Volatility forecast using HAR-RV model and partition by regimes

1. Consider sequence of observation of S&P500 index from 16th February 2001 to 27th August 2020 by 5-minutes intervals;

2. Find returns of S&P500 by 5-minutes intervals;

3. Compute Realized volatility for every day by formula

$$RV_t = \sqrt{\sum_{j=0}^{M-1} r^2_{t-j\Delta}} \qquad r_{t-j\Delta} = \ln\left(\frac{P_{t-j\Delta}}{P_{t-(j+1)\Delta}}\right),$$

   where t – index of a day (close moment of trading in day t), M – amount of high frequency returns in day t (computed by 5-minutes), $\Delta = 5\text{min}$, P_k – value of index in moment t. Note that we could use normal returns

$$r_{t-j\Delta} = (P_{t-j\Delta} - P_{t-(j+1)\Delta})/P_{t-(j+1)\Delta};$$

4. Divide sequence of realized volatilities $\{RV_t\}$ by 2 values using next scheme

$$V_t = \begin{cases} HV, & \text{if } RV_t > \alpha, \\ LV, & \text{if } RV_t \leq \alpha, \end{cases}$$

   where α - some threshold;

5. Using Baum-Welch algorithm find transition probability matrix and emission matrix of Hidden Markov Model (HMM);

6. Use algorithm Viterbi for finding the most likely sequence of hidden states;

7. Divide sequence $\{RV_t\}$ by regimes. The regime is in line with specified hidden state by Viterbi algorithm. I.e. we have k sequences $\{RV_t^{(i)}\}$, $i = 1,2,\ldots,k$, where k – amount of hidden states (regimes);

8. Compute realized volatility for weeks and months for every regimes by

$$RVW_t^{(i)} = \frac{1}{5}\sum_{j=0}^{4} RV_{t-j}^{(i)}, \qquad RVM_t^{(i)} = \frac{1}{22}\sum_{j=0}^{21} RV_{t-j}^{(i)}, \qquad i = 1,2,\ldots,k;$$

9. Estimate parameters of HAR-RV model by least squares estimation (LSE) on the different regimes. HAR-RV model for every regime is described by

$$RV_{t+1}^{(i)} = \beta_{0,i} + \beta_{1,i}RV_t^{(i)} + \beta_{2,i}RVW_t^{(i)} + \beta_{3,i}RVM_t^{(i)} + \varepsilon_{t+1}, \quad i = 1,2,\ldots,k,$$

where $\beta_{j,i}, j = \overline{0,3},\ i = \overline{1,k}$ – unknown parameters for every regimes. After estimate we have k vectors $\hat{\beta}^{(i)} = [\hat{\beta}_{0,i}, \hat{\beta}_{1,i}, \hat{\beta}_{2,i}, \hat{\beta}_{3,i}]^T$, $i = 1, \ldots, k$, where a^T – transpose operation for matrix a.

10. For forecasting of realized volatility of S&P500 by 1 day use next expression

$$\widehat{RV}_{t+1} = \beta_0^t + \beta_1^t RV_t + \beta_2^t RVW_t + \beta_3^t RVM_t,$$

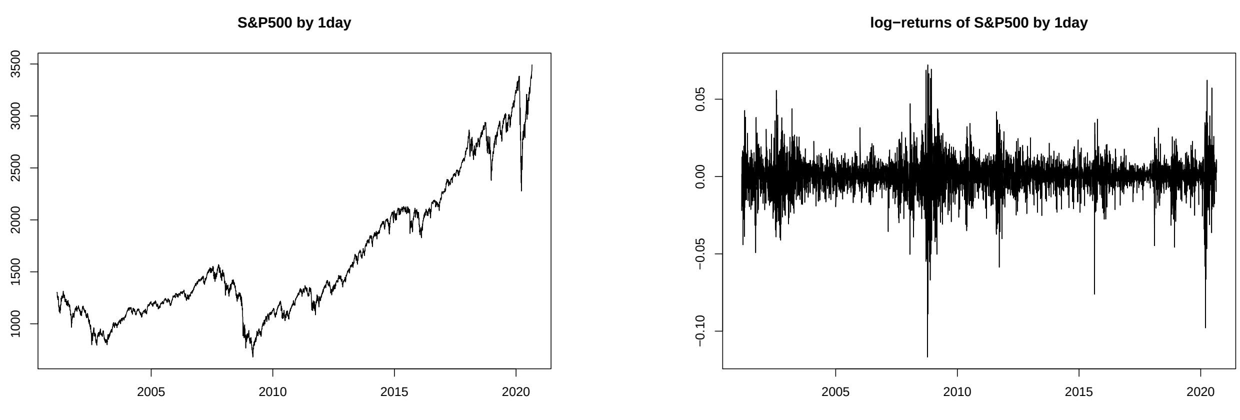
where

$$\beta_0^t = \hat{\beta}_{0,\,\mathrm{Reg}(RV_t)}, \qquad \beta_1^t = \hat{\beta}_{1,\,\mathrm{Reg}(RV_t)},$$
$$\beta_2^t = \frac{1}{5}\sum_{i=0}^{4} \hat{\beta}_{2,\,\mathrm{Reg}(RV_{t-i})}, \qquad \beta_3^t = \frac{1}{22}\sum_{i=0}^{21} \hat{\beta}_{3,\,\mathrm{Reg}(RV_{t-i})},$$

and $\mathrm{Reg}(RV_t)$ – index of regime in day t. I.e. coefficient β_0^t is estimation of parameter $\beta_{0,i}$ for regime i in day t, coefficient β_1^t is estimation of parameter $\beta_{1,i}$ for regime i in day t, coefficient β_2^t is mean of estimations for sequence of parameters $\beta_{2,i}$ for days from $t-4$ to t depending on regime, coefficient β_3^t is defined analogously to a coefficient β_2^t by month.

