General Sampling and Realized Variance

- We define the general sampling (GS) scheme (not necessarily equidistant in calendar time) by

$$\tau := (\tau_0, \ldots, \tau_M)$$

  with $0 = \tau_0 < \tau_1 < \ldots < \tau_M = T$
- $M \in \mathbb{N}$ is the number of subintervals $(\tau_{j-1}, \tau_j]$ for $j = 1, \ldots, M$.
- The intra-daily log-returns over a trading day are given by

$$r_j := r(\tau_{j-1}, \tau_j) = P(\tau_j) - P(\tau_{j-1}), \quad j = 1, \ldots, M. \tag{1}$$

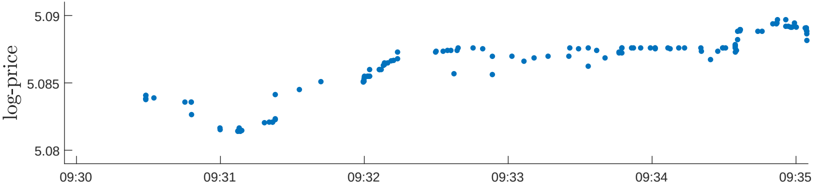
Realized Variance (Andersen and Bollerslev, 1998): For a given interval $[0, T]$, $T \geq 0$, the realized variance (RV) based on $M \in \mathbb{N}$ intraday log-returns is defined by

$$\text{RV}_{\boldsymbol{\tau}_M}(0, T) := \sum_{j=1}^{M} r_j^2, \tag{2}$$

where the index $\boldsymbol{\tau}_M$ highlights its dependence on the chosen grid.

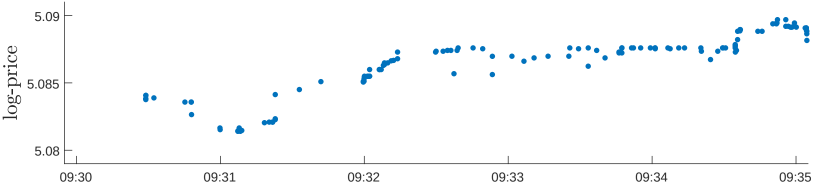
Standard Diffusion Setting

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Standard Diffusion Setting

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



$P(t)$ solves the stochastic differential equation

$$dP(t) = \mu(t)\,dt + \sigma(t)\,dB(t), \quad t \in [0, T], \tag{3}$$

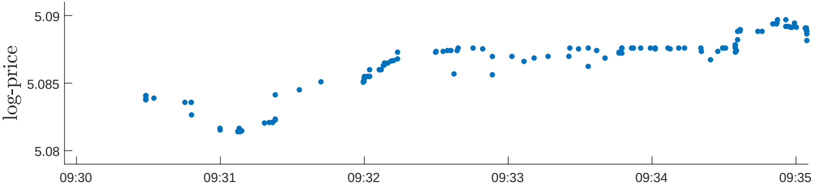
where $\mu(t)$ is of finite variation and $\mathcal{F}_t$-predictable, $\sigma(t)$ is spot variance, $\mathcal{F}_t$-predictable, is independent of $B(t)$ and $\mathbb{E}\big[\int_0^t \sigma^2(r)\,dr\big] < \infty$.

The integrated variance (IV) over $[0, T]$:

$$\mathrm{IV}(0, T) := \int_0^T \sigma^2_{spot}(r)\,dr = \int_0^T \sigma^2(r)\,dr.$$

Standard Diffusion Setting

Figure: Snippet of the IBM transaction log-price on July 2[nd], 2015.



$P(t)$ solves the stochastic differential equation

$$dP(t) = \mu(t)\,dt + \sigma(t)\,dB(t), \quad t \in [0, T], \tag{3}$$

where $\mu(t)$ is of finite variation and $\mathcal{F}_t$-predictable, $\sigma(t)$ is spot variance, $\mathcal{F}_t$-predictable, is independent of $B(t)$ and $\mathbb{E}\big[\int_0^t \sigma^2(r)\,dr\big] < \infty$.

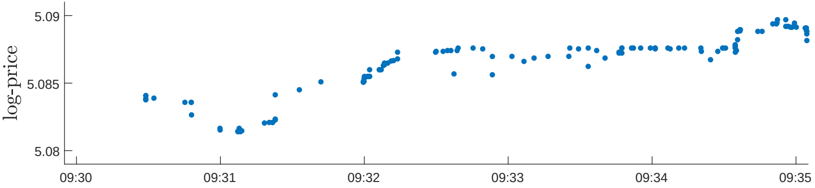
The integrated variance (IV) over $[0, T]$:

$$\mathrm{IV}(0, T) := \int_0^T \sigma^2_{spot}(r)\,dr = \int_0^T \sigma^2(r)\,dr.$$

RV is a consistent and asymptotically normal estimator of IV.

Compound Poisson process

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Oomen (2005, 2006): $P(t)$ follows a compound Poisson process (CPP) given by

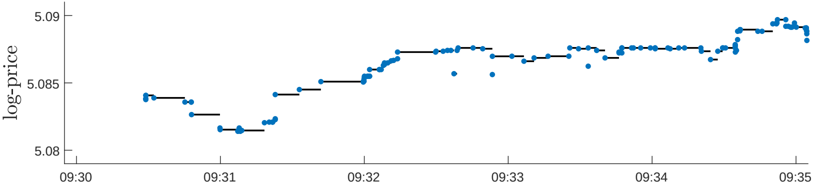
$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varepsilon_i, \tag{4}$$

$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{5}$$

where $N(t)$ is a doubly stochastic Poisson process with stochastic intensity $\lambda(t)$, and $N(t)$ and ε_i are independent.

Compound Poisson process

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Oomen (2005, 2006): $P(t)$ follows a compound Poisson process (CPP) given by

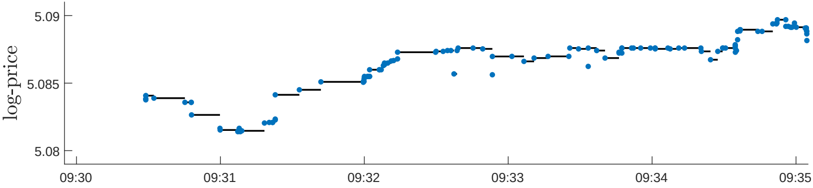
$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varepsilon_i, \qquad (4)$$

$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \qquad (5)$$

where $N(t)$ is a doubly stochastic Poisson process with stochastic intensity $\lambda(t)$, and $N(t)$ and ε_i are independent.

