

www.portfolioeffect.com
High Frequency Portfolio Analytics

User Manual

PortfolioEffectEstim MATLAB Toolbox

High Frequency Price Estimators & Models

Andrey Kostin
andrey.kostin@portfolioeffect.com

*Released Under BSD License
by Snowfall Systems, Inc.*

August 20, 2015

Contents

| | |
|---|----------|
| Contents | 1 |
| 1 Toolbox Installation | 2 |
| 1.1 Zip Archive (All OS) | 2 |
| 1.2 Install Wizard (Windows) | 2 |
| 2 Account Credentials | 3 |
| 2.1 Locate API Credentials | 3 |
| 2.2 Set API Credentials in MATLAB | 4 |
| 3 Price Variance | 5 |
| 3.1 Integrated Variance | 5 |
| 3.1.1 Returns | 5 |
| 3.2 Realized Variance | 5 |
| 3.2.1 Assumptions | 5 |
| 3.2.2 Estimator | 6 |
| 3.2.3 Properties | 6 |
| 3.3 Two Series Realized Variance | 6 |
| 3.3.1 Assumptions | 6 |
| 3.3.2 Estimator | 6 |
| 3.3.3 Properties | 7 |
| 3.4 Multiple Series Realized Variance | 7 |
| 3.4.1 Assumptions | 7 |
| 3.4.2 Estimator | 8 |

| | | |
|----------|---|-----------|
| 3.4.3 | Properties | 8 |
| 3.5 | Modulated Realized Variance | 8 |
| 3.5.1 | Assumptions | 8 |
| 3.5.2 | Estimator | 8 |
| 3.5.3 | Properties | 9 |
| 3.6 | Jump Robust Modulated Realized Variance | 9 |
| 3.6.1 | Assumptions | 9 |
| 3.6.2 | Estimator | 9 |
| 3.6.3 | Properties | 10 |
| 3.7 | Kernel-based Realized Variance | 10 |
| 3.7.1 | Assumptions | 10 |
| 3.7.2 | Kernel Types | 11 |
| 3.7.3 | Estimator | 11 |
| 3.7.4 | Properties | 12 |
| 3.8 | IV Estimators Comparison | 13 |
| 4 | Price Noise Variance | 14 |
| 4.1 | RV Noise Variance | 14 |
| 4.1.1 | Properties | 14 |
| 4.2 | Autocovariance Noise Variance | 14 |
| 4.2.1 | Properties | 15 |
| 5 | Price Quarticity | 16 |
| 5.1 | Integrated Quarticity | 16 |
| 5.2 | Realized Quarticity | 16 |
| 5.2.1 | Assumptions | 16 |
| 5.2.2 | Estimator | 16 |
| 5.2.3 | Properties | 16 |

1 Toolbox Installation

PortfolioEffect Estimators Toolbox for MATLAB is provided as a zip archive for all types of operating systems and as a self-install executable file for Windows.

1.1 Zip Archive (All OS)

Download zip archive with a toolbox from PortfolioEffect downloads or MATLAB Central File Exchange.

Once downloaded, unpack archive into the folder and add PortfolioEffectEstim folder to MATLAB's path using "Set Path" menu. Then call any method of the package in MATLAB editor to continue with the set-up:

```
+++++
Welcome to PortfolioEffect Estimators Toolbox.

Setup will download required binary files (~5mb).
Please, wait...
SUCCESS. File downloaded to:
/home/appadmin/.matlab/R2015a/portfolioeffect-quant-client-1.0-allinone.jar
Updating java class path file...
SUCCESS. Java class path updated.

Setup complete! Restart Matlab session now.
+++++
```

Restart MATLAB to complete installation.

1.2 Install Wizard (Windows)

Download self-installer executable for Windows from PortfolioEffect downloads. Follow the installation instructions. Once the wizard completes, add the PortfolioEffectEstim folder to MATLAB's path using "Set Path" menu. The PortfolioEffect Estimators Toolbox is now fully configured.

