Optimum Thresholding Using Mean and Conditional Mean Squared Error

José E. Figueroa-López

Department of Mathematics and Statistics

Washington University in St. Louis

2019 SIAM Conference on Financial Mathematics and Engineering

Toronto, Canada

June 7, 2019

(Joint work with Cecilia Mancini from University of Florence)

Outline

1. The Setting and Problem
2. Optimal Threshold Selection
   - Via Expected Number of Jump Misclassifications
   - Via conditional Mean Square Error (cMSE)
3. Ongoing and Future Research

Integrated Variance Estimation

Given discrete observations $X_{t_1}, \ldots, X_{t_n}$ of a continuous semimartingale

$$dX_t = a_t dt + \sigma_t dW_t + dJ_t,$$

where J_t is a pure-jump process, we consider the problem of estimating the integrated variance,

$$IV_T = \int_0^T \sigma_s^2 ds,$$

in a high-frequency and fixed time horizon sampling scheme:

$$h_n := t_i - t_{i-1} \to 0, \quad t_n \equiv T.$$

Realized Quadratic Variation

In the absence of jumps, the efficient estimator is the realized quadratic variation:

$$\widehat{IV}_n := \sum_{i=0}^{n-1} \left(X_{t_{i+1}} - X_{t_i}\right)^2.$$

In the presence of jumps,

$$\widehat{IV}_n \to \int_0^T \sigma_s^2 ds + \sum_{s \leq T} (\Delta X_s)^2.$$

So, the problem lies on finding ways to “disentangle" the jumps from the continuous component of the process.

Truncated Realized Variations (Mancini, 2003)

$$TRV_n(\varepsilon) := \sum_{i=0}^{n-1} \left(X_{t_{i+1}} - X_{t_i}\right)^2 \mathbf{1}_{\left\{\left|X_{t_{i+1}} - X_{t_i}\right| \leq \varepsilon_n\right\}},$$

where ε_n, called the threshold, is a parameter to be tuned by the statistician.

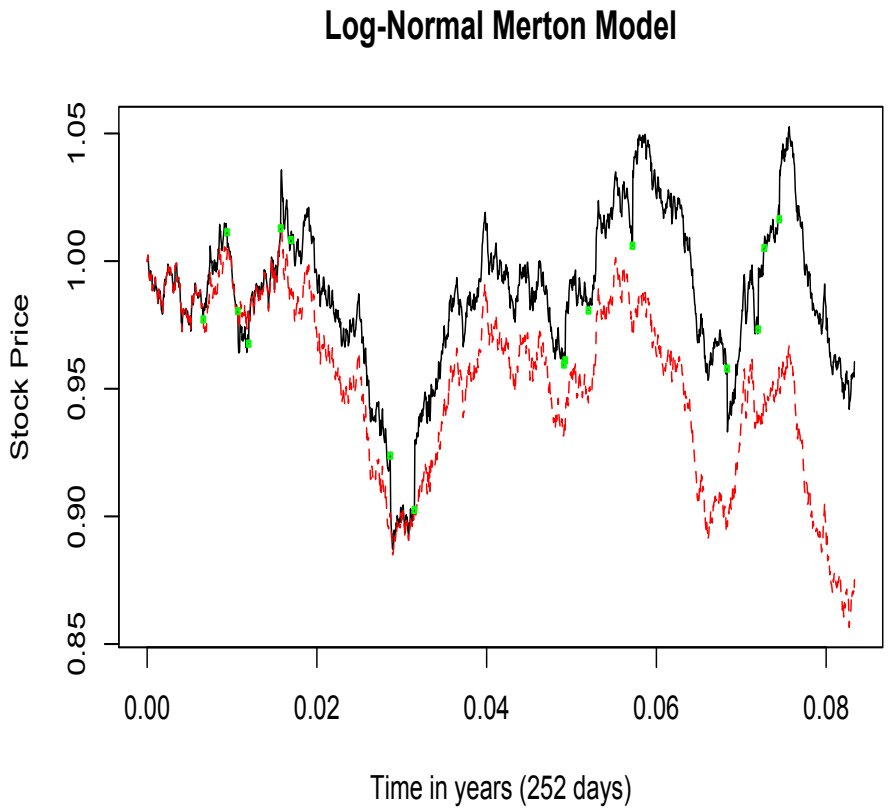
Idea: In a high-frequency setting, $X_{t_{i+1}} - X_{t_i}$ would be small in the absence of jumps. So, “large" increments are most likely due to jumps and should be removed from the realized variation.

Key Questions:

- How large is large?
- How do we choose the threshold parameter ε_n?

Illustration I: Log-normal Merton Model

$$X_t = at + \sigma W_t + \sum_{i:\tau_i \le t} \zeta_i, \quad \zeta_i \overset{\text{i.i.d.}}{\sim} N(\mu_{jmp}, \sigma^2_{jmp}), \quad \{\tau_i\}_{i \ge 1} \sim \text{Poisson}(\lambda)$$