Compound Poisson process

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Oomen (2005, 2006): $P(t)$ follows a compound Poisson process (CPP) given by

$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varepsilon_i, \tag{4}$$

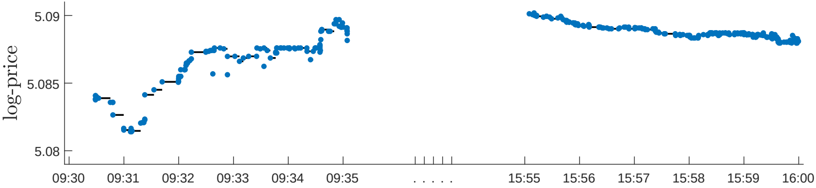
$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{5}$$

where $N(t)$ is a doubly stochastic Poisson process with stochastic intensity $\lambda(t)$, and $N(t)$ and ε_i are independent.

$$\text{IV}(0, T) := \int_0^T \sigma_{spot}^2(r)\, dr = \int_0^T \sigma_\varepsilon^2 \lambda(r)\, dr$$

Compound Poisson process

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Oomen (2005, 2006): $P(t)$ follows a compound Poisson process (CPP) given by

$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varepsilon_i, \tag{4}$$

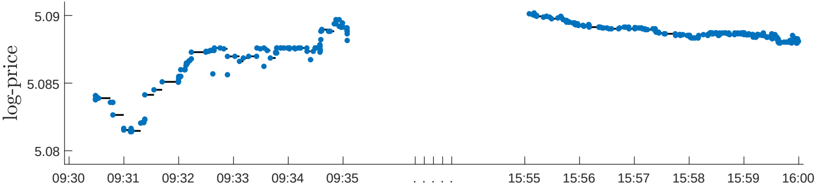
$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{5}$$

where $N(t)$ is a doubly stochastic Poisson process with stochastic intensity $\lambda(t)$, and $N(t)$ and ε_i are independent.

$$\text{IV}(0, T) := \int_0^T \sigma_{spot}^2(r)\, dr = \int_0^T \sigma_\varepsilon^2 \lambda(r)\, dr$$

Compound Poisson process

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Oomen (2005, 2006): $P(t)$ follows a compound Poisson process (CPP) given by

$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varepsilon_i, \qquad (4)$$

$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \qquad (5)$$

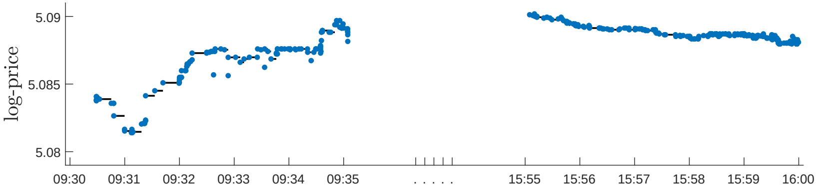
where $N(t)$ is a doubly stochastic Poisson process with stochastic intensity $\lambda(t)$, and $N(t)$ and ε_i are independent.

$$\text{IV}(0, T) := \int_0^T \sigma_{spot}^2(r)\, dr = \int_0^T \sigma_\varepsilon^2 \lambda(r)\, dr$$

Tick-Time Stochastic Volatility (TTSV) Model

:

Figure: Snippet of the IBM transaction log-price on July 2nd, 2015.



Our model: The log-price process $P(t)$ follows

$$P(t) = P(0) + \sum_{i=1}^{N(t)} \varsigma(t_i) U_i \tag{6}$$

- $N(t)$ is a general jump process with stochastic intensity $\lambda(t)$ and stochastic arrival times t_i, $i = 1, \ldots, N$; $N(t) = i$ for $t \in [t_i, t_{i+1})$.
- Conditionally that an arrival t_i has occurred, $U_i = B(N(t_i) - B(N(t_{i-1}))$, i.e. $U_i \overset{iid}{\sim} \mathcal{N}(0, 1)$

TTSV: $N(t)$ and $P(t)$

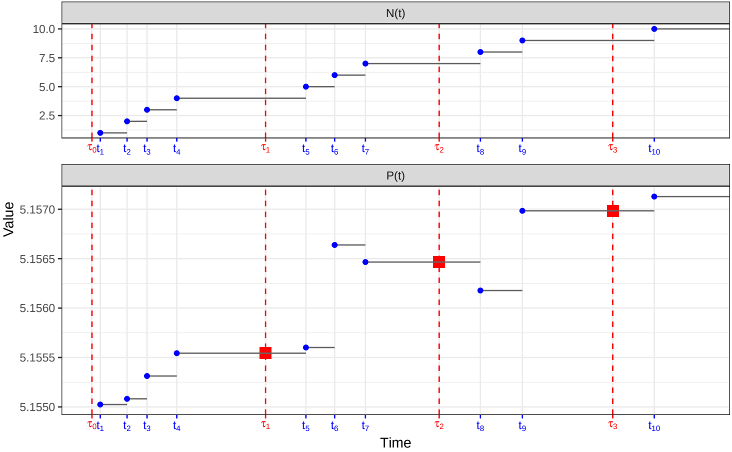


Figure: Upper panel: evolution of the jump process $N(t)$ generating the ticks (arrival times) t_i. Lower panel: the log-price process $P(t)$, which exhibits price jumps at the ticks t_i of $N(t)$ and is constant in between. Vertical red lines represent the sampling times of an exemplary sampling scheme $\boldsymbol{\tau}^T$ (that does not have to be equidistant in calendar time), and the red squares show the resampled prices based on the previous tick method.