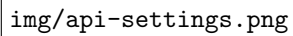
2 Account Credentials

All portfolio computations are performed on PortfolioEffect cloud servers. To obtain a free non-professional account, you need to follow a quick sign-up process on our website: www.portfolioeffect.com/registration.

Please use a valid sign-up address - it will be used to email your account activation link.

2.1 Locate API Credentials

Log in to you account and locate your API credentials on the main page

A large rectangular box representing a screenshot of a web page. Inside the box, the text 'img/api-settings.png' is visible in the bottom-left corner, indicating the location of the API settings within the application's image assets.

img/api-settings.png

2.2 Set API Credentials in MATLAB

Run the following commands to set your account API credentials for the PortfolioEffect Estimators Toolbox for MATLAB. You will need to do it only once as your credentials are stored between sessions on your local machine to speed up future logons.

You would need to repeat this procedure if you change your account password or install PortfolioEffect toolbox on another computer.

```
util_setCredentials('API Username', 'API Password', 'API Key');
```

You are now ready to call PortfolioEffect methods.

3 Price Variance

3.1 Integrated Variance

Assume that the logarithmic equilibrium price of a financial asset is given by the following diffusion process

$$X_t = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s) \quad (3.1)$$

where

- W_t is a standard Brownian Motion,
- the mean process μ is continuous and of finite variation,
- $\sigma(t) > 0$ denotes the cadlag instantaneous volatility.

The object of interest is the integrated variance (IV), i.e. the amount of variation at time point t accumulated over a past time interval Δ according to [?, Pigorsch et al.]:

$$IV_t = \int_{t-\Delta}^t \sigma^2(s)ds \quad (3.2)$$

3.1.1 Returns

Suppose there exist m intraday equilibrium returns, the i th intraday return is then defined as:

$$r_i^{X(m)} = X_{i/m} - X_{(i-1)/m}, \quad i = 1, 2, \dots, m. \quad (3.3)$$

3.2 Realized Variance

3.2.1 Assumptions

The equilibrium price process.

1. The logarithmic equilibrium price process p_t^X is a continuous stochastic volatility semimartingale. Specifically,

$$p_t^X = \alpha_t + m_t; \quad (3.4)$$

where α_t (with $\alpha_0 = 0$) is a predictable drift process of finite variation and m_t is a continuous local martingale defined as $\int_0^t \sigma_s dW_s$ with $W_t : t \geq 0$ denoting standard Brownian motion.

2. The spot volatility process σ_t is cadlag and bounded away from zero.
3. The process $\int_0^t \sigma_s^4 ds$ is bounded almost surely for all $t < \infty$.

3.2.2 Estimator

For discretely observed path X_{t_i} , $i = 0, \dots, n$ of the X , the realized quadratic variation could be estimated consistently using the realized variance measure, defined as [?, Zu and Boswijk, 2008]:

$$RV_t = \sum_{i=1}^n (X_{t_i} - X_{t_{(i-1)}})^2 \xrightarrow{P} IV \quad (3.5)$$

However the realized variance estimator for integrated volatility is not consistent when data is contaminated by market microstructure noise. In particular, when we only observe data with microstructure noise, the realized variance measure will diverge. When sampling frequency increases, realized variance actually estimates the sum of infinite many variances of noises.

3.2.3 Properties

- Convergence speed: $m^{1/2}$ (m - number of observation)
- Unbiased: no
- Consistent: no
- Jump Robust: no
- Allows for time dependence noise: no
- Allows for endogenous noise: no

3.3 Two Series Realized Variance

3.3.1 Assumptions

The microstructure noise.

1. The microstructure noise, $\epsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.
2. The noise is independent of the price process.
3. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.