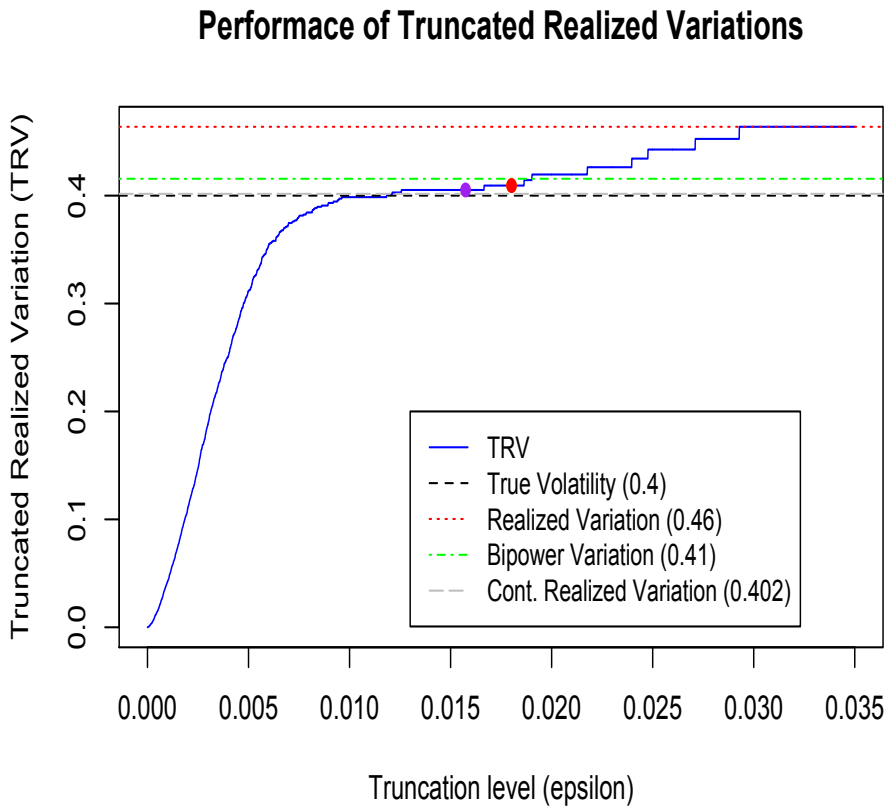
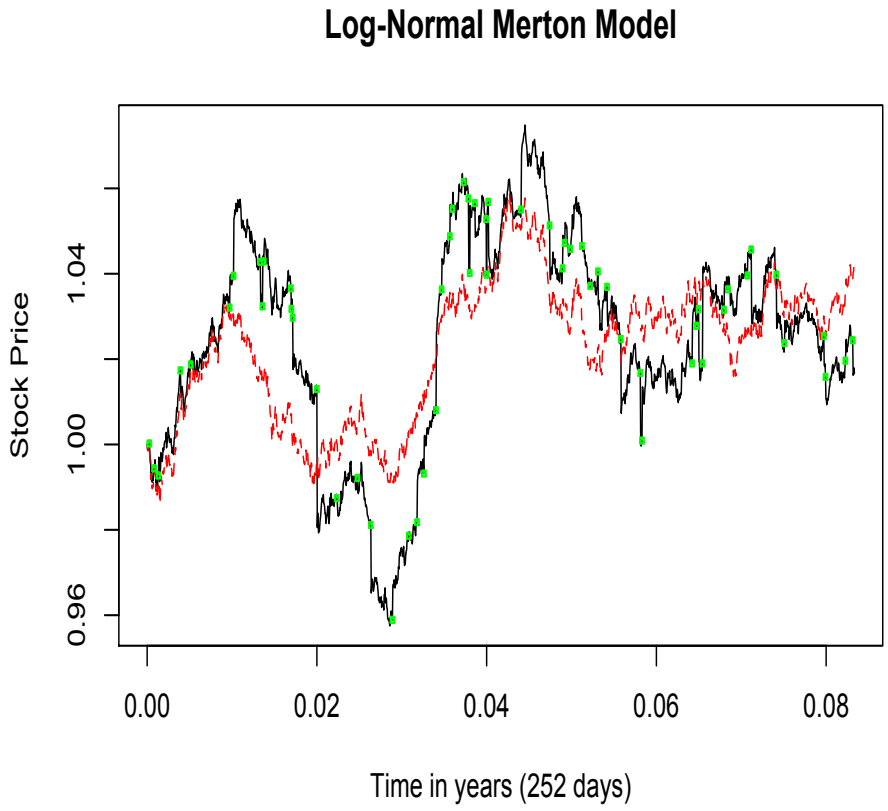


Figure: (left)1-month 5-min log-normal observations X_{t_i} with $\sigma = 0.4$, $\sigma_{jmp} = 3\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 200$; (right) TRV values wrt the truncation ε_n

