Jump Returns and the Truncated Variance

1 Setup and Review

We assume that log-prices evolve according to a jump-diffusion process:

$$dX_t = \sqrt{c_t} dW_t + dJ_t$$

where c_t is the variance process and J_t is a jump process (of finite activity). Notice that the process above does not have a drift term. This is because the drift is negligible at small time periods.

The process is sampled at equi-distant intervals:

$$X_{i\Delta_n} \text{ for } i = 0, 1, 2, \ldots, nT$$

where T is the number of days for which the process is observed, and n counts the number of observations within a day.

Remember that the process X has two parts: one that is continuous (diffusion), and one that is discontinuous (jump). Indeed:

$$\begin{aligned} X_t^{'} &\equiv \sum_{0\leq s\leq t} \Delta X_t \\ X_t &= \underbrace{(X_t - X_t^{'})}_{\text{continuous}} + \underbrace{X_t^{'}}_{\text{jumps}} \\ &= X_t^c + X_t^d \end{aligned}$$

Let's assume $T = 1$. Then the daily integrated variance is given by:

$$IV = \int_0^1 c_s ds$$

Given the geometric returns $r_i = \Delta_i^n X$ for $i = 1, 2, \ldots, n$, we studied two estimators for the integrated variance. The realized volatility (RV) and the bipower variance (BV):

$$\begin{aligned} RV &\equiv \sum_{i=1}^{n} r_i^2 \\ BV &\equiv \frac{\pi}{2} \sum_{i=2}^{n} |r_i r_{i-1}| \end{aligned}$$

RV and BV are estimators for IV, but RV is affected by the presence of jumps, while BV is jump robust. Next, we will discuss the theory for separating moves in the price (returns) that come from the continuous part of the process (X^c) and moves that come from the discontinuous part of the process (X^d).

2 Separating Diffusive Returns from Jump Returns

Let's denote the return at the i-th interval of day t by $r_{t,i}$:

$$\begin{aligned} r_{t,i} &\equiv X_{(i+(t-1)n)\Delta_n} - X_{(i-1+(t-1)n)\Delta_n} \\ &= \Delta^n_{i+(t-1)n} X \end{aligned}$$

for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, n$. To simplify the notation, let $t_i \equiv (i + (t-1)n)\Delta_n$. Then, the return on any interval is given by

$$\begin{aligned} r_{t,i} &= X_{t_i} - X_{t_{i-1}} \\ &= X_{t_i} - X_{t_i - \Delta_n} \end{aligned}$$

The idea is to classify each of the moves (returns) as continuous (diffusive) or discontinuous (jump):

$$r_{t,i} = r^c_{t,i} + r^d_{t,i}$$

Using the jump-diffusion process, we can write:

$$X_{t_i} = X_{t_i - \Delta_n} + \int_{t_i - \Delta_n}^{t_i} \sqrt{c_s} dW_s + \sum_{t_i - \Delta_n < s \le t_i} \Delta J_s$$

If there were no jumps:

$$\begin{aligned} r_{t,i} &= X_{t_i} - X_{t_i - \Delta_n} \\ &= \int_{t_i - \Delta_n}^{t_i} \sqrt{c_s} dW_s \\ &\approx \underbrace{\sqrt{\int_{t_i - \Delta_n}^{t_i} c_s ds}}_{\equiv \sigma_{t,i}} Z_{t,i} \\ &\overset{d}{\sim} \mathcal{N}\left(0, \sigma^2_{t,i}\right) \end{aligned}$$

If we could observe (we do not) the path of the variance process c_t on the interval $[t_i - \Delta_n, t_i]$, then we could compute $\sigma_{t,i}$. Now, the return over such a short time interval is approximately normal:

$$r_{t,i} \approx \sigma_{t,i} Z_{t,i}$$

The probability that $r_{t,i}$ is between 3 standard deviations of the mean is about 99.73%. The probability that $r_{t,i}$ is between 4 standard deviations of the mean is about 99.99%. That is, if there are no jumps in the time interval $[t_i - \Delta_n, t_i]$, then we expect most of the returns to fall within 4 standard deviations of the mean. However, we know jumps can occur at any time interval. If a jump occurs, then the magnitude of $r_{t,i}$ will be dominated by the jump and will far exceed 4 standard deviations of the mean.

This analysis motivates the following rule:

$$\begin{aligned} |r_{t,i}| &\le \alpha_n \sigma_{t,i} : \text{return is diffusive} \\ |r_{t,i}| &> \alpha_n \sigma_{t,i} : \text{return is jump} \end{aligned}$$

Where α_n is some threshold that decreases as $\Delta_n \to 0$. The use of a threshold to separate continuous returns from jump returns is due to Cecilia Mancini (2001) and C. Mancini (2009), and was refined in Jacod and Protter (2012). The formal proof that the threshold correctly classifies diffusive moves from jump moves was developed in Li, Todorov, and Tauchen (2017).

The current practice is the following:

$$r^c_{t,i} = r_{t,i} 1_{\left\{|r_{t,i}| \leq \alpha \Delta_n^{0.49} \sqrt{\tau_i BV_t}\right\}}$$
$$r^d_{t,i} = r_{t,i} 1_{\left\{|r_{t,i}| > \alpha \Delta_n^{0.49} \sqrt{\tau_i BV_t}\right\}}$$

where $\alpha \in [3.5, 4.5]$.

Let's understand the substitution of $\sigma_{t,i}$ by the term $\Delta_n^{0.49}\sqrt{\tau_i BV_t}$. First, we do not observe the path of c_t, so we cannot directly compute $\sigma_{t,i}$. However, we can estimate the total variation of a day (IV_t) using the bipower variance estimator (BV_t). But using the total variation leads to a problem: on a small interval the variation is not given by IV_t but only a fraction of it. This can be fixed by appropriately redistributing the total variation. Thus, instead of writing $r_{t,i} \approx \sigma_{t,i} Z_{t,i}$ we have:

$$\begin{aligned} r_{t,i} &\approx \sqrt{\Delta_n IV_t} Z_{t,i} \\ &= \Delta_n^{0.50} \sqrt{IV_t} Z_{t,i} \end{aligned}$$

When redistributing the total variance there is another problem that needs to be addressed: the variance process exhibits an intraday pattern. This intraday pattern refers to the fact that in the mornings the variance is higher than during lunch time, and after lunch time the variance increases again. To adjust for this effect, we scale IV_t depending on the time of the day, leading to:

$$r_{t,i} \approx \Delta_n^{0.50} \sqrt{\tau_i IV_t} Z_{t,i}$$

We call τ_i the diurnal pattern or the time-of-day factor.

The time-of-day factor is used to account for the intraday pattern of the volatility. To compute τ_i we first estimate the average (across days) bipower factors at each time interval:

$$b_i \equiv \frac{1}{T} \sum_{t=1}^{T} |r_{t,i} r_{t,i-1}| \text{ for } i = 2, 3, \ldots, n$$
$$b_1 \equiv b_2$$

Then, the time-of-day factor is defined as:

$$\tau_i \equiv \frac{b_i}{\frac{1}{n}\sum_{j=1}^{n} b_j}$$

That is, the time-of-day factor is just a re-scaled version of b_i so that the mean of τ_i over a day is 1.

Now, the local return ($r_{t,i}$) is approximately normally distributed:

$$r_{t,i} \overset{d}{\sim} \mathcal{N}(0, \Delta_n \tau_i IV_t)$$

To correctly separate the diffusive from the jump returns, we need a threshold that is slightly bigger than $\Delta_n^{0.50}\sqrt{\tau_i IV_t}$. Alternatively, we need a threshold that decreases slower than $\Delta_n^{0.50}\sqrt{\tau_i IV_t}$ when $\Delta_n \to 0$ to allow all diffusive moves to get through the threshold and exclude the jump moves. To achieve that we inflate the cutoff by:

$$\begin{aligned}\Delta_n^{-0.01} &= \left(\frac{1}{n}\right)^{-0.01} \\ &= n^{0.01} \\ &> 1\end{aligned}$$

In summary, we separate returns using the cuttoff:

$$\text{cutoff}_{t,i} \equiv \alpha_n \Delta_n^{0.49}\sqrt{\tau_i BV_t}$$

The exponent for Δ_n does not necessarily have to be 0.49, but can actually be any number close to 0.50. The literature often writes this threshold as:

$$\text{cutoff}_{t,i} \equiv \alpha_n \Delta_n^{\varpi}\sqrt{\tau_i BV_t} \text{ for } 0 < a < \varpi < 0.50$$

where a is a lower bound determined by some other conditions (unimportant here).

3 Truncated Variance

Now that we can separate returns coming from the diffusive part of the model from the returns coming from the jump part, we can study a new estimator for the integrated variance.

The truncated variance is an estimator for the integrated variance that is robust to jumps. It is defined as:

$$\begin{aligned}TV_t &\equiv \sum_{i=1}^{n}(\Delta_{i+(t-1)n}^n X)^2 1_{\left\{\left|\Delta_{i+(t-1)n}^n X\right| \leq \text{cutoff}_{t,i}\right\}} \\ &= \sum_{i=1}^{n}(r_{t,i}^c)^2\end{aligned}$$

This estimator throws away moves with big jumps, retaining only the moves that are diffusive to estimate the variance.

Under regularity conditions, it is possible to show that:

$$\Delta_n^{-1/2}(TV_t - IV_t) \xrightarrow{d} \mathcal{N}\left(0, 2\int_{t-1}^{t} c_s^2 ds\right)$$

Next, let's compare the truncated variance to the previous estimators.

4 Comparison of IV Estimators

The table below compares the asymptotic distribution of the IV estimators studied up to now:

Estimator	Definition	Rate of Convergence	Asymptotic Distribution	Robust to Jumps
RV_t	$\sum_{i=1}^{n} r_{t,i}^2$	$\Delta_n^{-0.50}$	$\mathcal{N}\left(IV_t, 2\int_0^1 c_s^2 ds\right)$	No (bias: $\sum_s \Delta J_s^2$)
BV_t	$\sum_{i=2}^{n} \lvert r_{t,i} r_{t,i-1} \rvert$	$\Delta_n^{-0.50}$	$\mathcal{N}\left(IV_t, 2.61\int_0^1 c_s^2 ds\right)$	Yes
TV_t	$\sum_{i=1}^{n} (r_{t,i}^c)^2$	$\Delta_n^{-0.50}$	$\mathcal{N}\left(IV_t, 2\int_0^1 c_s^2 ds\right)$	Yes

Notice that RV is the only estimator that is affected by jumps. Under the presence of jumps RV estimates the integrated variance, but also adds in the squared jump returns. However, if there are no jumps, then the asymptotic variance of RV is smaller than that of BV. The TV estimator converges to the integrated variance whether there are jumps or not. Its asymptotic variance is smaller than that of BV, making it a more efficient jump robust estimator for the integrated variance.

References

Jacod, Jean and Philip Protter (2012). *Discretization of Processes*. Springer-Verlag. URL: `http://dx.doi.org/10.1007/978-3-642-24127-7`.

Li, Jia, Viktor Todorov, and George Tauchen (2017). “Jump Regressions”. In: **Econometrica** 85, pp. 173–195.

Mancini, C. (2009). “Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps”. In: *Scandinavian Journal of Statistics* 36, pp. 270–296.

Mancini, Cecilia (2001). “Disentangling the jumps of the diffusion in a geometric jumping Brownian motion”. In: *Giornale della Istituto Italiano degli Attuari* 64.19-47, p. 44.