3.3.2 Estimator

Two Scale Realized Variance (TSRV) estimates integrated volatility consistently. The idea is to use realized variance type estimators over two time scales to correct the effect of market microstructure noise. Define as:

$$[Y, Y]_t^{avg} = \frac{1}{K} \sum_{i=K}^n (Y_{t_i} - Y_{t_{(i-K)}})^2 \quad (3.6)$$

$$[Y, Y]_t^{all} = \sum_{i=1}^n (Y_{t_i} - Y_{t_{(i-1)}})^2 \quad (3.7)$$

$$\bar{n} = \frac{n - K + 1}{K}, \quad (3.8)$$

the TSRV estimator is defined as [?, Zu and Boswijk, 2008]:

$$TSRV_t = [Y, Y]_t^{(avg)} - \frac{\bar{n}}{n} [Y, Y]_t^{(all)} \xrightarrow{p} IV \quad (3.9)$$

A small sample refinement to the estimator give correction:

$$TSRV_t^{adjust} = \left(1 - \frac{\bar{n}}{n}\right)^{-1} TSRV_t \xrightarrow{p} IV. \quad (3.10)$$

If we have possibly dependent noise we should use an alternative estimator that is also based on the two time scales idea.

To pin down the optimal sampling frequency K [?, Zhang, Mykland, and Ait-Sahalia, 2005], one can minimize the expected asymptotic variance and to obtain

$$c^* = \left(\frac{16(E\epsilon^2)^2}{TE\eta^2} \right)^{1/3} \quad (3.11)$$

which can be consistently estimated from data in past time periods (before time $t_0 = 0$), using $E\epsilon^2$ and an estimator of η^2 . η^2 can be taken to be independent of K as long as one allocates sampling points to grids regularly. Hence one can choose c , and so also K , based on past data.

3.3.3 Properties

- Convergence speed: $m^{1/6}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: no
- Allows for time dependence noise: no
- Allows for endogenous noise: no

3.4 Multiple Series Realized Variance

3.4.1 Assumptions

Dependent Noise Structure [?, Podolskij and Vetter, 2009]:

1. The microstructure noise, $\epsilon_{t,i}$, has a zero mean, stationary, and strong mixing stochastic process, with the mixing coefficients decaying exponentially. In addition, $E[(\epsilon_{t,i})^{4+\kappa}]$, for some $\kappa > 0$.
2. The noise is independent of the price process.
3. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.

3.4.2 Estimator

Under most assumptions, this estimator violates the sufficiency principle, whence we define the average lag j realized volatility as

$$[Y, Y]_T^{(J)} = \frac{1}{J} \sum_{r=0}^{J-1} [Y, Y]_T^{(J,r)} = \frac{1}{J} \sum_{t=0}^{n-J} (Y_{t+J} - Y_t)^2 \quad (3.12)$$

A generalization of TSRV can be defined for $1 \leq J < K \leq n$ as

$$MSRV_t = [Y, Y]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} [Y, Y]_T^{(J)} \xrightarrow{p} IV \quad (3.13)$$

thereby combining the two time scales J and K . Here $\bar{n}_K = (n - K + 1)/K$ and similarly for \bar{n}_J .

3.4.3 Properties

- Convergence speed: $m^{1/4}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: no
- Allows for time dependence noise: yes
- Allows for endogenous noise: no

3.5 Modulated Realized Variance

3.5.1 Assumptions

The microstructure noise.

1. The microstructure noise, $\epsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.
2. The noise is independent of the price process.
3. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.