Illustration II: Log-normal Merton Model



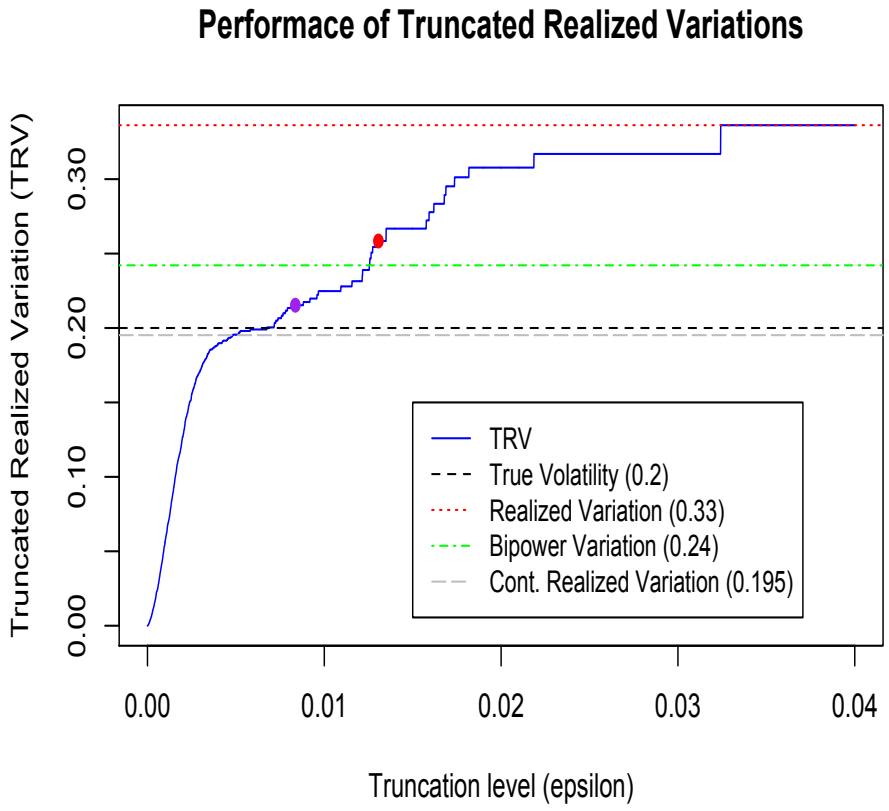


Figure: (left) 5 minute Merton observations with $\sigma = 0.2$, $\sigma_{jmp} = 1.5\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 1000$; (right) TRV performance wrt the truncation level

Approach For Threshold Selection

1. Fix a sensible metric of the estimation error;
2. Show the existence of an optimal threshold $\varepsilon_n^\star$ that minimizes the estimation error;
3. Analyze the (infill) asymptotic behavior $\varepsilon_n^\star$ (when $n \to \infty$) with the goal of
   - infer its explicit dependence on the underlying parameters of the model
   - devise iterative plug-in type calibrations of $\varepsilon_n^\star$ by estimating those parameters (if possible).

Via expected number of jump misclassifications

1. Notation:

$$\Delta_i^n X := X_{t_i} - X_{t_{i-1}}$$

$$\Delta_i^n N := \#\text{ of jumps during } (t_{i-1}, t_i]$$

2. Estimation Loss Function: (F-L & Nisen, SPA 2013)

$$Loss_n(\varepsilon) := \mathbb{E} \sum_{i=1}^{n} \left(\mathbf{1}_{[|\Delta_i^n X| > \varepsilon, \Delta_i^n N = 0]} + \mathbf{1}_{[|\Delta_i^n X| \leq \varepsilon, \Delta_i^n N \neq 0]} \right).$$

3. Underlying Principle

   The estimation error of the TRV depends on the ability of the threshold to detect jumps.

Existence and Infill Asymptotic Characterization

Theorem (FL & Nisen, SPA 2013)

Let X be a Finite Jump Activity Lévy process

$$X_t = at + \sigma W_t + \sum_{i=1}^{N_t} \xi_i.$$

1. *The loss function $\text{Loss}_n(\varepsilon)$ is quasi-convex with a unique global minimum $\varepsilon_n^\star$.*
2. *As $n \to \infty$, the optimal threshold sequence $(\varepsilon_n^\star)_n$ is such that*

$$\varepsilon_n^\star = \sqrt{3\sigma^2 h_n \log\left(\frac{1}{h_n}\right)} + \text{h.o.t.},$$

where hereafter h.o.t. *refers to 'higher order terms'.*

Remarks

1. Why $\sqrt{h \log(1/h)}$?

   This is proportional to modulus of continuity of the B.M.:

$$\limsup_{h \to 0} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{s,t \in [0,1]: |t-s| < h} |W_t - W_s| = 1.$$

2. Practically,

$$\varepsilon_n^{\star 1} := \sqrt{3\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}$$

   provides us with a “blueprint" for devising threshold sequences with good estimation properties!

A Feasible Implementation based on $\varepsilon_n^{\star 1}$

(i) Get a “rough" estimate of σ^2 via, e.g., the realized QV:

$$\hat{\sigma}^2_{n,0} := \frac{1}{T}\sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$

(ii) Use $\hat{\sigma}^2_{n,0}$ to estimate the optimal threshold

$$\hat{\varepsilon}^{\star}_{n,0} := \left(3\,\hat{\sigma}^2_{n,0} h_n \log(1/h_n)\right)^{1/2}$$

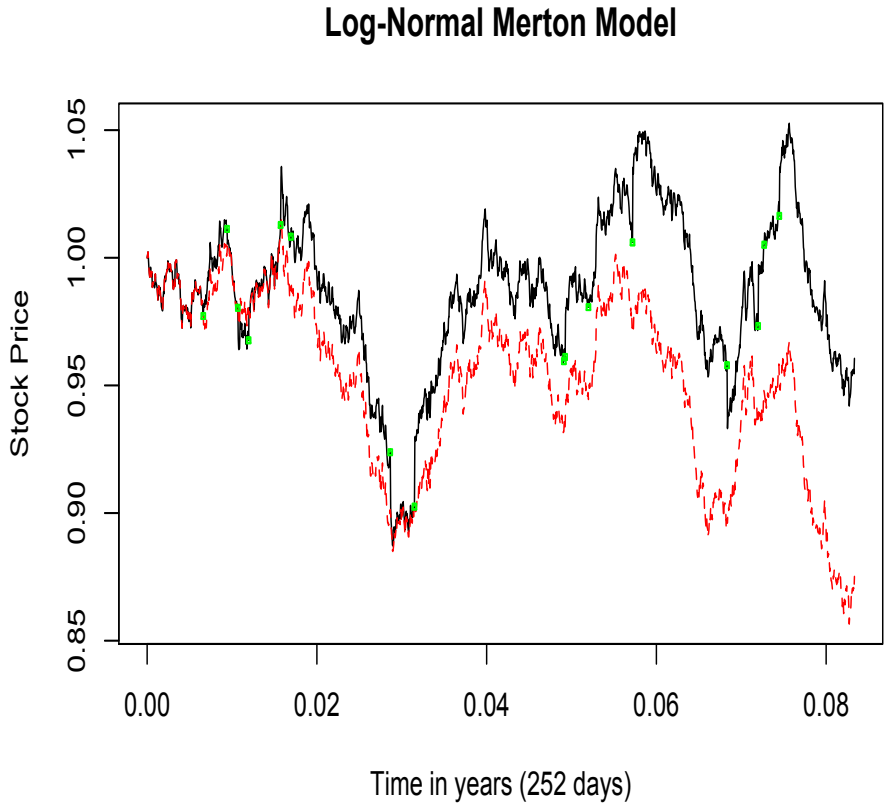
(iii) Refine $\hat{\sigma}^2_{n,0}$ using thresholding,

