Realized variance

April 5, 2020

A short document for the RV computation. Please let me know if you have any comments :)

0. create one dataset out of all .csv files

note 0: they don't say anything about averaging bid-ask quotes, but I assume they do? Also they say they got it from the Institute of Financial Markets but we can't find this institution. Use TAQ instead.

note 0_Ste: It's probably better to check if it is relevant and, in case, I'll download again the first part of the intraday prices. Since it seems that the Institute of Financial Markets doesn't exist I think it's the best choice.

1. create 5-min prices

Start with first return and keep only those that at 5-min distance.

Note 1: Is that also how you would do it? Problem here is also that we have rather not equispaced times in the raw data.

Note 1_Ste: We have to make a choice: I would use the first price and as a second one the first after ≥ 5 minutes. Then continue in this way until the end of the way. I know it's not super efficient but it seems that if you use returns closer than 5 minutes you risk to fuck up everything due to micro noises.

2. calculate realized variance

2.a. split in daily price and overnight

overnight: 4:00 PM to 9:30 AM

so here: don't use 5-min values but return between close and open the next day only.

Note Ste: yes.

2.b. calculate the total, upside and downside realized variance

total realized variance (during trading day)

$$RV_t = \sum_{j=1}^{n_t} r_{j,t}^2$$

with $r_{j,t}^2$ the j^{th} intraday log return and n_t the number of intraday returns observed that day.

However at one point they say they obtain the realized variance by just adding upside and downside realized variance (p.13). I'll try both and see if they are the same (should be if we didn't misunderstand the rescaling).

decomposition (during trading day)

$$RV_t^U(\kappa) = \sum_{j=1}^{n_t} r_{j,t}^2 \mathbb{1}_{[r_{j,t} > \kappa]}$$

$$RV_t^D(\kappa) = \sum_{j=1}^{n_t} r_{j,t}^2 \mathbb{1}_{[r_{j,t} \leq \kappa]}$$

with κ a threshold, in the paper $\kappa = 1$

overnight returns

$$r_{overnight} = log(p_{open,t}) - log(p_{close,t-1})$$

Add $r^2_{overnight}$ to total RV,
if $r_{overnight} > \kappa$: add squared return to $RV^U_t(\kappa)$,
if $r_{overnight} \leq \kappa$: add squared return to $RV^D_t(\kappa)$.

Note Ste: ok probably I am tired but I don't know what you mean with 'both'. I think that k=0 and not 1 (but also here I am tired). For the rest, it seems fine.

3. apply scaling

σ_r .. sample variance of daily log returns
μ_{RV} .. sample average of the (pre-scaled) realized variance
$\tilde{RV}^{U/D}_t$ .. unscaled realized variance, sum of squared log returns
$RV^{U/D}_t$ .. scaled realized variance, the one we work with
U/D .. U or D

$$RV_t = \tilde{RV}_t \frac{1}{\sigma_r}$$
$$RV^{U/D}_t = \tilde{RV}^{U/D}_t \frac{\sigma_r}{\mu_{RV}}$$

textsource: "we scale the RV_t series to ensure that the sample average realized variance euqlas the sample variance of daily log- returns. [...] we apply the same scale to the two components of the realized variance. Specifically, we multiply both components by the ratio of the sample variance of daily log-returns over the sample average of the (pre-scaled) realized variance."
(we dropped the argument κ for ease of notation)

why?
total realized variance: makes sense that realized variance should have the same variance as daily log returns
upside and downside variance: ??

other references: they refer on p.14 to Hansen and Lunde (2006) wo discuss various approaches to adjusting open-to-close RVs but I would try not too go too deep here, could be floorless.

Note 2: Is that also how you would read it? But anyway the reasoning why they do it is still an open question..

Note 2_Ste: Ok, I might have a good interpretation of this fact.
Why do they want to compute the realized variance daily? To estimate the variance of the returns day by day.
Why do they want to do this rescaling? Because on (daily) average, we can say that the realized variance is equal to the variance of the daily returns.
Why is this still strange to me? Because I still don't know how they give 'proper' weights to 'good' and bad' variances.

4. accumulating

h .. periods

$$RV_{t,h}(\kappa) = \sum_{j=1}^{h} RV_{t+j}(\kappa)$$
$$RV^U_{t,h}(\kappa) = \sum_{j=1}^{h} RV^U_{t+j}(\kappa)$$
$$RV^D_{t,h}(\kappa) = \sum_{j=1}^{h} RV^D_{t+j}(\kappa)$$

Interpretation: I.e. if h = 1 month, we take all daily RV of one month and sum them up. The periods should be *nonoverlapping* (hence we sum up january, february etc)

Note 3: Is that also how you would read it?

Note 3_Ste: It makes sense how you read it. However, I have a big doubt about this. How is it possible that the sum starts from 1 and the asymptotic result (the last two equations before section 1.2) involves the integral from t (and not t+1)? I know I haven't analyzed properly how they proved that result, but it seems strange to me that a variable without any information about time t gives a result with information about time t.

They use for h: 1,2,3,6,9,12,18,24 month.

5. construction of p-expectation

In the paper, they use 3 specifications: random walk, u/d-har, m-har, but only report the random walk, so I will focus for this atm.

random walk

$$\mathbb{E}_t^{\mathbb{P}}[RV_{t,h}^{U/D}(\kappa)] = RV_{t-h,h}^{U/D}(\kappa)$$

x. calculate variance risk premium

$$VRP_{t,h} = \mathbb{E}_t^{\mathbb{Q}}[RV_{t,h}] - \mathbb{E}_t^{\mathbb{P}}[RV_{t,h}]$$

hence assume the p-expectation for a period h (e.g. one month, say february) is just the realized variance from the past period (e.g. january same year)

Note 4: We don't have the Q expectation yet so I didn't deal with this so far.