3.5.2 Estimator

Modulated Bipower Variation is written as [?, Podolskij and Vetter, 2009]:

$$MBV(Y, r, l)_n = n^{(r+l)/4-1/2} \sum_{m=1}^M |\bar{Y}_m^{(K)}|^r |\bar{Y}_{m+1}^{(K)}|^l, \quad r, l \geq 0; \quad (3.14)$$

$$Y_m^{(K)} = \frac{1}{n/M - K + 1} \sum_{i=(m-1)n/M}^{mn/M-K} (Y_{(i+K)/n} - Y_{i/n}) \quad (3.15)$$

with

$$K = c_1 n^{1/2}, \quad M = \frac{n}{c_2 K} = \frac{n^{1/2}}{c_1 c_2} \quad (3.16)$$

We can choose the constants c_1 and c_2 from specific process:

$$c_1 = 0.25, \quad c_2 = 2. \quad (3.17)$$

Modulated Realized Variance is written as:

$$MRV(Y)_n = \frac{c_1 c_2 MBV(Y, 2, 0)_n - \nu_2 \hat{\omega}^2}{\nu_1} \xrightarrow{p} IV \quad (3.18)$$

where

$$\nu_1 = \frac{c_1(3c_2 - 4 + (2 - c_2)^3 \vee 0)}{3(c_2 - 1)^2} \quad (3.19)$$

$$\nu_2 = \frac{2((c_2 - 1) \wedge 1)}{c_1(c_2 - 1)^2} \quad (3.20)$$

3.5.3 Properties

- Convergence speed: $m^{1/4}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: no
- Allows for time dependence noise: no
- Allows for endogenous noise: no

3.6 Jump Robust Modulated Realized Variance

3.6.1 Assumptions

The microstructure noise.

1. The microstructure noise, $\epsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.
2. The noise is independent of the price process.
3. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.
4. The price is $Z = Y + J$ where Y is a noisy diffusion process and J denotes a finite activity jump process, that is, J exhibits finitely many jumps on compact intervals.

3.6.2 Estimator

We can construct consistent estimates for the integrated volatility, which are robust to noise and finite activity jumps [?, Podolskij and Vetter, 2009].

$$MBV(Y, r, l)_n = \frac{(c_1 c_2 / \mu_1^2)(MBV(Z, 1, 1)_n - \nu_2 \hat{\omega}^2)}{\nu_1} \xrightarrow{p} IV \quad (3.21)$$

where

$$\nu_1 = \frac{c_1(3c_2 - 4 + (2 - c_2)^3 \vee 0)}{3(c_2 - 1)^2} \quad (3.22)$$

$$\nu_2 = \frac{2((c_2 - 1) \wedge 1)}{c_1(c_2 - 1)^2} \quad (3.23)$$

Mixed normal distribution with conditional variance [?, Podolskij and Vetter, 2009]:

$$\beta^2 = \frac{2c_1c_2}{\nu_1^2} \int_0^1 (\nu_1\sigma_u^2 + \nu_2\omega^2)^2 du \quad (3.24)$$

Consistent estimator of β is:

$$\beta_n^2 = \frac{2c_1^2c_2^2}{3\nu_1^2} MBV(Y, 4, 0)_n \quad (3.25)$$

We can choose the constants c_1 and c_2 that minimise the conditional variance. In order to compare our asymptotic variance with the corresponding results of other methods we assume that the volatility process σ is constant. In that case the conditional variance β^2 is minimised by

$$c_1 = \sqrt{\frac{18}{(c_2 - 1)(4 - c_2)}} \cdot \frac{\omega}{\sigma}, \quad c_2 = \frac{8}{5} \quad (3.26)$$

We can choose the constants c_1 and c_2 from specific process:

$$c_1 = 0.25, \quad c_2 = 2. \quad (3.27)$$

3.6.3 Properties

- Converges to integrated variance
- Convergence speed: $m^{1/6}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: yes
- Allows for time dependence noise: no
- Allows for for endogenous noise: no

3.7 Kernel-based Realized Variance

3.7.1 Assumptions

The microstructure noise.

1. The microstructure noise, $\epsilon_{t,i}$, has zero mean and is distributed random variable.
2. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.

