Financial Econometrics

Lecture 2

Advanced volatility modeling

Beyond GARCH

Roberto Renò
Università di Siena
March 12, 2011

Roberto Renò, 2011 ©
1 ARFIMA models

A motivating figure

Stationarity

- Strict stationarity:
  \[(X_1, X_2, \ldots, X_n) \overset{d}{=} (X_{1+k}, X_{2+k}, \ldots, X_{n+k})\]
  for every integers \(n > 1, k\).
- Weak/second-order/covariance stationarity:
  \[- Ex_t = \mu\]
  \[- EX_t - \mu^2 = \sigma^2 < +\infty\] (i.e. constant and independent of \(t\))
  \[- E(X_t - \mu)(X_{t+k} - \mu) = \gamma(|k|)\] (i.e. independent of \(t\) for each \(k\))
- Interpretation:
  - unconditional mean and variance are constant
  - mean reversion
  - shocks are transient
  - covariance between \(X_t\) and \(X_{t-k}\) tends to 0 as \(k \to \infty\)
Financial Econometrics, Advanced volatility modeling

White noise

- weak (uncorrelated)
  - $\mathbb{E}(\varepsilon_t) = 0$ $\forall t$
  - $V(\varepsilon_t) = \sigma^2$ $\forall t$
  - $\rho(\varepsilon_t, \varepsilon_s) = 0$ $\forall s \neq t$ where $\rho \equiv \frac{\gamma(t-s)}{\gamma(0)}$

- strong (independence)
  - $\varepsilon_t \sim I.I.D.(0, \sigma^2)$

- Gaussian (weak=strong)
  - $\varepsilon_t \sim N.I.D.(0, \sigma^2)$

Lag operator

- the Lag operator is defined as:
  $$LX_t \equiv X_{t-1}$$

- is a linear operator:
  $$L(\beta X_t) = \beta \cdot LX_t = \beta X_{t-1}$$
  $$L(X_t + Y_t) = LX_t + LY_t = X_{t-1} + Y_{t-1}$$

- and admits power exponent, for instance:
  $$L^2X_t = L(LX_t) = LX_{t-1} = X_{t-2}$$
  $$L^kX_t = X_{t-k}$$
  $$L^{-1}X_t = X_{t+1}$$

- Some examples:
  $$\Delta X_t = X_t - X_{t-1} = X_t - LX_t = (1 - L)X_t$$
  $$y_t = (\theta_1 + \theta_2 L)LX_t = (\theta_1 L + \theta_2 L^2)X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2}$$

- Expression like
  $$(\theta_0 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_n L^n)$$
  with possibly $n = \infty$, are called lag polynomial and are indicated as $\theta(L)$
Fractional differentiation

- The $k$-difference operator $(1 - L)^n$ with integer $n$ can be generalized to a fractional difference operator $(1 - L)^d$ with defined by the binomial expansion

$$(1 - L)^d = 1 - dL + d(d - 1)L^2/2! - d(d - 1)(d - 2)L^3/3! + ...$$

obtaining a fractionally integrated process of order $d$ i.e. $I(d)$.

- If $d = 0$ it becomes an ARMA$(p, q)$ process.
- If $d < 0.5$ the process is second-order stationary and admits an AR$(\infty)$ representation.
- The usefulness of a fractional filter $(1 - L)^d$ is that it produces hyperbolic decaying autocorrelations i.e. the so called long memory.
- The class of ARFIMA models has been introduced by Granger and Joyeux (1980) and Hosking (1981).

ARFIMA model

- We can write an ARFIMA$(p,d,q)$ model as:

$$\phi(L)Y_t = \theta(L)(1 - L)^{-d}\epsilon_t$$

where $\phi(L) = (1 + \phi_1L + \ldots + \phi_pL^p)$ and $\theta(L) = (1 + \theta_1L + \ldots + \theta_qL^q)$.

- the autocorrelation functions is proportional to

$$\rho(k) \approx ck^{2d-1}$$

- If $d < 0.5$, the process is second-order stationary
- If $d > -0.5$, the process is invertible
- If $0.5 \leq d < 1$, the process is non-stationary but mean-reverting.

2 The HAR model

Long Memory Models

- In std GARCH and SV models volatility shocks decay with exponential rate:

$$\rho_h \sim \gamma^h \text{ with } 0 < \gamma < 1$$

while an alternative possibility is an hyperbolic decay rate:

$$\rho_h \sim h^\gamma \text{ with } 0 < \gamma < 1$$

- The model is said to have long memory if the integral of the autocovariance function is infinite. One example is the ARFIMA class.
Financial Econometrics, Advanced volatility modeling

Fractional integration

- **Fractional Integration**: generalize the usual differencing of $I(1)$ series $y_t$ to get an $I(0)$ $\epsilon_t$ as:

\[(1 - L)^d y_t = \epsilon_t\]

$I(d)$ gives an infinite MA representation:

\[y_t = \sum_{k=0}^{\infty} a_k(d) \epsilon_{t-k}\]

with $a_k(d) = \frac{\Gamma(k+d)}{\Gamma(k+1) \Gamma(d)}$ which displays long memory with $\gamma = 2d - 1$.

Fractional Integration + ARMA $\rightarrow$ ARFIMA

Fractional Integration + GARCH $\rightarrow$ FIGARCH

Fractality & Multifractality

- Self-Similar, Fractal or Scaling Process:

\[(Y_{t1}, Y_{t2}, Y_{t3}, \ldots) \overset{d}{=} c^{-H} (Y_{c1}, Y_{c2}, Y_{c3}, \ldots)\]

which in terms of "generalized volatilities" implies:

\[E[|r(\Delta t)|^q] \sim \Delta t^{\mathcal{H}(q)}\]

- if $\mathcal{H}(q)$ is linear i.e. $\mathcal{H}(q) = Hq \rightarrow$ Unifractal or Monofractal process: - Brownian Motion ($H = 0.5$) - Fractionally Integrated processes ($H = d - 0.5$)

- if $\mathcal{H}(q)$ is nonlinear $\rightarrow$ Multifractal process: different scaling of different generalized volatility (Ding et al. 1993, Lux 1996, Andersen and Bollerslev 1997 + "econophysicists"). - Multifractal Model of Asset Returns (MMAR), Calvet and Fisher (2002):

\[X(t) \equiv B[\theta(t)] \quad \text{where } \theta(t) = \text{c.d.f. of multifractal measure}\]
Volatility cascades

- Are linked to the Heterogeneous Market Hypothesis of Muller et al. (1997)
- *Asymmetric propagation of volatility*: volatility over longer time intervals have stronger influence on those at shorter time intervals than conversely.
- Induced some authors to propose analogies with energy cascades of turbulent fluids, borrowing from Kolmogorov model of hydrodynamic turbulence: the so-called Stochastic Multiplicative Cascade (SMC)

Stylized Facts and Volatility Models

Standard volatility models are not able to reproduce all the stylized facts:

- GARCH and SV (one factor):
  - No long memory
  - No scaling
  - No volatility cascade

- Fractionally Integrated models:
  - No multi-scaling
  - No volatility cascade

Models Summary
### A different approach: Heterogeneity

- Additive processes with *heterogeneous components* can generate those stylized facts!

- It is an intuition of Clive Granger (1980): combination of only AR processes can display *apparent long memory*.

- If the aggregation level is not \( \gg \) than the lowest frequency component \( \Rightarrow \) asymptotically short memory models can be mistaken for long memory i.e. they are *empirically indistinguishable*.

### Model Ingredients

1. Heterogeneous Market Hypothesis (Müller et al. 1993): Main heterogeneity: difference in time horizons \( \Rightarrow \) agents perceive, react and cause different volatility components \( \tilde{\sigma}_t^{(i)} \)

2. *Volatility Cascade*: hierarchical process from Low to High Frequency.


\[ \downarrow \text{Cascade of Few Heterogeneous Realized Volatility Components} \]
we consider only 3 partial volatility components: daily $\tilde{\sigma}_t^{(d)}$, weekly $\tilde{\sigma}_t^{(w)}$, monthly $\tilde{\sigma}_t^{(m)}$

The HAR-RV Model of Corsi (2009)

- we work with logs to avoid negativity issues and get approximately Normal distributions.
- Consider the log RV$_t$ aggregated, as follow:

$$\log \text{RV}_t^{(n)} = \frac{1}{n} (\log \text{RV}_t + \ldots + \log \text{RV}_{t-n+1})$$

at the 3 different horizons: daily $d = 1$, weekly $w = 5$, monthly $m = 22$

Hence the model reads:

$$r_t = \tilde{\sigma}_t^{(d)} z_t$$
$$\log \tilde{\sigma}_{t+m}^{(m)} = c^{(m)} + \phi^{(m)} \log \text{RV}_t^{(m)} + \tilde{\omega}_{t+m}^{(m)}$$
$$\log \tilde{\sigma}_{t+w}^{(w)} = c^{(w)} + \phi^{(w)} \log \text{RV}_t^{(w)} + \gamma^{(w)} E_t[\log \tilde{\sigma}_{t+m}^{(m)}] + \tilde{\omega}_{t+w}^{(w)}$$
$$\log \tilde{\sigma}_{t+d}^{(d)} = c^{(d)} + \phi^{(d)} \log \text{RV}_t^{(d)} + \gamma^{(d)} E_t[\log \tilde{\sigma}_{t+w}^{(w)}] + \tilde{\omega}_{t+d}^{(d)}$$

A possible economic interpretation

each market components forms expectation for the next period volatility based on:

- the current RV experienced at the same time scale ("AR(1) part")
- the expectation for the next longer horizon partial volatility (Hierarchical part)

HAR-RV Model

- By straightforward recursive substitution

$$\log \sigma_{t+1}^{(d)} = c + \beta^{(d)} \log \text{RV}_t^{(d)}$$
$$+ \beta^{(w)} \log \text{RV}_t^{(w)}$$
$$+ \beta^{(m)} \log \text{RV}_t^{(m)} + \epsilon_t^{(d)}$$
Moreover, being:

$$\log \sigma_{t+1d} = \log RV_{t+1d} + \tilde{\epsilon}_{t+1d}$$

where $\tilde{\epsilon}_t$ is the measurement errors of $\log RV$, we get

$$\log RV_{t+1}^{(d)} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + \epsilon_{t+1d}^{(d)}$$

a simple AR-type model in the RV with the feature of considering volatilities realized over different interval sizes.

### 3 Jumps

**Jumps**

- The need for the introduction of discontinuous variations (jumps) come from a statistical basis: some movements of asset prices are totally incompatible with the observed volatility.
- Example 1: the black Friday, 1987
- Example 2: October 2009 (vs. the VIX)
- Presence of jumps is fundamental to explain the smile at low maturities.
Modelling the occurrence of rare events

- Assume that an event (the jump, the default of a bond) may occur, between time \( t \) and \( t + dt \), with probability \( \lambda dt \) for some constant \( \lambda \):

\[
P(t < \text{jump} < t + dt) = \lambda dt
\]

that is, the probability of jumping is infinitesimal constant

- Now consider a decay process starting at \( N_0 \) with constant infinitesimal decay. This means that:

\[
dN_t = N_t p(t < \text{decay} < t + dt) = N_t \lambda dt
\]

which implies:

\[
N_t = N_0 \exp(-\lambda t)
\]

- An infinitesimal constant decay probability implies an exponential distribution of the decay process.

Exponential random variables

**Definition 1** (Exponential random variable). A random variable \( \tau \) is called exponential if its pdf is given by:

\[
f(t) = \begin{cases} 
\lambda e^{-\lambda t} & t \geq 0 \\
0 & t < 0
\end{cases}
\]

where \( \lambda \) is a positive constant.

- The cdf is:

\[
F(t) = P(\tau \leq t) = 1 - e^{-\lambda t}
\]

which implies:

\[
P(\tau > t) = e^{-\lambda t}
\]

- The expected value is:

\[
E[\tau] = \frac{1}{\lambda}
\]
Poisson process

- Consider a collection $\tau_1, \ldots, \tau_n$ of iid exponential random variables, with parameter $\lambda$.
- Define the arrival times as:

$$S_n = \sum_{i=1}^{n} \tau_i$$

**Definition 2** (Poisson process). The Poisson process is the counting process:

$$N_t = n \text{ if } S_n \leq t < S_{n+1}$$

- The parameter $\lambda$ is called the intensity of the Poisson process, whose interpretation is: the probability of a jump between time $t$ and $t + dt$ is $\lambda dt$.
- The distribution of $S_n$ has a gamma density:

$$g_n(s) = \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s}$$

- The discrete distribution of the Poisson process is:

$$\mathcal{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

- It can be seen as the limit, as $p \to 0$, of a Binomial distribution with $k$ attempts (law of rare events).

A *Bernoulli random variable* can take the value 1 with probability $p$ and 0 with probability $1 - p$. A *binomial random variable* is the sum of $n$ independent Bernoulli. The distribution of a binomial random variable is given by:

$$P(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$
To see that it converges to the Poisson, define
\[ \lambda = p \cdot n \]

Now write:
\[
\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{k!} n(n-1) \cdots (n-k+1) \lambda^k \left( \frac{1-\lambda}{n} \right)^n \left( \frac{1-\lambda}{n} \right)^{-k}
\]

The last term converges, for a finite \( \lambda \) and \( p \to 0 \) (which is equivalent to \( n \to \infty \)), to:
\[
\frac{\lambda^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} \left( \frac{1-\lambda}{n} \right)^n \left( \frac{1-\lambda}{n} \right)^{-k} \to 1
\]

\[ \to \lambda^k e^{-\lambda} \frac{1}{k!} \].

### Mean and variance of a Poisson Process

Consider the increments of a Poisson process \( N_t - N_s \). We have:

- \( \mathbb{E}[N_t - N_s] = \lambda (s-t) \)
- \( \text{Var}[N_t - N_s] = \lambda (s-t) \)
- The mean and the variance are the same!

### Thumb rule

If the outcome of a counting process is \( N \), the standard error on that counting is between \( \sqrt{N}/2 \) and \( \sqrt{N} \).

**Exercise:** You are running for US presidency and according to a poll with \( N \) (hopefully independent) interviews, you have the 51.5% of votes. For what number \( N \) you will be 95% confident that you will be elected?

Computation of the mean and the variance of the Poisson process.

\[
\text{Mean} = \sum_{k=0}^{+\infty} \frac{k(\lambda t)^k}{k!} e^{-\lambda t}
\]
\[
= e^{-\lambda t} \sum_{k=1}^{+\infty} \frac{\lambda t(\lambda t)^{k-1}}{(k-1)!}
\]
(\text{using Taylor expansion of the exponential})
\[
= \lambda t e^{-\lambda t} e^{\lambda t} = \lambda t
\]
Second moment \[= \sum_{k=0}^{+\infty} k^2 \frac{\lambda t}{k!}e^{-\lambda t} \]
\[= e^{-\lambda t} \sum_{k=1}^{+\infty} \frac{(\lambda t)^k}{(k-1)!} \]
\[= e^{-\lambda t} \sum_{k=2}^{+\infty} \frac{(\lambda t)^{k-2}}{(k-2)!} + e^{-\lambda t} \sum_{k=1}^{+\infty} \frac{(\lambda t)(\lambda t)^{k-1}}{(k-1)!} \]
\[= (\text{using Taylor expansion of the exponential}) \]
\[= \lambda t + (\lambda t)^2 \]

Solution of the exercise. Assume that the true probability of being elected is \( p = 50\% \), and that the polls are independent: this is the null hypothesis. I want to see if the value \( \hat{p} = 51.5\% \) rejects it. According to this assumption, the probability distribution is given by a binomial distribution with \( N \) attempts. Since \( N \) is large, we approximate the binomial distribution with the Normal distribution by the central limit theorem (it is the sum of independent Bernoulli). The variance is then \( p(1-p)/N = 0.25/N \). We are confident that \( p > 0.5 \) at 95% confidence band (one sided) at 1.6449 standard deviations. Thus, we reject the null if

\[0.015 > 1.6449 \sqrt{\frac{0.25}{N}} \rightarrow N > 3006\]

and in this case I will be confident, at 95% confidence level, to win.

Now, we repeat the same reasoning using the thumb rule! Assume that you interviewed \( N = 3000 \) persons. You immediately compute that the error is \( \sqrt{N} \approx 50 \). Thus, if \( 3000 \cdot 51.5\% = 1545 \) persons declared the intentions to vote for you, you know that the margin of 45 is not enough to be safe. In this case, the standard error is very close to \( \sqrt{N}/2 \) since \( p = 0.5 \), which is not a “rare” event. It is much closer to \( \sqrt{N} \) if \( p \) is close to zero. As you can see, it can immediately give you a “sense” of the statistical significance of the numbers.

Wiener process and Poisson process
Wiener process:
- \( W_0 = 0 \)
- If \( t < s < u \), \( W_u - W_s \) is independent from \( W_s - W_t \).
- \( W_s - W_t \sim \mathcal{N}(0, s - t) \)
- The trajectory \( W_t \) is continuous
- \( W_t \) is a martingale

Poisson process:
- \( N_0 = 0 \)
- If \( t < s < u \), \( N_u - N_s \) is independent from \( N_s - N_t \).
- \( N_s - N_t \sim \frac{(\lambda(s-t))^k}{k!}e^{-\lambda(s-t)} \)
- The trajectory \( N_t \) has purely discontinuous variations
- \( N_t \) is not a martingale.

Compensated Poisson process
- The non-martingality of \( N_t \) is not an issue.
- Indeed, define the process:
  \[
  M_t = N_t - \lambda t
  \]
- It is easy to verify that \( M_t \) is a martingale.

Compound Poisson process
- We can easily allow jumps to have arbitrary sizes.
- Let \( N_t \) be a Poisson process and \( Y_1, \ldots, Y_n \) iid random variables.

**Definition 3** (Compound Poisson process). A Compound Poisson process is defined as:

\[
Q_t = \sum_{i=1}^{N_t} Y_i
\]
• The mean of the increments of a compound Poisson process is:
\[ E[Q_s - Q_t] = E[Y_i] \lambda (s - t) \]

• It is easy to verify that, if \( s < t < u \), the increments \( Q_u - Q_s \) and \( Q_s - Q_t \) are independent.

• Also in this case, we can define a compensated version which is a martingale:
\[ J_t = Q_t - E[Y_i] \lambda t \]

• We could also allow for a stochastic intensity. In this case, we should replace in all formulas \( \lambda (s - t) \) with \( E[\int_t^s \lambda_s ds] \)

2.28 Power variation in the general case
In the more general case, we have:
\[
p - \lim_{\delta \to 0} \mu^{-1} PV_\delta(X)_t^{[\gamma]} = \begin{cases} 
\int_0^t \sigma_s^\gamma ds & \text{if } \gamma < 2 \\
[X]_t & \text{if } \gamma = 2 \\
+\infty & \text{if } \gamma > 2
\end{cases}
\]

However, it is impossible to get an estimate of the integrated volatility using power variation.

2.29 Bipower variation
We define realized bipower variation of order \([\gamma_1, \gamma_2]\) as:
\[
BPV_\delta(X)_t^{[\gamma_1, \gamma_2]} = \delta^{1 - \frac{1}{2}(\gamma_1 + \gamma_2)} \sum_{j=2}^{[t/\delta]} |\Delta_{j-1}X|^{\gamma_1} \cdot |\Delta_jX|^{\gamma_2}
\]

In the case of no jumps we have:
\[
(J = 0) \quad p - \lim_{\delta \to 0} BPV_\delta(X)_t^{[\gamma, \gamma]} = \mu_1^2 \int_0^t \sigma_s^{2\gamma} ds
\]

2.30 Bipower variation and continuous quadratic variation
• In the more general case, we have:
\[
p - \lim_{\delta \to 0} \mu_1^{-1} \mu_2^{-1} BPV_\delta(X)_t^{[\gamma_1, \gamma_2]} = \begin{cases} 
\int_0^t \sigma_s^{\gamma_1 + \gamma_2} ds & \text{if } \max(\gamma_1, \gamma_2) < 2 \\
\text{something finite} & \text{if } \max(\gamma_1, \gamma_2) = 2 \\
+\infty & \text{if } \max(\gamma_1, \gamma_2) > 2
\end{cases}
\]
This result means that we can use bipower variation to estimate the integrated variance even in the presence of jumps: that is, choosing $\gamma_1 < 2$ and $\gamma_2 = 2 - \gamma_1$.

**Standard bipower variation**

The $[1, 1]$-order bipower variation, when it exists, is defined as:

$$BPV(X)^{[1,1]}_t = p - \lim_{\delta \to 0} \frac{[t/\delta]}{\delta} \sum_{j=2}^{[t/\delta]} |\Delta_{j-1}X| \cdot |\Delta_jX|$$

In the case $\mu = 0$ and $\sigma$ independent from $W_t$, we have:

$$BPV(X)^{[1,1]}_t = \mu_1 \int_0^t \sigma^2 s \, ds = \mu_1^2 [X^c]_t$$

where

$$\mu_1 = \frac{\sqrt{2}}{\sqrt{\pi}}$$

This result can be extended to the case $\mu \neq 0$.

**Realized volatility and bipower variation**


[Graphs showing realized variance and bipower variation over time]

**Bipower variation on simulated data**

Bipower variation and data

Confidence intervals for bipower variation

Confidence intervals can be obtained using

\[ \delta^{-\frac{1}{2}} \left( BPV_\delta(X)_t^{[1,1]} - BPV(X)_t^{[1,1]} \right) \xrightarrow{L} \mu_1^2 \sqrt{2 + \vartheta} \int_0^t \sigma_s^2 dW_s' \]

where

\[ \vartheta = \frac{\pi^2}{4} + \pi - 5. \]
Testing for jumps

- Bipower variation can be used to test for jumps.
- Linear jump statistic:
  \[
  \frac{\delta^{-1/2} \left( \mu_1^{-2} BPV_\delta(X)_{t}^{[1,1]} - RV_\delta(X)_{t} \right)}{\sqrt{\int_0^t \sigma_s^4 ds}} \to_{L} \mathcal{N}(0, \vartheta)
  \]

- Ratio jump statistic:
  \[
  \frac{\delta^{-1/2} \left( \mu_1^{-2} BPV_\delta(X)_{t}^{[1,1]} \right)}{RV_\delta(X)_{t}^{[1,1]}} - 1
  \frac{1}{\sqrt{\left( \int_0^t \sigma_s^2 ds \right)^2}} \to_{L} \mathcal{N}(0, \vartheta)
  \]

Multipower variation

- We need an estimate of \( \int_0^t \sigma_s^4 ds \) which is robust to jumps!
- We define the \textit{realized multipower variation} as:
  \[
  MPV_{\delta}^{\gamma_1,\ldots,\gamma_N}(X)_{t} = \delta^{1-\frac{1}{2}(\gamma_1+\ldots+\gamma_N)} \sum_{j=N}^{[t/\delta]} \prod_{k=0}^{N-1} |\Delta_{j-k}X|
  \]

- when \( \max(\gamma_1, \ldots, \gamma_N) < 2 \), we have:
  \[
  p - \lim_{\delta \to 0} MPV_{\delta}^{\gamma_1,\ldots,\gamma_N}(X)_{t} = \left( \prod_{k=1}^{N} \mu_{\gamma_k} \right) \int_0^t \sigma_s^{\gamma_1+\ldots+\gamma_N} ds
  \]

Examples of multipower variation

- \textit{quadpower variation}:
  \[
  MPV_{\delta}^{2,2,2,2}(X)_{t} = \frac{1}{\delta} \sum_{j=4}^{[t/\delta]} |\Delta_{j-3}X| \cdot |\Delta_{j-2}X| \cdot |\Delta_{j-1}X| \cdot |\Delta_jX|
  \to \mu_4 \int_0^t \sigma_s^4 ds
  \]

- \textit{tripower quarticity}:
  \[
  MPV_{\delta}^{1,2,\frac{3}{2},\frac{3}{2}}(X)_{t} = \frac{1}{\delta} \sum_{j=3}^{[t/\delta]} |\Delta_{j-2}X|^{\frac{3}{2}} \cdot |\Delta_{j-1}X|^{\frac{3}{2}} \cdot |\Delta_jX|^{\frac{3}{2}}
  \to \mu_3 \int_0^t \sigma_s^4 ds
  \]
Detecting jumps

Our idea to disentangle diffusion from jumps is based on the **modulus of continuity** of the Brownian motion:

$$r(\delta) = \sqrt{2\delta \log \frac{1}{\delta}}$$

which has the following property, as established by Lévy:

$$\mathcal{P} \left[ \limsup_{\delta \to 0} \max_{|t-s| \leq \delta} \frac{|W(t) - W(s)|}{r(\delta)} = 1 \right] = 1$$

It measures the *speed* at which the Brownian motion shrinks to zero.

**The intuition**

When $\delta \to 0$, diffusive variations go to zero, while jumps do not.

Moreover, we know the rate at which the diffusive variations shrink to zero: the modulus of continuity.

Thus, we can identify the jumps as those variations which are larger than a suitable threshold $\vartheta(\delta)$ which goes to zero, as $\delta \to 0$, slower than $r(\delta)$. 
The theorem (Mancini, 2009)

Suppose
\[ X = Y + J \]
where \( Y \) is a Brownian martingale plus drift and \( J \) is a jump process with counting process \( N \) with \( E[N_T] < \infty \) and time horizon \( T < \infty \).

If \( \vartheta(\delta) \) is a real deterministic function such that
\[
\lim_{\delta \to 0} \vartheta(\delta) = 0 \quad \text{and} \quad \lim_{\delta \to 0} \frac{\delta \log \frac{1}{\delta}}{\vartheta(\delta)} = 0
\]
than for \( P \)-almost all \( \omega \), \( \exists \tilde{\delta}(\omega) \) such that \( \forall \delta < \tilde{\delta}(\omega) \) we have
\[
\forall i = 1, \ldots, n, \quad I_{\{\Delta N = 0\}}(\omega) = I_{\{|\Delta X|^2 \leq \vartheta(\delta)\}}(\omega).
\]

Threshold realized variance

We define the threshold realized variance as:
\[
TRV_{\delta}(X)_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} (\Delta_j X)^2 I_{\{|\Delta_j X|^2 \leq \vartheta(\delta)\}}
\]

Using the above result, we can prove that, for every threshold \( \vartheta(\cdot) \) vanishing slower than the modulus of continuity of the Brownian motion, \( TRV_{\delta}(X)_t \) is a consistent estimator, as \( \delta \to 0 \), of the integrated volatility, and this result is robust under infinite activity jumps.

Estimating \( \int \sigma_s^2 ds \): Threshold Bipower Variation

Bipower variation can be heavily biased in small samples, potentially distorting the analysis of realized jumps.

Inspired by the results of Mancini (2009), Corsi, Pirino and Renò (2010) proposed the **Threshold Bipower Variation**:

\[
TBPV_t = \frac{\pi}{2n - 1 - nJ} \sum_{j=0}^{n-2} |\Delta_t, jX| \cdot |\Delta_t, j+1X| I_{\{|\Delta_t, jX|^2 \leq \vartheta_{j-1}\}} I_{\{|\Delta_t, j+1X|^2 \leq \vartheta_j\}}
\]

where \( \vartheta_t \) is a threshold function proportional to the local variance (for an estimator of the same spirit, see Andersen, Dobrev and Schaumburg, 2009).
It combines threshold estimation for large jumps and bipower variation for smaller jumps (*twin blade*). It is expected not to be too sensitive to the threshold.

### References