Some Assumptions

Assumptions 1:

- The filtration $\mathcal{F}_t := \sigma\big(Z_1(s), \ldots, Z_m(s);\ 0 \le s \le t\big)$ is generated by a fixed number $m \in \mathbb{N}$ of given stochastic factors $Z_1(t), \ldots, Z_m(t)$ that govern the randomness of the price processes.
- The counting process $\{N(t)\}_{t\ge 0}$ is an $\mathcal{F}_t$-adapted jump process, which has an $\mathcal{F}_t$-predictable intensity process $\{\lambda(t)\}_{t\ge 0}$ that is left-continuous with right-hand limits and $\int_0^t \lambda(s)ds < \infty$ a.s. for all $t \ge 0$.
- $\{B(n)\}_{n\ge 0}$ is a Brownian motion such that $B(N(t))$ is adapted to $\mathcal{F}_t$.
- The tick volatility $\{\varsigma(t)\}_{t\ge 0}$ is a positive, continuous and $\mathcal{F}_t$-predictable process.

Assumptions 2:

- The process $\{B(n)\}_{n\ge 0}$ is *independent* from $\{N(t)\}_{t\ge 0}$ and from $\{\varsigma(t)\}_{t\ge 0}$.

Assumptions 3:

- The expectations $\mathbb{E}\big[\int_t^T \varsigma^2(r)\lambda(r)dr \mid \mathcal{F}_t\big]$, $\mathbb{E}\left[\varsigma^4(t)\right]$ and $\mathbb{E}\big[\int_0^t \varsigma^4(r)\,\lambda(r)\,dr\big]$ exist and are finite for all $t \in [0, T]$.

Assumptions 4:

- The counting process $\{N(t)\}_{t\ge 0}$ is a doubly stochastic Poisson process, adapted to $\mathcal{F}_t$, which has a positive, $\mathcal{F}_t$-measurable and continuous intensity $\{\lambda(t)\}_{t\ge 0}$ such that $\int_0^t \lambda(s)ds < \infty$ a.s. for all $t \ge 0$;
- The processes $\{N(t)\}_{t\ge 0}$ and $\{\varsigma(t)\}_{t\ge 0}$ are *independent*.

TTSV Model

The TTSV model can be written as:

$$dP\,(t) = \varsigma\,(t)\,dB\,(N\,(t))\,, \quad t \in [0, T], \tag{7}$$

- We show that
$$\sigma^2_{spot}\,(t) = \varsigma^2\,(t)\,\lambda\,(t+)\,,$$
where $\lambda(t_+) := \lim_{\delta \searrow 0} \lambda(t + \delta)$;
- For continuous intensities $\lambda(t)$, the decomposition simplifies to $\sigma^2(t) = \varsigma^2(t)\lambda(t)$.
- It holds that
$$\mathbb{E}\left[r^2_{\text{daily}} - \text{IV}\,(0, T)\right] = 0.$$
- We get
$$\text{IV}(0, T) = \int_0^T \sigma^2_{spot}\,(r)\,dr = \int_0^T \varsigma^2\,(r)\,\lambda\,(r)\,dr.$$

A new concept of IV and unbiasedness results

- Define the following two information sets for $t \in [0, T]$,

$$\mathcal{F}_t^{\lambda,\varsigma} := \sigma\big(\lambda(s), \varsigma(s); \quad 0 \leq s \leq t\big) \subset \mathcal{F}_t, \qquad \text{and}$$
$$\mathcal{F}_t^{\lambda,\varsigma,N} := \sigma\big(\lambda(s), \varsigma(s), N(s); \quad 0 \leq s \leq t\big) \subset \mathcal{F}_t,$$

- Define the realized IV (rIV) as $\mathrm{rIV}(0, T) := \int_0^T \varsigma^2(r) dN(r) = \sum_{0 \leq t_i \leq T} \varsigma^2(t_i)$ that can be interpreted as a jump-process based and hence "realized" version of the classical IV.

Theorem 1 (Unbiasedness)

(a) Under Assumptions 1-3, it holds that $\mathbb{E}\left[\mathrm{RV}(\boldsymbol{\tau}^T) \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma,N}\right] = \mathrm{rIV}(0, T)$.

(b) Under Assumptions 1-4, it holds that $\mathbb{E}\left[\mathrm{RV}(\boldsymbol{\tau}^T) \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma}\right] = \mathrm{IV}(0, T)$.

Efficiency Results I

Theorem 2 (Efficiency)

(a) *Under Assumptions 1–3 it holds that*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma,N}\right] = \left(\mathrm{rIV}(0,T) - \mathrm{IV}(0,T)\right)^2 + \underline{2\sum_{j=1}^{M(T)} \mathrm{rIV}(\tau_{j-1}^T, \tau_j^T)^2}$$

.

(b) *Under Assumptions 1–4 it holds that*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma}\right] = 3\,\mathrm{IQ}(0,T) + \underline{2\sum_{j=1}^{M(T)} \mathrm{IV}(\tau_{j-1}^T, \tau_j^T)^2},$$

where $\mathrm{IQ}(s,t) := \int_s^t \varsigma^4(r)\lambda(r)dr$ *denotes the Integrated Quarticity of the TTSV model.*

Efficiency Results I

Theorem 2 (Efficiency)

(a) *Under Assumptions 1–3 it holds that*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma,N}\right] = \left(\mathrm{rIV}(0,T) - \mathrm{IV}(0,T)\right)^2 + \underline{2\sum_{j=1}^{M(T)} \mathrm{rIV}(\tau_{j-1}^T,\tau_j^T)^2}$$

.