3.7.2 Kernel Types

| Flat-Top Kernel | |
|-------------------------------|---|
| Bartlett Kernel | $k(x) = 1 - x$ |
| Epanichnikov Kernel | $k(x) = 1 - x^2$ |
| Second order Kernel | $k(x) = 1 - 2x + x^2$ |
| Non-Flat-Top Kernel | |
| Cubic Kernel | $k(x) = 1 - 3x^2 + 2x^3$ |
| Parzen Kernel | $k(x) = \begin{cases} 2(1-x)^3 & x > 0.5 \\ 1 - 6x^2 + 6x^3 & x < 0.5, \end{cases}$ |
| Tukey Hanning Kernel | $k(x) = \frac{(1 + \sin(\pi/2 - \pi x))}{2}$ |
| Tukey Hanning Modified Kernel | $k(x) = \frac{(1 - \sin(\pi/2 - \pi(1-x)^2))}{2}$ |
| Fifth Order Kernel | $k(x) = 1 - 10x^3 + 15x^4 - 6x^5$ |
| Sixth Order Kernel | $k(x) = 1 - 15x^4 + 24x^5 - 10x^6$ |
| Seventh Order Kernel | $k(x) = 1 - 21x^5 + 35x^6 - 15x^7$ |
| Eighth Order Kernel | $k(x) = 1 - 28x^6 + 48x^7 - 21x^8$ |

3.7.3 Estimator

Kernel-based Realized Variance is [?, Barndorff-Nielsen et al., 2006]:

$$\tilde{K}(X_\delta) = \gamma_0(X_\delta) + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) \tilde{\gamma}_h(X_\delta) \quad (3.28)$$

where $k()$ is kernel function and:

$$\tilde{\gamma}_h(X_\delta) = \gamma_h(X_\delta) + \gamma_{-h}(X_\delta) \quad (3.29)$$

$$\gamma_h(X_\delta) = \sum_{j=1}^n (X_{\delta j} - X_{\delta(j-1)})(X_{\delta(j-h)} - X_{\delta(j-h-1)}) \quad (3.30)$$

with $h = -H, \dots, -1, 0, 1, \dots, H$ and $n = \lceil 1/\delta \rceil$.

Bandwidth for all kernrel is:

$$H = cn^{2/3} \quad (3.31)$$

In this case we have the asymptotic distribution given:

$$n^{1/6} \{ \tilde{K}(X_\delta) - \int_0^t \sigma_u^2 du \} \xrightarrow{L_\S} MN \left[0, 4ck_{\bullet}^{0,0}t \int_0^t \sigma_u^4 du + 4\omega^4 c^{-2} \{k'(0)^2 + k'(1)^2\} \right] \quad (3.32)$$

where $\omega = (1, 1, k(\frac{1}{H}), \dots, k(\frac{H-1}{H}))^T$

$$k_{\bullet}^{0,0} = \int_0^1 k(x)^2 dx, \quad k_{\bullet}^{0,2} = \int_0^1 k(x)k''(x)dx \quad k_{\bullet}^{0,4} = \int_0^1 k(x)k''''(x)dx \quad (3.33)$$

The value of c which minimises the asymptotic variance is

$$c = d \frac{\omega^{4/3}}{(t \int_0^t \sigma_u^4 du)^{1/3}} \quad (3.34)$$

where

$$d = \left[\frac{2k^2(0)^2 + k'(1)^2}{k_{\bullet}^{0,0}} \right]^{1/2} \quad (3.35)$$

The lower bound for the asymptotic variance is $6dk_{\bullet}^{0,0}\omega^{4/3}(t \int_0^t \sigma_u^4 du)^{2/3}$

But for non-flat-top kernel we can review special case.