$$\hat{\sigma}^2_{n,1} = \frac{1}{T}\sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2 \mathbf{1}_{\left[|X_{t_i} - X_{t_{i-1}}| \le \hat{\varepsilon}^{\star}_{n,0}\right]}$$

(iv) Iterate Steps (ii) and (iii):

$$\hat{\sigma}^2_{n,0} \to \hat{\varepsilon}^{\star}_{n,0} \to \hat{\sigma}^2_{n,1} \to \hat{\varepsilon}^{\star}_{n,1} \to \hat{\sigma}^2_{n,2} \to \cdots \to \hat{\sigma}^2_{n,\infty}$$

Illustration I



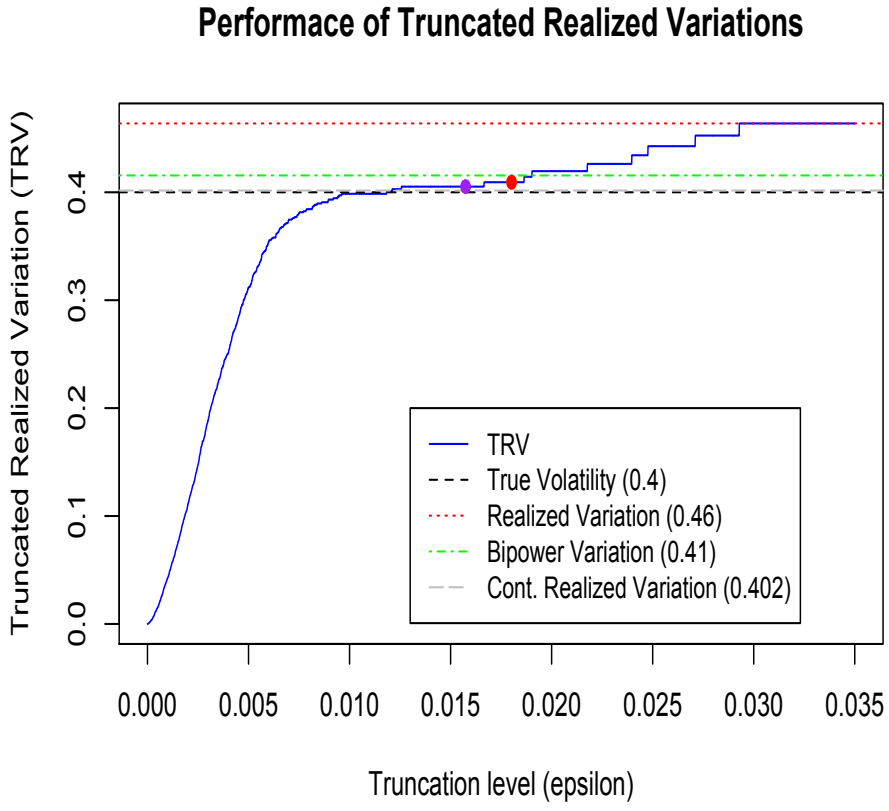
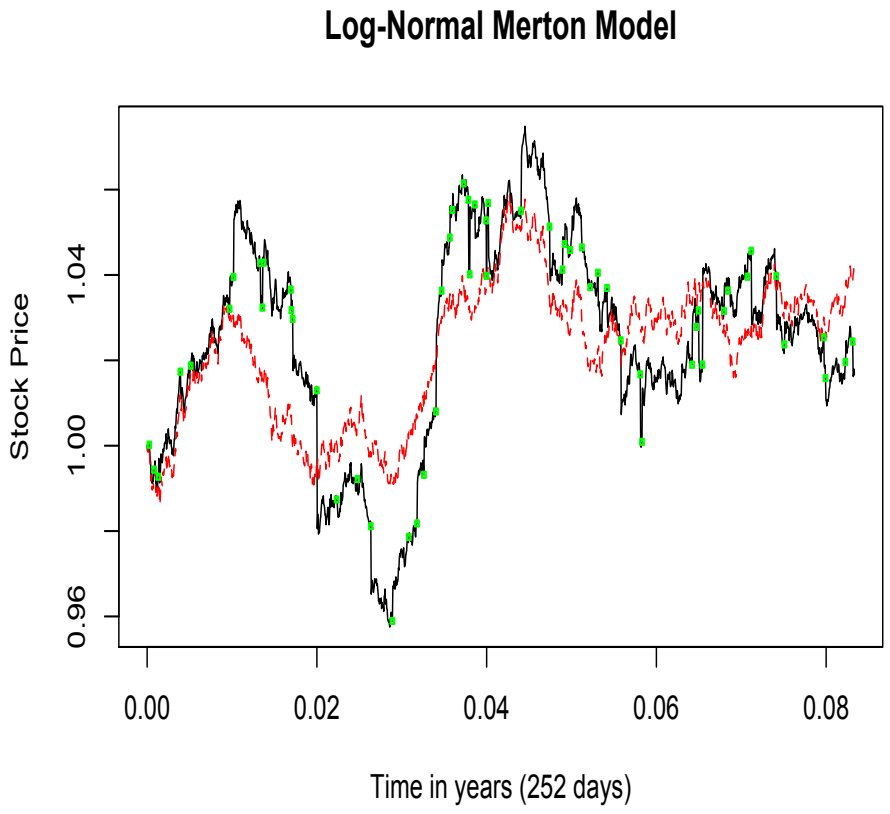


Figure: (left) Merton Model with $\sigma = 0.4$, $\sigma_{jmp} = 3\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 200$; (right) TRV performance wrt the truncation level. Red dot is $\hat{\sigma}_{n,1} = 0.409$, while purple dot is the limiting estimator $\hat{\sigma}_{n,\infty} = 0.405$

Illustration II



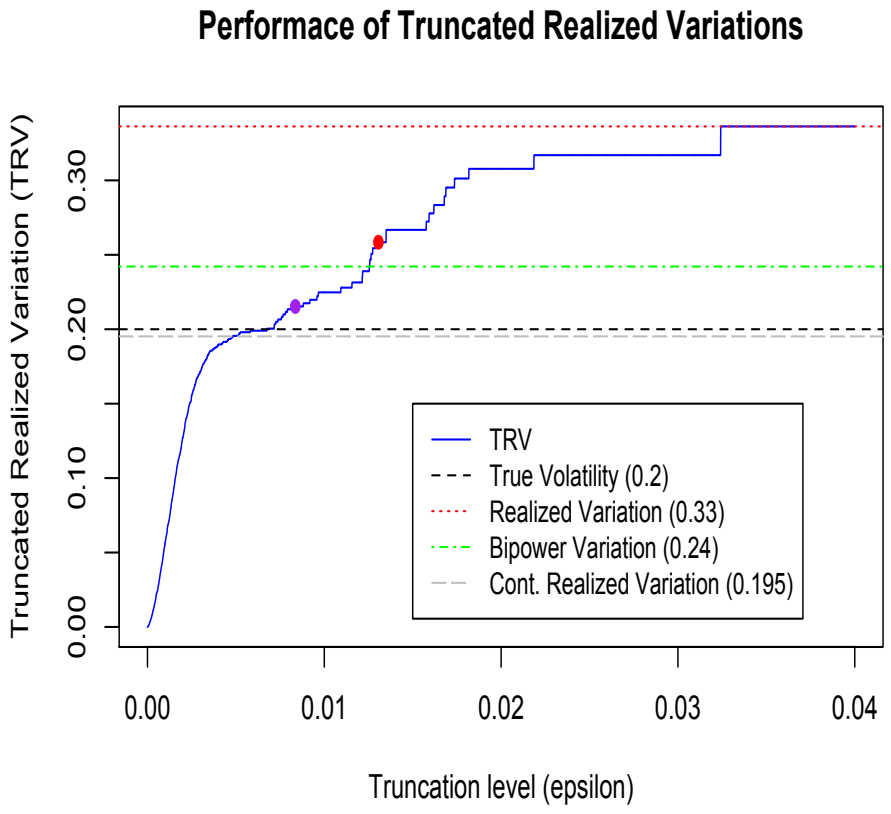


Figure: (left) Merton Model with $\sigma = 0.2$, $\sigma_{jmp} = 1.5\sqrt{h}$, $\mu_{jmp} = 0$, $\lambda = 1000$; (right) TRV performance wrt the truncation level. Red dot is $\hat{\sigma}_{n,1} = 0.336$, while purple dot is the limiting estimator $\hat{\sigma}_{n,\infty} = 0.215$

Via conditional Mean Square Error (cMSE)

1. We now propose a second approach in which we aim to control the conditional MSE:

$$MSE_c(\varepsilon) := \mathbb{E}\left[\left(TRV_n(\varepsilon) - \int_0^T \sigma_s^2 ds\right)^2 \middle| \sigma_\cdot, J_\cdot\right].$$

2. Assumptions:

   $\sigma_t > 0$, $\forall\ t$, and σ and J are independent of W (no leverage).

Infill Asymptotics under Finite Jump Activity

Theorem (F-L & Mancini (2019))

Suppose that $\sigma_t \equiv \sigma$ is constant and J is an arbitrary finite jump activity process:

$$X_t = \sigma W_t + \sum_{j=1}^{N_t} \zeta_j$$

Then, as $n \to \infty$, the optimal threshold $\varepsilon_n^{\star\star}$ is such that

$$\varepsilon_n^{\star\star} \sim \sqrt{2\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}$$

A Feasible Implementation of $\varepsilon_n^{\star\star}$

(i) Get a “rough" estimate of σ^2 via the realized QV:

$$\tilde{\sigma}^2_{n,0} := \frac{1}{T}\sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2$$

(ii) Use $\tilde{\sigma}^2_{n,0}$ to estimate the optimal threshold

$$\hat{\varepsilon}^{\star\star}_{n,0} := \left(2\,\tilde{\sigma}^2_{n,0} h_n \log(1/h_n)\right)^{1/2}$$

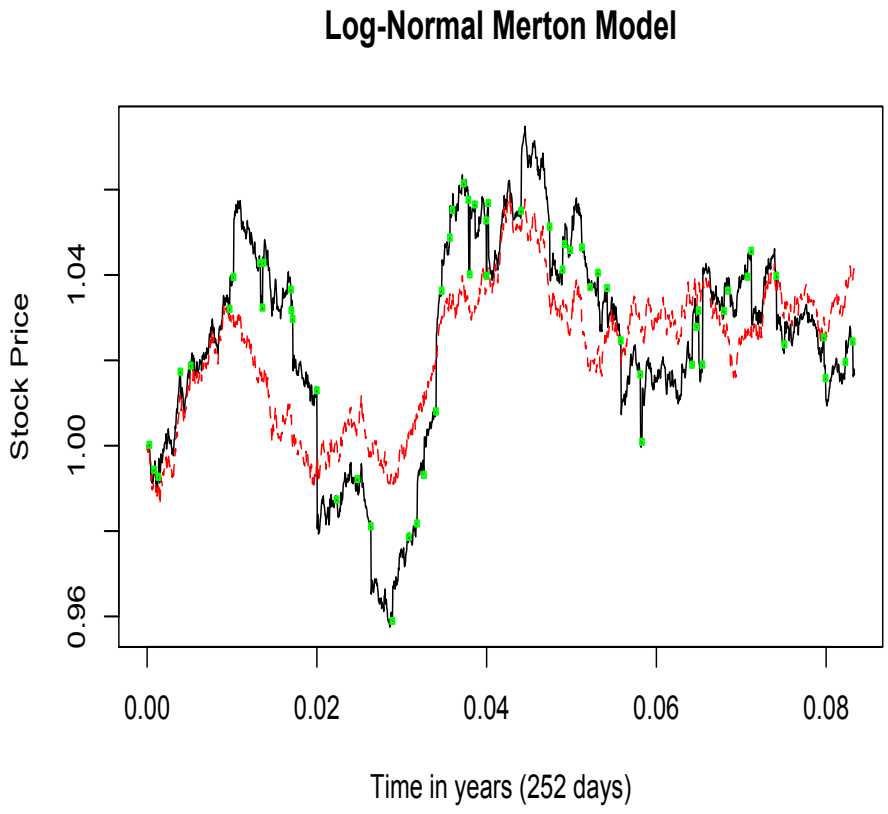
(iii) Refine $\tilde{\sigma}^2_{n,0}$ using thresholding,

$$\tilde{\sigma}^2_{n,1} = \frac{1}{T}\sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2 \mathbf{1}_{\left[|X_{t_i}-X_{t_{i-1}}| \le \hat{\varepsilon}^{\star\star}_{n,0}\right]}$$

(iv) Iterate Steps (ii) and (iii):

$$\tilde{\sigma}^2_{n,0} \to \hat{\varepsilon}^{\star\star}_{n,0} \to \tilde{\sigma}^2_{n,1} \to \hat{\varepsilon}^{\star\star}_{n,1} \to \tilde{\sigma}^2_{n,2} \to \cdots \to \tilde{\sigma}^2_{n,\infty}$$

Illustration II. Continued...



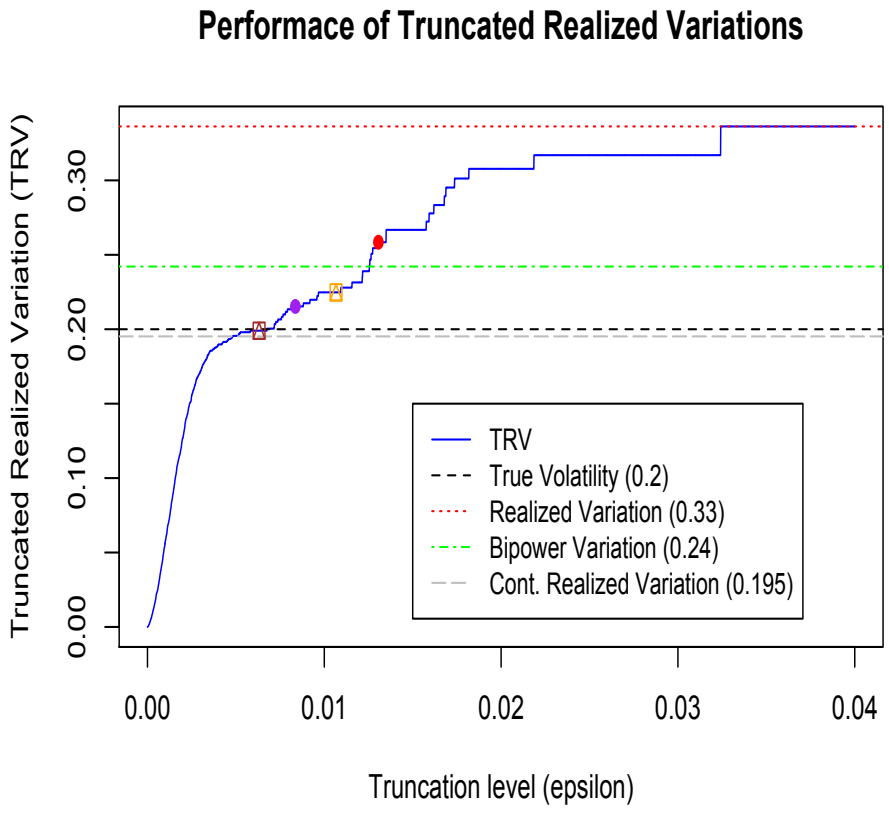


Figure: (left) Merton Model with $\lambda = 1000$. Red dot is $\hat{\sigma}_{n,1} = 0.336$, while purple dot is the limiting $\hat{\sigma}_{n,k} = 0.215$. Orange square is $\tilde{\sigma}_{n,1} = 0.225$, while brown square is the limiting estimator $\tilde{\sigma}_{n,\infty} = 0.199$

Monte Carlo Simulations

Estimator	$\overline{\hat{\sigma}}$	std$(\hat{\sigma})$	$\overline{\text{Loss}}$	$\overline{\varepsilon}$	$\overline{N}$
RQV	0.3921	0.0279			
$\hat{\sigma}_{n,1}$	0.29618	0.02148	70.1	0.01523	1
$\hat{\sigma}_{n,\infty}$	0.23	0.0108	49.8	0.00892	5.86
$\tilde{\sigma}_{n,1}$	0.265	0.0163	62.6	0.0124	1
$\tilde{\sigma}_{n,\infty}$	0.211	0.00588	39.1	0.00671	5.10
BPV	0.2664	0.0129			

Table: Estimation of the volatility $\sigma = 0.2$ for a log-normal Merton model based on 10000 simulations of 5-minute observations over a 1 month time horizon. The jump parameters are $\lambda = 1000$, $\sigma^{Jmp} = 1.5\sqrt{h}$ and $\mu^{Jmp} = 0$. Loss is the number of jump misclassifications and N is the number of iterations. bar is used to denote average.

Problems

1. What can we do in the case of time-varying or stochastic volatility $t \to \sigma_t$?
2. What can we say for infinite jump activity processes J?
3. Can we introduce leverage (correlation between the price process S and the volatility process σ)?

Analysis: Finite-Jump Activity J

Let

$$b_i(\varepsilon) := \mathbb{E}\left[(\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \le \varepsilon\}} | \sigma_\cdot, J_\cdot\right]$$

We have that

$$b_i(\varepsilon) = -\frac{\sigma_i}{\sqrt{2\pi}}\left(e^{-\frac{(\varepsilon - m_i)^2}{2\sigma_i^2}}(\varepsilon + m_i) + e^{-\frac{(\varepsilon + m_i)^2}{2\sigma_i^2}}(\varepsilon - m_i)\right)$$
$$+ (m_i^2 + \sigma_i^2)\int_{\frac{\varepsilon - m_i}{\sigma_i}}^{\frac{\varepsilon + m_i}{\sigma_i}} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

where $\sigma_i^2 = \int_{t_{i-1}}^{t_i} \sigma_s^2 ds$ and $m_i = J_{t_i} - J_{t_{i-1}}$.

Furthermore, $b_i(\varepsilon)$ has the following remarkable property:

$$\frac{db_i(\varepsilon)}{d\varepsilon} = \varepsilon^2 a_i(\varepsilon), \quad \text{with} \quad a_i(\varepsilon) := \frac{e^{-\frac{(\varepsilon - m_i)^2}{2\sigma_i^2}} + e^{-\frac{(\varepsilon + m_i)^2}{2\sigma_i^2}}}{\sigma_i\sqrt{2\pi}}$$

Conditional Mean Square Error (MSE_c)

Theorem (F-L & Mancini (2017))

The conditional MSE,

$$MSE_c(\varepsilon) := \mathbb{E}\left[(TRV_n(\varepsilon) - IV)^2 \middle| \sigma, J \right],$$

is such that

$$\frac{d\, MSE_c(\varepsilon)}{d\varepsilon} = \varepsilon^2 G(\varepsilon),$$

where

$$G(\varepsilon) := \sum_i a_i(\varepsilon)\left(\varepsilon^2 + 2\sum_{j \neq i} b_j(\varepsilon) - 2IV \right).$$

Furthermore, there exists an optimal threshold $\varepsilon_n^{\star\star}$ that minimizes $MSE_c(\varepsilon)$ and is such that $G(\varepsilon_n^{\star\star}) = 0$.

Asymptotic Behavior of $\varepsilon_n^{\star\star}$ I

In the case that $m_j = 0$, we have that

$$b_j(\varepsilon) = \sigma_j^2 - \frac{2\sigma_j}{\sqrt{2\pi}}\varepsilon e^{-\frac{\varepsilon^2}{2\sigma_j^2}} + \text{h.o.t.},$$

Using this result,

$$\begin{aligned}
g_i(\varepsilon) &:= \varepsilon^2 + 2\sum_{j\neq i} b_j(\varepsilon) - 2\sum_{j=1}^{n}\sigma_j^2 \\
&= \varepsilon^2 + 2\sum_{j\neq i:m_j=0}\left[b_j(\varepsilon) - \sigma_j^2\right] + 2\sum_{j\neq i:m_j\neq 0} b_j(\varepsilon) - 2\sum_{j:m_j\neq 0}\sigma_j^2 \\
&= \varepsilon^2 - 4\varepsilon\sum_{j\neq i:m_j=0}\frac{\sigma_j}{\sqrt{2\pi}}e^{-\frac{\varepsilon^2}{2\sigma_j^2}} + \text{h.o.t.} \\
&= \varepsilon^2 - 4\varepsilon\sum_{j=1}^{n}\frac{\sigma_j}{\sqrt{2\pi}}e^{-\frac{\varepsilon^2}{2\sigma_j^2}} + \text{h.o.t.}
\end{aligned}$$

Asymptotic Behavior of $\varepsilon_n^{\star\star}$ II

Then, up to higher-order terms, $\varepsilon_n^{\star\star}$ is a solution of the equation:

$$\varepsilon_n - 4\sum_{j=1}^{n} \frac{\sigma_j}{\sqrt{2\pi}} e^{-\frac{\varepsilon_n^2}{2\sigma_j^2}} = 0.$$

In the case of constant σ, we have:

$$\varepsilon_n^{\star\star} - 4n\sqrt{h_n}\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\varepsilon_n^{\star\star 2}}{2\sigma^2 h_n}} + \text{h.o.t.} = 0,$$

which implies

$$\varepsilon_n^{\star\star} \sim \sqrt{2\sigma^2 h_n \ln(1/h_n)}, \quad n \to \infty.$$

Asymptotic Behavior of $\varepsilon_n^{\star\star}$ III

In the general case, we can proceed as follows (recall that $\sigma_j^2 := \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \sigma_{s_j}^2 h_n$, for some $s_{j_n} \in (t_{j-1}, t_j)$):

$$
\begin{aligned}
0 &= \varepsilon_n - 4 \sum_{j=1}^{n} \frac{\sigma_j}{\sqrt{2\pi}} e^{-\frac{\varepsilon_n^2}{2\sigma_j^2}} + \text{h.o.t.} \\
&= \varepsilon_n - 4 \frac{1}{\sqrt{h_n}} \sum_{j=1}^{n} \frac{\sigma_{s_j}}{\sqrt{2\pi}} e^{-\frac{\varepsilon_n^2}{2\sigma_{s_j}^2 h_n}} h_n + \text{h.o.t.} \\
&= \varepsilon_n - \frac{4}{\sqrt{h_n}} \int_0^T \frac{\sigma_s}{\sqrt{2\pi}} e^{-\frac{w_n}{2\sigma_s^2}} ds + \text{h.o.t.}, \qquad w_n := \varepsilon_n^2 / h_n.
\end{aligned}
$$

Generalization II

It can be shown that $w_n = \varepsilon_n^2 / h_n \to \infty$. Then, by the Laplace method,

$$\int_0^T \frac{\sigma_s}{\sqrt{2\pi}} e^{-\frac{w_n}{2\sigma_s^2}} ds \sim \frac{\sigma_{s_0}^{5/2}}{\sqrt{|\sigma''_{s_0}|}} \frac{1}{\sqrt{w_n}} e^{-\frac{w_n}{2\sigma_{s_0}^2}}$$

where $\sigma_{s_0} := \max_{s \in [0,T]} \sigma_s$.

Then, for some constant K and $\bar{\sigma} := \max_{s \in [0,T]} \sigma_s$,

$$0 = \varepsilon_n - K \frac{1}{\varepsilon_n} e^{-\frac{\varepsilon_n^2}{2\bar{\sigma}^2 h_n}} + \text{h.o.t.},$$

which again implies

$$\varepsilon_n^{\star\star} \sim \sqrt{2\bar{\sigma}^2 h_n \ln(1/h_n)}, \quad n \to \infty.$$

Ongoing and Future Research

1. In principle, we can apply the constant-volatility method for varying volatility $t \to \sigma_t$ by localization; i.e., applying it to periods where σ is approximately constant.
2. This may also help us to estimate $\bar{\sigma} = \max_{s \in [0,T]} \sigma_s$ and, then, apply the asymptotic equivalence

$$\varepsilon_n^{\star\star} \sim \sqrt{2\bar{\sigma}^2 h_n \ln(1/h_n)}.$$

3. Of course, having twice differentiable volatility, as needed for the Laplace method, is too strong. Can we relax such a condition? In other words, what is the asymptotic behavior of

$$\int_0^T \frac{\sigma_s}{\sqrt{2\pi}} e^{-\frac{w_n}{2\sigma_s^2}} ds$$

as $w_n \to \infty$ for different classes of volatility processes σ?

Ongoing and Future Research

4. As it turns, for a Lévy process J and constant σ, the expected mean square error, $MSE(\varepsilon) := \mathbb{E}\left[(TRV_n(\varepsilon) - IV)^2\right]$, is such that

$$\frac{d}{d\varepsilon} MSE(\varepsilon) = n\varepsilon^2 \mathbb{E}[a_1(\varepsilon)]\left(\varepsilon^2 + 2(n-1)\mathbb{E}[b_1(\varepsilon)] - 2nh_n\sigma^2\right)$$

Therefore, there exists a unique minimum point $\bar{\varepsilon}_n^\star$ which is the solution of the equation

$$\varepsilon^2 + 2(n-1)\mathbb{E}[b_1(\varepsilon)] - 2nh_n\sigma^2 = 0.$$

Ongoing and Future Research

5. It can be shown that
   - for a finite jump activity process,
   $$\bar{\varepsilon}_n^\star \sim \sqrt{2\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}$$
   - But, surprisingly, if J is a Y-stable Lévy process (IA),
   $$\bar{\varepsilon}_n^\star \sim \sqrt{(2-Y)\sigma^2 h_n \log\left(\frac{1}{h_n}\right)}$$

   Thus the higher the jump activity is, the lower the optimal threshold has to be to discard the higher noise represented by the jumps.

Problems: Does the above asymptotics holds for the minimizer $\varepsilon_n^\star$ of the cMSE? Can we generalize it to Lévy processes with stable like jumps?

References

J.E. Figueroa-López & C. Mancini.

Optimum thresholding using mean and conditional mean square error.

Journal of Econometrics Vol. 208, issue 1, p. 179-210, 2019.

J.E. Figueroa-López & J. Nisen.

Optimally Thresholded Realized Power Variations for Lévy Jump Diffusion Models.

Stochastic Processes and their Applications 123(7), 2648-2677, 2013.