(b) *Under Assumptions 1–4 it holds that*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma}\right] = 3\,\mathrm{IQ}(0,T) + \underline{2\sum_{j=1}^{M(T)} \mathrm{IV}(\tau_{j-1}^T,\tau_j^T)^2},$$

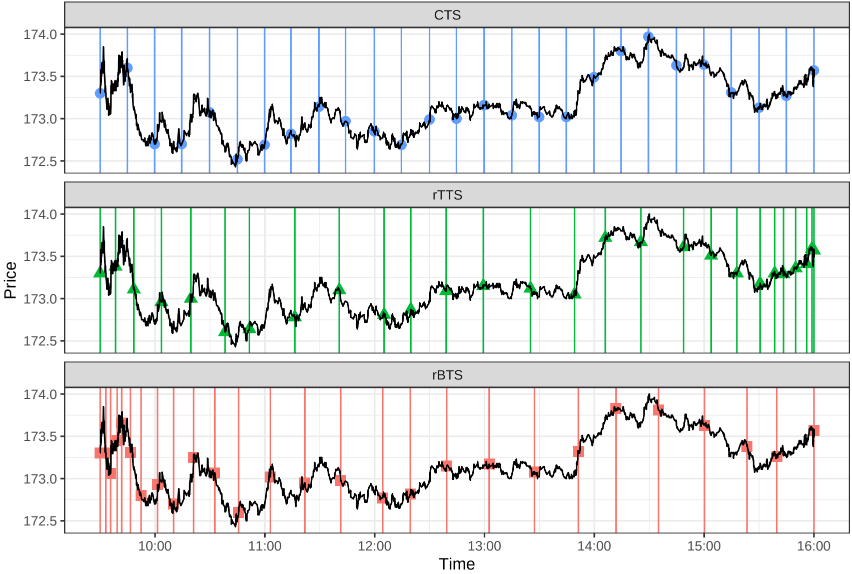
where $\mathrm{IQ}(s,t) := \int_s^t \varsigma^4(r)\lambda(r)dr$ *denotes the Integrated Quarticity of the TTSV model.*

- MSE is bounded below and depends on the choice of sampling scheme $\boldsymbol{\tau}^T$

Sampling Schemes $\boldsymbol{\tau}^T$

- **Calendar Time Sampling (CTS)**: the sampling times are equidistant in calendar time.
- **Intensity Transaction Time Sampling (iTTS)**: data sampled equidistantly in the integrated (latent) *trading intensity*. It is $\mathcal{F}_T^{\lambda,\varsigma}$-measurable.
- **Realized Transaction Time Sampling (rTTS)**: data sampled equidistantly in the *observed number of transactions*. It is $\mathcal{F}_T^{\lambda,\varsigma,N}$-measurable.
- **Intensity Business Time Sampling (iBTS)**: data sampled equidistantly in integrated (latent) *spot variance*. It is $\mathcal{F}_T^{\lambda,\varsigma}$-measurable and incorporates the tick variance and the trading intensity, both latent.
- **Realized Business Time Sampling (rBTS)**: data sampled equidistantly in *the observed number of transactions weighted by their tick variance*. It is a $\mathcal{F}_T^{\lambda,\varsigma,N}$-measurable extension of iBTS when accounting for the sample path of $N(t)$.

Sampling Schemes $\boldsymbol{\tau}^T$



Efficiency Results II

Corollary 3

(a) *Under Assumptions 1–3 and given that the sampling times* $\boldsymbol{\tau}^T$ *are* $\mathcal{F}_T^{\lambda,\varsigma,N}$*-measurable,*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma,N}\right] \geq \mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^{rBTS}) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma,N}\right],$$

with equality if and only if $\boldsymbol{\tau}^T \equiv \boldsymbol{\tau}^{rBTS}$*.*

(b) *Under Assumptions 1–4 and given that the sampling times* $\boldsymbol{\tau}^T$ *are* $\mathcal{F}_T^{\lambda,\varsigma}$*-measurable,*

$$\mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma}\right] \geq \mathbb{E}\left[\left(\mathrm{RV}(\boldsymbol{\tau}^{iBTS}) - \mathrm{IV}(0,T)\right)^2 \,\middle|\, \mathcal{F}_T^{\lambda,\varsigma}\right],$$

with equality if and only if $\boldsymbol{\tau}^T \equiv \boldsymbol{\tau}^{iBTS}$*.*

Asymptotic results for jump-based sampling without MMN

- In-fill asymptotic as in diffusion models is not possible within TTSV, as the amount of observed ticks on the fixed interval $[0, T]$ is bounded in probability.
- Our asymptotic is based on increasing the number of ticks by letting T diverge to ∞ (Dahlhaus, 1997; Dahlhaus and Tunyavetchakit, 2016).
- To allow for jump-based sampling, we make assumptions related to Fukasawa (2010) and Mykland and Zhang (2006).

Asymptotic results for jump-based sampling without MMN

Theorem 4

Under the following assumptions

- $\mathbb{E}[\tau_j^T - \tau_{j-1}^T \mid \mathcal{F}^T_{\tau^T_{j-1}}] = \left(f\phi\left(\frac{\tau^T_{j-1}}{T}\right)\right)^{-1} + o_p(1)$ *uniformly in* $j \in \{1,\dots,M(T)\}$ *for some integrable sampling intensity function* $\phi : [0,1] \to \mathbb{R}_{>0}$ *such that* $\int_0^1 \phi(s)ds = 1$,
- $\mathrm{Var}\left[N_T(\tau_j^T) - N_T(\tau_{j-1}^T) \mid \mathcal{F}^T_{\tau^T_{j-1}}\right] = \mu\left(\frac{\tau^T_{j-1}}{T}\right)\mathbb{E}[\tau_j^T - \tau_{j-1}^T \mid \mathcal{F}^T_{\tau^T_{j-1}}] + o_p(1)$ *uniformly in* $j \in \{1,\dots,M(T)\}$ *for some integrable function* $\mu : [0,1] \to \mathbb{R}_{>0}$

and some further technical assumptions, it holds that

$$\sqrt{T}\left(\mathrm{RV}(\boldsymbol{\tau}^T) - \mathrm{IV}\right) \xrightarrow{d} \mathcal{N}\left(0, V_\phi + V_\mu + \mathrm{IQ}\right),$$

where

$$V_\phi = \frac{2}{f}\int_0^1 \frac{\varsigma^4(r)\lambda^2(r)}{\phi(r)}dr, V_\mu = 2\int_0^1 \varsigma^4(r)\mu(r)dr,$$

$$\mathrm{IQ} = \int_0^1 \varsigma^4(r)\lambda(r)dr, f = \mathrm{plim}_{T\to\infty}\frac{M(T)}{T} > 0.$$

- CTS, iTTS, iBTS, rTTS satisfy the assumptions
- Conjecture: rBTS also satisfies the assumptions and it minimizes the asymptotic variance.

Market Microstructure Noise (MMN)

We observe $\tilde{P}(t)$ which follows the decomposition

$$\widetilde{P}_T(\tau_j^T) = P_T(\tau_j^T) + v_j^T, \qquad j = 1, \ldots, M(T), \tag{8}$$

and the corresponding noisy log-returns are given by

$$\tilde{r}_{j,T} = \widetilde{P}_T(\tau_j^T) - \widetilde{P}_T(\tau_{j-1}^T) = r_{j,T} + v_j^T - v_{j-1}^T.$$

where

- $P_T(\cdot)$ is the true, efficient price that follows the TTSV model
- v_j^T is the noise component s.t.
  (a) is i.i.d. with $\mathbb{E}[v_j^T] = 0$, $\mathrm{Var}(v_j^T) = \omega^2$ and $\mathbb{E}\left[(v_j^T)^4\right] = \theta\omega^4$ for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$,
  (b) The sequences v_j^T and $P_T(\tau_j^T)$ are independent for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$,
  (c) v_j^T is $\mathcal{F}^T_{\tau_j^T}$-measurable for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$.

Market Microstructure Noise (MMN)

We observe $\tilde{P}\,(t)$ which follows the decomposition

$$\widetilde{P}_T(\tau_j^T) = P_T(\tau_j^T) + v_j^T, \qquad j = 1, \dots, M(T), \tag{8}$$

and the corresponding noisy log-returns are given by

$$\tilde{r}_{j,T} = \widetilde{P}_T(\tau_j^T) - \widetilde{P}_T(\tau_{j-1}^T) = r_{j,T} + v_j^T - v_{j-1}^T.$$

where

- $P_T(\cdot)$ is the true, efficient price that follows the TTSV model
- v_j^T is the noise component s.t.
  - (a) is i.i.d. with $\mathbb{E}[v_j^T] = 0$, $\mathrm{Var}(v_j^T) = \omega^2$ and $\mathbb{E}\left[(v_j^T)^4\right] = \theta\omega^4$ for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$,
  - (b) The sequences v_j^T and $P_T(\tau_j^T)$ are independent for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$,
  - (c) v_j^T is $\mathcal{F}^T_{\tau_j^T}$-measurable for all $j \in \{0, .., M(T)\}$ and $T \in \mathbb{N}$.

- To deal with MMN, we implement the pre-averaged RV estimator of Podolskij and Vetter (2009)

Asymptotic results for intensity-based sampling with MMN

Theorem 5

Given that the price process is contaminated by MMN, that the sampling times $\boldsymbol{\tau}^T$ are intensity-based, and that the pre-averaging bandwidth $H = \delta\sqrt{T}$, it holds that

$$T^{1/4}\left(\overline{\mathrm{RV}}(\boldsymbol{\tau}^T) - \mathrm{IV}\right) \xrightarrow{d} \mathcal{N}\left(0, \delta\eta_A^2 + \frac{1}{\delta}\eta_B^2 + \frac{1}{\delta^3}\eta_C^2\right),$$

with

$$\eta_A^2 = \frac{2}{f}\int_0^1 \frac{\varsigma^4(r)\lambda^2(r)}{\phi(r)}dr, \quad \eta_B^2 = 4\frac{g_2'}{g_2}\omega^2\int_0^1 \varsigma^2(r)\lambda(r)dr, \quad \textit{and} \quad \eta_C^2 = 2f\left(\frac{g_2'}{g_2}\right)^2\omega^4. \tag{9}$$

- The iBTS scheme minimizes the asymptotic variance

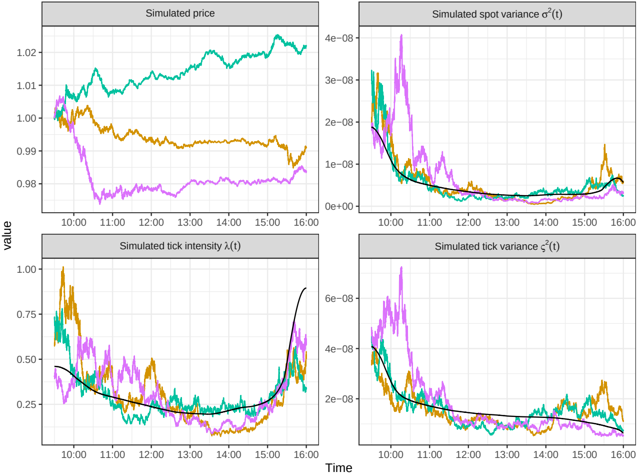
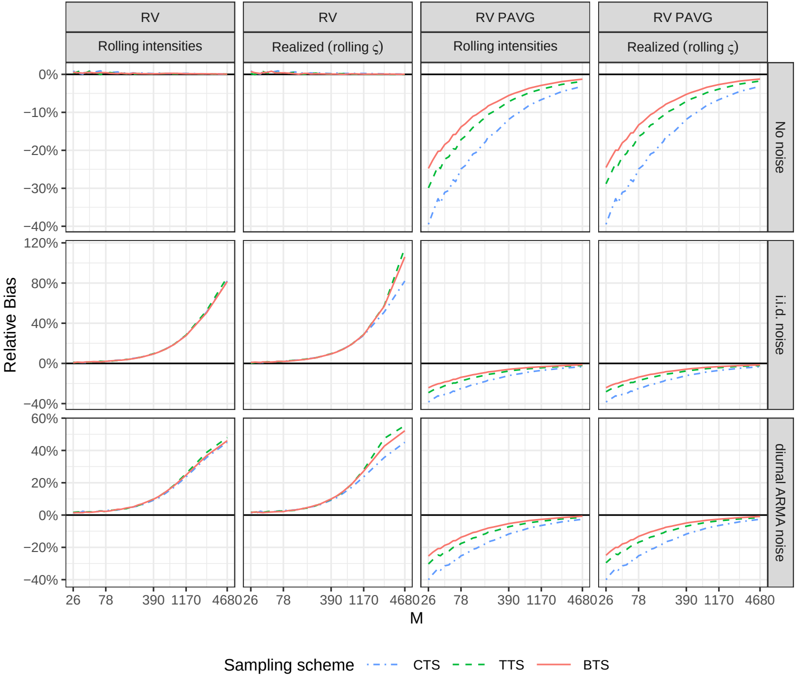
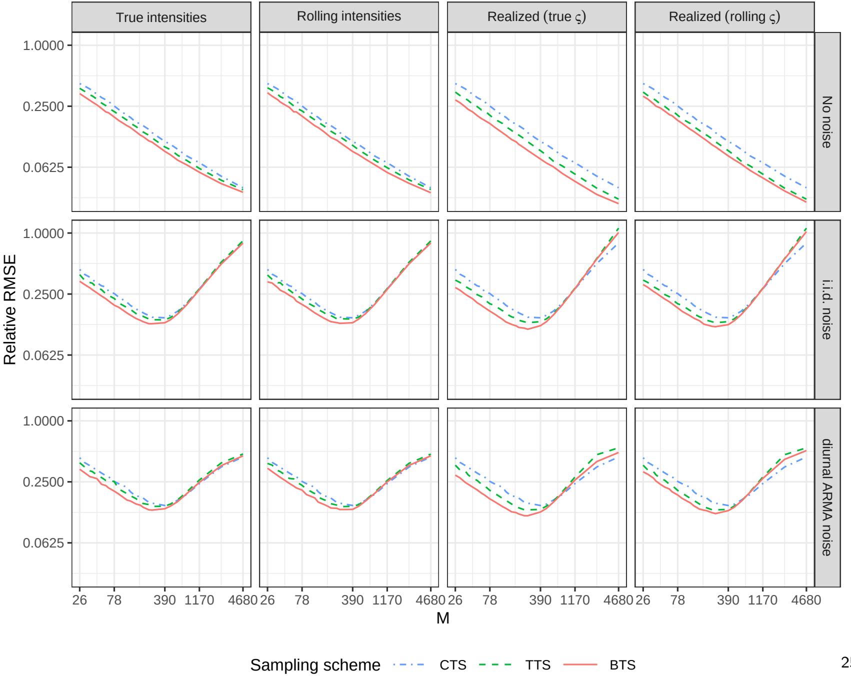


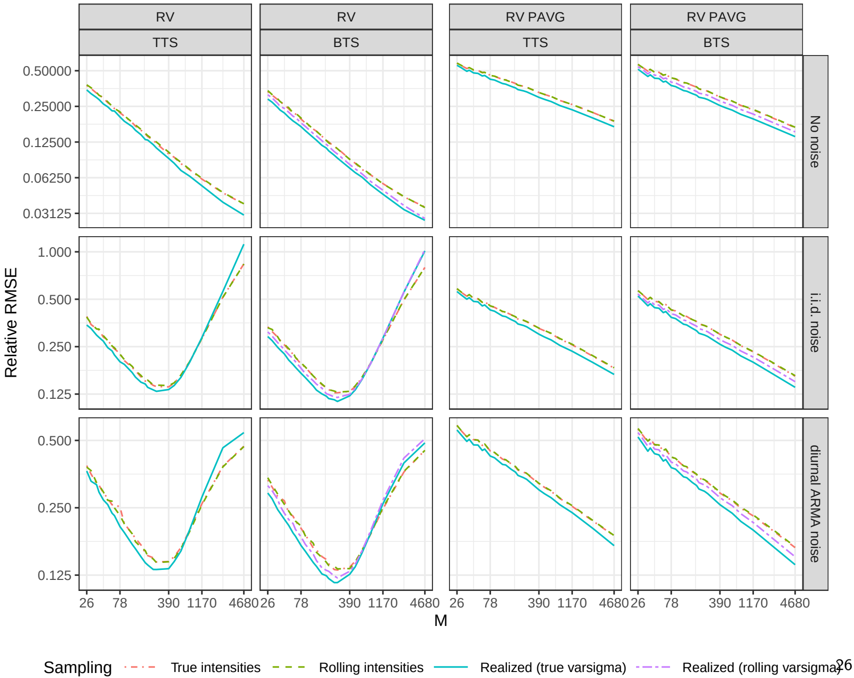
Figure: Simulated paths of the asset price according to TTSV model, the spot variance $\sigma^2(t)$, the trading intensity $\lambda(t)$, and the tick variance $\varsigma^2(t)$ for three exemplary days in colors. Simulate $D = 5000$ days from the TTSV model using $T = 23400$.

$$\lambda(t) = \lambda_{\text{det}}(t) \exp\left(0.01\lambda^*(t) - \bar{\lambda}^*\right), \quad \text{where} \quad d\lambda^*(t) = -0.0002\lambda^*(t)dt + dB_1(t),$$

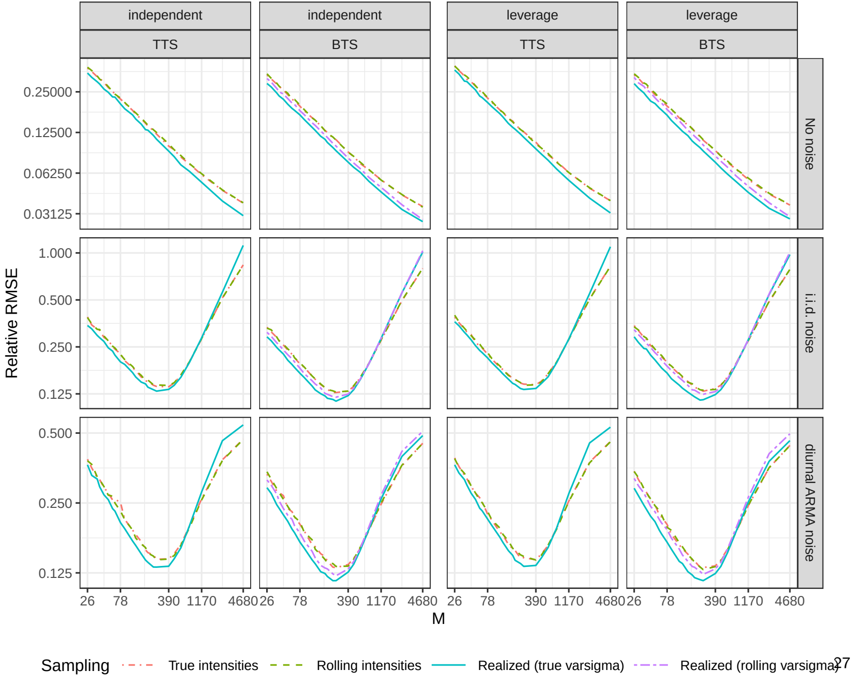
$$\varsigma(t) = \varsigma_{\text{det}}(t) \exp\left(0.005\varsigma^*(t) - \bar{\varsigma}^*\right), \quad \text{where} \quad d\varsigma^*(t) = -0.0002\varsigma^*(t)dt + dB_2(t),$$