Bandwidth for non-flat-top kernel is:

$$H = cn^{1/2} \quad (3.36)$$

In this case the asymptotic distribution is given

$$n^{1/4} \{ \tilde{K}(X_\delta) - \int_0^t \sigma_u^2 du \} \xrightarrow{L_\S} MN \left[0, 4ck_{\bullet}^{0,0}t \int_0^t \sigma_u^4 du - 8c^{-1}k_{\bullet}^{0,2}\omega^2 \left(\int_0^t \sigma_u^2 du + \frac{\omega^2}{2} \right) + 4\omega^4 c^{-3} \{k'''(0) + k_{\bullet}^{0,4}\} \right] \quad (3.37)$$

For this class of kernels the value of \hat{c} which minimises the asymptotic variance is

$$c \approx \sqrt{\frac{1}{k_{\bullet}^{0,0}} \left\{ -k_{\bullet}^{0,2} + \sqrt{(k_{\bullet}^{0,2})^2 + 3k_{\bullet}^{0,0}f} \right\}} \quad (3.38)$$

3.7.4 Properties

Flat-Top Kernel

- Convergence speed: $m^{1/6}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: no
- Allows for time dependence noise: no
- Allows for endogenous noise: no

Non-Flat-Top Kernel

- Convergence speed: $m^{1/4}$ (m - number of observation)
- Unbiased: yes
- Consistent: yes
- Jump Robust: no
- Allows for time dependence noise: yes
- Allows for for endogenous noise: yes

3.8 IV Estimators Comparison

Figure 3.1: Table Estimators

4 Price Noise Variance

4.1 RV Noise Variance

Assume that the observed (log) price is contaminated by market microstructure noise u (or measurement error), i.e.:

$$Y_{i/m} = X_{i/m} + u_{i/m}, \quad i = 1, 2, \dots, m \quad (4.1)$$

where $X_{i/m}$ is the latent true, or so-called efficient, price that follows the semimartingale. In this case, the observed intraday return is given by:

$$r_i^{Y(m)} = r_i^{X(m)} + \epsilon_i^{(m)}, \quad i = 1, 2, \dots, m, \quad (4.2)$$

i.e. by the efficient intraday return $r_i^{X(m)} = X_{i/m} - X_{(i-1)/m}$ and the intraday noise increment $\epsilon_i^{(m)} = u_{i/m} - u_{(i-1)/m}$.

The noise variance $E\epsilon^2$ can be estimated consistently by normalized realized variance over the whole interval $[0, 1]$ for noisy data [?, Zhang et al., 2005]:

$$\hat{\omega}^2 = \frac{1}{2n} \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2 \quad (4.3)$$

4.1.1 Properties

- Convergence speed: $m^{1/2}$ (m - number of observation)
- Unbiased: no
- Consistent: yes
- Jump Robust: yes
- Allows for time dependence noise: no

4.2 Autocovariance Noise Variance

The noise variance can be estimated as the negative of the first order autocovariance of observed returns [?, Oomen, 2005]:

$$\hat{\omega}^2 = \max \left\{ 0, -\frac{1}{n-1} \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})(Y_{t_{i-1}} - Y_{t_{i-2}}) \right\} \quad (4.4)$$

4.2.1 Properties

- Convergence speed: $m^{1/2}$ (m - number of observation)
- Unbiased: no
- Consistent: yes
- Jump Robust: yes
- Allows for time dependence noise: no

5 Price Quarticity

5.1 Integrated Quarticity

The integrated quarticity is as described in [?, Pigorsch et al.]:

$$IQ = \int_{t-1}^t \sigma^4(s) ds. \quad (5.1)$$

5.2 Realized Quarticity

5.2.1 Assumptions

The microstructure noise.

1. The microstructure noise, $\epsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.
2. The noise is independent of the price process.
3. The variance of $\nu_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.

5.2.2 Estimator

The realized fourth-power variation or realized quarticity, defined as [?, Corsi et al., 2005]:

$$RQ_t = \frac{M}{3} \sum_{j=1}^M r_{t,j}^4 \xrightarrow{P} IQ \quad (5.2)$$

where M is sampling frequency.

5.2.3 Properties

- Convergence speed: $m^{1/n}$ (m - number of observation)
- Unbiased:
- Consistent:
- Jump Robust:
- Allows for time dependence noise:
- Allows for endogenous noise: