Inference for the Integrated Variance

Under a Jump-Diffusion model, we can write the price of an asset as the combination of a continuous term and a discontinuous term:

$$X_t = X_t^c + X_t^d$$

Using the bipower variance estimator, we learned how to separate prices moves (returns) coming from the continuous part, from those returns coming from the discontinuous part. We wrote the returns of an asset at day t and time i as:

$$r_{t,i} = r_{t,i}^c + r_{t,i}^d$$

We can use the continuous returns to estimate the integrated variance. The estimate is known as the truncated variance estimator and it is jump-robust:

$$TV_t = \sum_{i=1}^{n} (r_{t,i}^c)^2$$

The theory indicates that for large n:

$$\Delta_n^{0.50}(TV_t - IV_t) \overset{d}{\sim} \mathcal{N}\left(0, 2\int_{t-1}^{t} c_s^2 ds\right)$$

In order to do inference for the integrated variance, we need an estimator for its asymptotic variance.

1 QIV: Quartic Integrated Variation

Define:

$$QIV_t \equiv \int_{t-1}^{t} c_s^2 ds$$

QIV stands for Quartic Integrated Variation. It can be estimated by:

$$\widehat{QIV}_t \equiv (3\Delta_n)^{-1} \sum_{i=1}^{n} (r_{t,i}^c)^4$$

Let's verify the convergence and understand why the term $(3\Delta_n)^{-1}$ shows up. Remember that:

$$r_{t,i}^c = \sqrt{\int_{t_i-\Delta_n}^{t_i} c_s ds}\, Z_{t,i}$$

To simplify the argument, consider that c_s is constant:

$$r^c_{t,i} = \sqrt{\Delta_n c} Z_{t,i}$$

Then, let's study the convergence of just the sum of the continuous returns to the 4th power:

$$\begin{aligned}\sum_{i=1}^{n} (r^c_{t,i})^4 &= \sum_{i=1}^{n} \Delta_n^2 c^2 Z_{t,i}^4 \\ &= (c^2 \Delta_n) \left(\Delta_n \sum_{i=1}^{n} Z_{t,i}^4 \right)\end{aligned}$$

We can use the law of large numbers on the right most term:

$$\Delta_n \sum_{i=1}^{n} Z_{t,i}^4 \to \mathbb{E}\left[Z^4\right] = 3$$

The left most term converges to 0:

$$c^2 \Delta_n \to 0$$

Putting both together we would get:

$$\sum_{i=1}^{n} (r^c_{t,i})^4 \to 0$$

But what we want to estimate is c^2. In order to fix that, we multiply by the term $(3\Delta_n)^{-1}$:

$$(3\Delta_n)^{-1} \sum_{i=1}^{n} (r^c_{t,i})^4 = \frac{c^2}{3} \left(\Delta_n \sum_{i=1}^{n} Z_{t,i}^4 \right) \to c^2$$

These adjustments ensure that:

$$\widehat{QIV}_t \to \int_{t-1}^{t} c_s^2 ds$$

2 Confidence Interval for the Integrated Variance

Now, we discuss how to compute confidence intervals for the integrated variance, given the truncated variance estimator and the quartic integrated variance estimator.

The first approach is using the asymptotic theory, which tells us that TV is asymptotically normal and centered around IV. This approach is computationally faster, but requires the user to have derived the asymptotic distribution result.

The second approach is based on the bootstrap method. The idea is to obtain samples of returns from the same distribution, use these new samples to compute the truncated variance estimator multiple times, which gives us the asymptotic distribution of TV by brute force. While this approach does not rely on having derived the asymptotic results, it is computationally intensive.