RV
Rolling intensities
RV
Realized (rolling ς)
RV PAVG
Rolling intensities
RV PAVG
Realized (rolling ς)
No noise
i.i.d. noise
diurnal ARMA noise
Relative Bias
0%
−10%
−20%
−30%
−40%
120%
80%
40%
0%
−40%
60%
40%
20%
0%
−20%
−40%
26
78
390
1170
4680
M
Sampling scheme
CTS
TTS
BTS


True intensities
Rolling intensities
Realized (true ς)
Realized (rolling ς)
No noise
i.i.d. noise
diurnal ARMA noise
Relative RMSE
1.0000
0.2500
0.0625
26
78
390
1170
4680
M
Sampling scheme
CTS
TTS
BTS

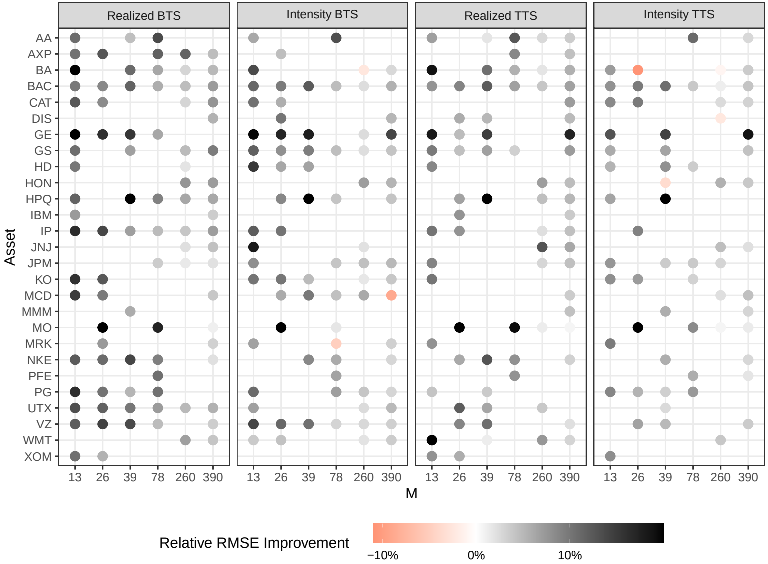


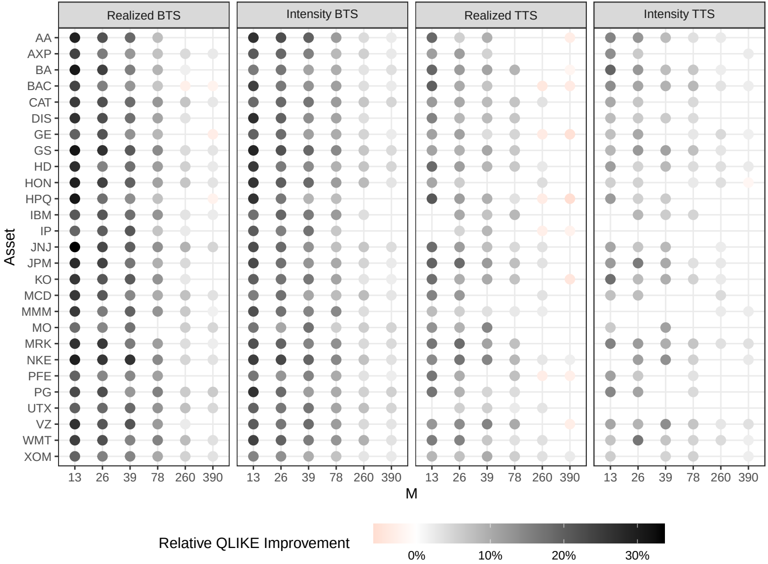
RV
TTS
RV
BTS
RV PAVG
TTS
RV PAVG
BTS
No noise
i.i.d. noise
diurnal ARMA noise
Relative RMSE
0.50000
0.25000
0.12500
0.06250
0.03125
1.000
0.500
0.250
0.125
0.500
0.250
0.125
26
78
390
1170
4680
M
Sampling
True intensities
Rolling intensities
Realized (true varsigma)
Realized (rolling varsigma)

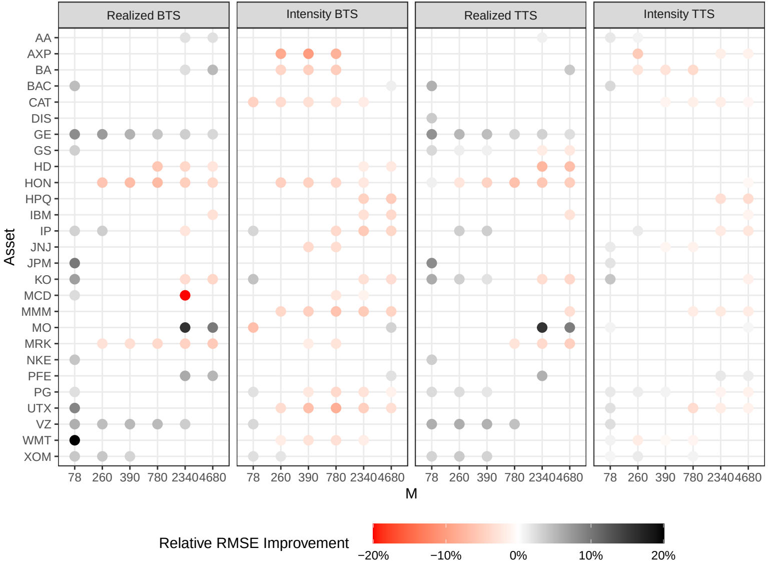

independent
TTS
independent
BTS
leverage
TTS
leverage
BTS
No noise
i.i.d. noise
diurnal ARMA noise
Relative RMSE
0.25000
0.12500
0.06250
0.03125
1.000
0.500
0.250
0.125
0.500
0.250
0.125
26
78
390
1170
4680
M
Sampling
True intensities
Rolling intensities
Realized (true varsigma)
Realized (rolling varsigma)