2.1 Using the Asymptotic Distribution

The estimator $\widehat{QIV}_t$ can be used to construct confidence intervals for IV_t. The asymptotic theory implies that for large n we have:

$$\Delta_n^{-\frac{1}{2}} \frac{TV_t - IV_t}{\sqrt{2\widehat{QIV}_t}} \overset{d}{\sim} \mathcal{N}(0,1)$$

We can use this result to construct a confidence interval for IV_t:

$$IV_t \in \left[TV_t + q_Z(\alpha/2)\sqrt{\Delta_n 2\widehat{QIV}_t}, TV_t + q_Z(1-\alpha/2)\sqrt{\Delta_n 2\widehat{QIV}_t}\right]$$

with probability $1-\alpha$, where $q_Z(\alpha/2)$ is the $\alpha/2$ quantile of the standard normal (e.g.: $q_Z(5\%/2) = -1.96$).

Notice that we can simplify the confidence interval by opening the expression inside the square-root:

$$\begin{aligned} \sqrt{\Delta_n 2\widehat{QIV}_t} &= \sqrt{2\Delta_n \frac{1}{3\Delta_n}\sum_{i=1}^{n}(r_{t,i}^c)^4} \\ &= \sqrt{\frac{2}{3}\sum_{i=1}^{n}(r_{t,i}^c)^4} \end{aligned}$$

Using this simplification we can write the confidence interval as:

$$CI(IV_t,\alpha) \equiv \left[TV_t + q_Z(\alpha/2)\sqrt{\frac{2}{3}\sum_{i=1}^{n}(r_{t,i}^c)^4},\ TV_t + q_Z(1-\alpha/2)\sqrt{\frac{2}{3}\sum_{i=1}^{n}(r_{t,i}^c)^4}\right]$$

2.2 Using the Bootstrap Method

The bootstrap is an alternative way to create confidence intervals for a statistic of interest. We will discuss a bootstrapping scheme based on Hounyo (2013),Gonçalves and Meddahi (2009) and Dovonon et al. (2014).

The idea of the bootstrap is to use the data that we already have to obtain "new" samples from the same distribution. The idea is to draw random samples (with replacement) from the observed data. That is, we resample with replacement the original sample, generating a "new" sample from the same distribution. Then we use these new samples to compute the estimator of interest multiple times. We then use the fluctuations of the estimators to construct the confidence interval.

Let's fix $T=1$, then we have n observations of continuous returns in a day:

$$r_1^c, r_2^c, \dots, r_n^c$$

We want to draw new samples from these returns. However, the new samples must respect the heteroskedasticity (time-varying variance) over the day. To do so, we will partition the day into non-overlapping segments of length k_n, so that there are M divisions in a day.

$$\underbrace{r_1^c, r_2^c, \dots, r_{k_n}^c}_{\text{1st}}, \underbrace{r_{k_n+1}^c, r_{k_n+2}^c, \dots, r_{2k_n}^c}_{\text{2nd}}, \dots, \underbrace{r_{(M-1)k_n+1}^c, r_{(M-1)k_n+2}^c, \dots, r_{Mk_n}^c}_{\text{Mth}}$$

We must have $Mk_n \leq n$. In practice we will divide the day into an integer number of segments so that $Mk_n = n$.

Now, to generate a new random sample we draw k_n returns with replacement from each of the M divisions:

$$\underbrace{\tilde{r}_1^c, \tilde{r}_2^c, \ldots, \tilde{r}_{k_n}^c}_{\text{1st}}, \underbrace{\tilde{r}_{k_n+1}^c, \tilde{r}_{k_n+2}^c, \ldots, \tilde{r}_{2k_n}^c}_{\text{2nd}}, \ldots, \underbrace{\tilde{r}_{(M-1)k_n+1}^c, \tilde{r}_{(M-1)k_n+2}^c, \ldots, \tilde{r}_{Mk_n}^c}_{\text{Mth}}$$

Use this new random sample to compute the truncated variance estimator:

$$\widetilde{TV} = \sum_{i=1}^{Mk_n} (\tilde{r}_i^c)^2$$

Store $\widetilde{TV}$ and repeat the process: draw a new random sample from the original data and compute a new $\widetilde{TV}$ estimator. After repeating this a sufficient number of times (say 10000), we have a collection of estimators:

$$BS(TV) \equiv \left\{\widetilde{TV}_1, \widetilde{TV}_2, \ldots, \widetilde{TV}_{10000}\right\}$$

To construct the confidence interval for IV we compute the quantiles of the set above. For example, if we compute the 2.5% and 97.5% quantiles ($q_{BS(TV)}$) of the set, then the confidence interval is given by:

$$CI(TV, 5\%) \equiv \left[q_{BS(TV)}(2.5\%), q_{BS(TV)}(97.5\%)\right]$$

The bootstrap is useful because it avoids having to compute the asymptotic variance of the estimator, which can be quite hard in some cases. However, it does require a lot of computational power, since you need to draw a big number of random samples and re-compute the estimator every time, and then repeat the process for every day of your sample. It is important to know that the confidence intervals computed via bootstrap are no better or worse than the ones from the asymptotic theory. Additionally, there are stringent conditions to guarantee that the bootstrap confidence interval is valid. In this context, arguments from Li, Todorov, and Tauchen (2015) justify the use of the bootstrap.

3 Confidence Interval for the Annualized Integrated Variance

The estimates for the integrated variance are often converted to annualized measures. To obtain the confidence intervals for the annualized integrated variance we need to make use of the Delta method. The Delta Method is a useful result that allows us to quickly change the units of measurement of an estimator and adapt its limiting distribution.

Suppose we have an estimator $\hat{\theta}_n$ of some statistic we care about θ. Additionally, suppose we know its limiting distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

Given a function $g : \mathbb{R} \to \mathbb{R}$ that is differentiable at θ, then:

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, g'(\theta)^2\sigma^2\right)$$

That is, by rescaling the estimator with a smooth function we only change the variance of the asymptotic distribution. This result is known as the Delta method.

We can use the Delta Method to obtain the asymptotic distribution of annualized versions of the estimators we have studied so far. For example, if there are no jumps:

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

Notice that:

$$\begin{aligned}\Delta_n^{-\frac{1}{2}} &= \left(\frac{1}{n}\right)^{-\frac{1}{2}} \\ &= \sqrt{n}\end{aligned}$$

Define the functions:

$$g : x \mapsto \sqrt{252 \times x}$$

Then the asymptotic distribution of the annualized RV is:

$$\Delta_n^{-\frac{1}{2}}(\sqrt{252RV_t} - \sqrt{252IV_t}) \xrightarrow{d} \mathcal{N}\left(0, g'(IV_t)^2 \cdot 2\int_{t-1}^{t} c_s^2 ds\right)$$

Using the notation from last lecture:

$$\Delta_n^{-\frac{1}{2}}(\sqrt{252RV_t} - \sqrt{252IV_t}) \xrightarrow{d} \mathcal{N}\left(0, g'(IV_t)^2 \cdot 2QIV_t\right)$$

Plugging in estimators for the new asymptotic variance we get:

$$\Delta_n^{-\frac{1}{2}} \frac{\sqrt{252RV_t} - \sqrt{252IV_t}}{g'(TV_t)\sqrt{2\widehat{QIV}_t}} \xrightarrow{d} \mathcal{N}(0, 1)$$

We can use the asymptotic distribution to create confidence intervals for the annualized integrated variance.

References

Dovonon, P. et al. (2014). *Bootstrapping High Frequency Jump Tests*. Tech. rep. Toulouse School of Economics.

Gonçalves, Silvia and Nour Meddahi (2009). “Bootstrapping Realized Volatility”. In: *Econometrica* 77, pp. 283–306.

Hounyo, Ulrich (2013). *Bootstrapping Realized Volatility and Realized Beta under a Local Gaussianity Assumption*. Tech. rep. University of Oxford.

Li, Jia, Viktor Todorov, and George Tauchen (2015). “Mixed-scale Jump Regressions”. In: *Working paper, Duke and Northwestern*.