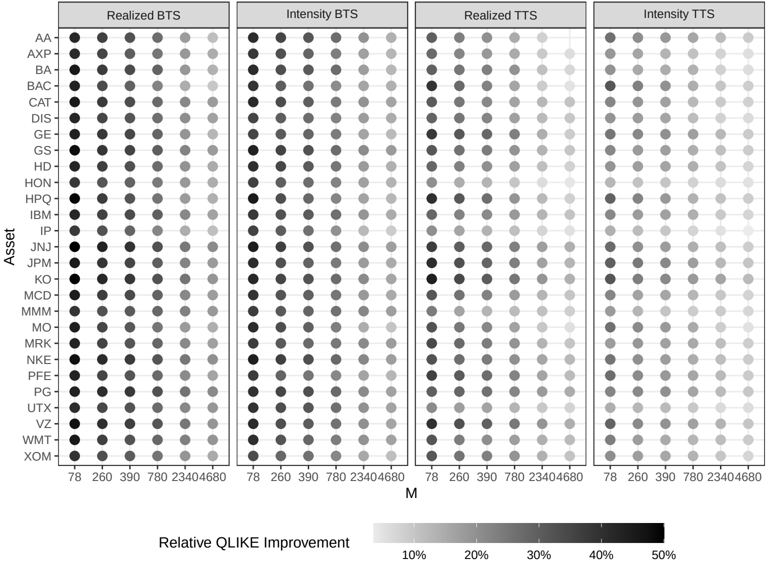
Data

- 27 liquid stocks traded at the NYSE
- Data from January 01, 2001 and March 31, 2019
- For RV, $M \in \{13, 26, 39, 78, 260, 390\}$
- For pre-averaged RV, $M \in \{78, 260, 390, 780, 2340, 4680\}$
- Evaluate RV estimates against a standard 5-minute CTS RV (Patton, 2011)


Realized BTS
Intensity BTS
Realized TTS
Intensity TTS
Asset
AA
AXP
BA
BAC
CAT
DIS
GE
GS
HD
HON
HPQ
IBM
IP
JNJ
JPM
KO
MCD
MMM
MO
MRK
NKE
PFE
PG
UTX
VZ
WMT
XOM
13
26
39
78
260
390
M
Relative RMSE Improvement
−10%
0%
10%


Realized BTS
Intensity BTS
Realized TTS
Intensity TTS
Asset
AA
AXP
BA
BAC
CAT
DIS
GE
GS
HD
HON
HPQ
IBM
IP
JNJ
JPM
KO
MCD
MMM
MO
MRK
NKE
PFE
PG
UTX
VZ
WMT
XOM
13 26 39 78 260 390
M
Relative QLIKE Improvement
0%
10%
20%
30%


Realized BTS
Intensity BTS
Realized TTS
Intensity TTS
Asset
AA
AXP
BA
BAC
CAT
DIS
GE
GS
HD
HON
HPQ
IBM
IP
JNJ
JPM
KO
MCD
MMM
MO
MRK
NKE
PFE
PG
UTX
VZ
WMT
XOM
78
260
390
780
2340
4680
M
Relative RMSE Improvement
−20%
−10%
0%
10%
20%


Realized BTS
Intensity BTS
Realized TTS
Intensity TTS
Asset
AA
AXP
BA
BAC
CAT
DIS
GE
GS
HD
HON
HPQ
IBM
IP
JNJ
JPM
KO
MCD
MMM
MO
MRK
NKE
PFE
PG
UTX
VZ
WMT
XOM
78 260 390 780 2340 4680
M
Relative QLIKE Improvement
10%
20%
30%
40%
50%

Evaluation target	Sampling	MSE			QLIKE		
		avg. rk.	avg.	winning	avg. rk.	avg.	winning
(a) RV CTS $M = 78$	rBTS	1.92	0.30	0.53	1.48	0.18	0.74
	iBTS	3.32	0.41	0.09	2.70	0.19	0.10
	rTTS	2.52	0.31	0.20	2.58	0.18	0.10
	iTTS	3.44	0.43	0.08	3.80	0.19	0.02
	CTS	3.81	0.49	0.10	4.44	0.21	0.04
(b) Squared return	rBTS	2.18	1.25	0.42	1.87	1.44	0.52
	iBTS	3.48	1.36	0.09	3.30	1.46	0.06
	rTTS	2.37	1.26	0.29	2.25	1.45	0.29
	iTTS	3.38	1.38	0.07	3.44	1.45	0.06
	CTS	3.59	1.44	0.12	4.14	1.47	0.06
(c) Individual estimator	rBTS	2.90	0.26	0.30	1.57	0.18	0.64
	iBTS	3.20	0.38	0.16	1.98	0.19	0.33
	rTTS	3.04	0.28	0.13	3.24	0.19	0.01
	iTTS	2.96	0.41	0.08	3.55	0.20	0.02
	CTS	2.90	0.48	0.33	4.65	0.22	0.01

Conclusion

- Exploit HF data from intrinsic time perspective to improve the estimation of daily variance.
- Show theoretically that (r)BTS provides the best RV estimates
- General price model: time-changed diffusion model
- Allow to disentangle **theoretically** between intraday different patterns
- Brings together two strands of literature: duration models and stochastic volatility
- (r)BTS: best RV in simulations and real data

Model Extensions

- Relaxation of independence assumptions of TTSV.
- Possibility of modelling $\lambda(t)$ to explicitly account for trade clustering (e.g. by self-exciting time-varying Hawkes process).
- Application to broader sets of asset classes and markets.

Thank you for your attention!