Note that

$$RVW_t = \frac{1}{5}\sum_{j=0}^{4} RV_{t-j}, \qquad RVM_t = \frac{1}{22}\sum_{j=0}^{21} RV_{t-j}.$$



a) b)

Figure 1 — a) Sequence of $S\&P500$ (from 16.02.2001 to 27.08.2020) b) Sequence of log-returns of $S\&P500$ (from 19.02.2001 to 27.08.2020)

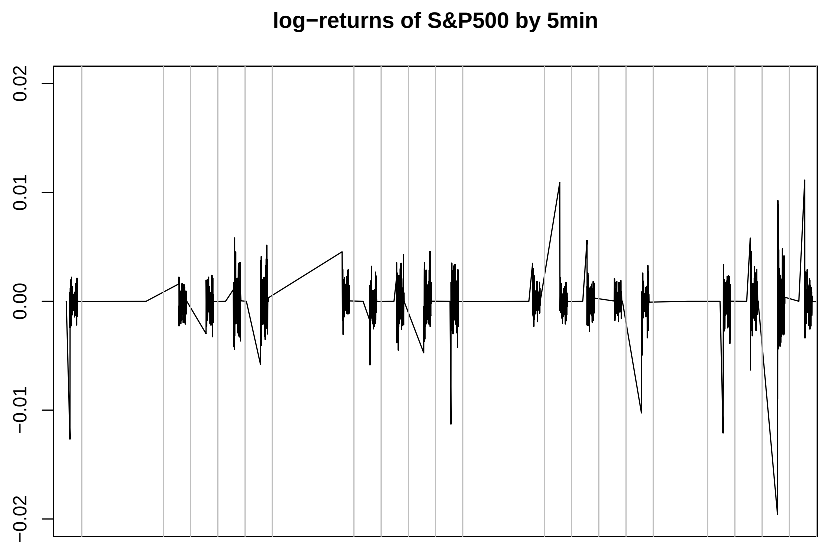


Figure 2 — Sequence of log-returns of $S\&P500$ by 5-minutes intervals from 17.02.2001 to 15.03.2001 (grey lines is close moment of trading)

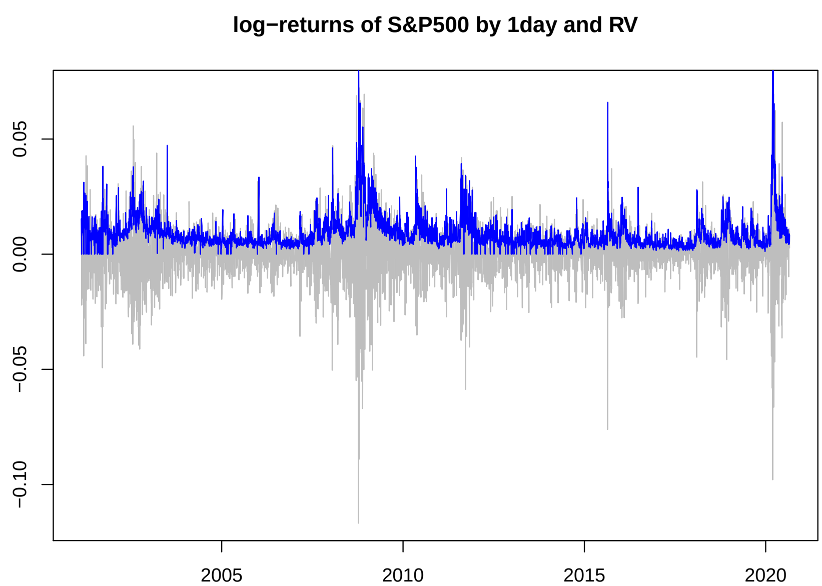
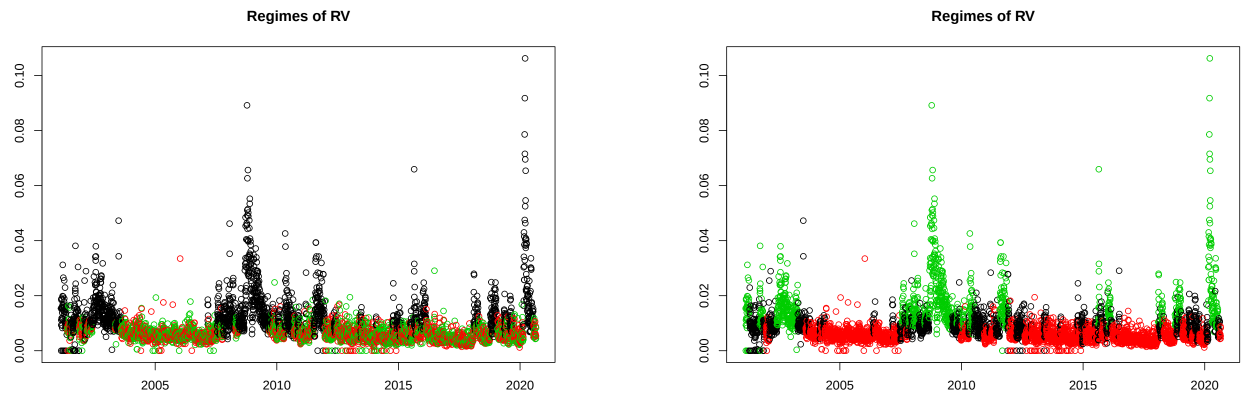
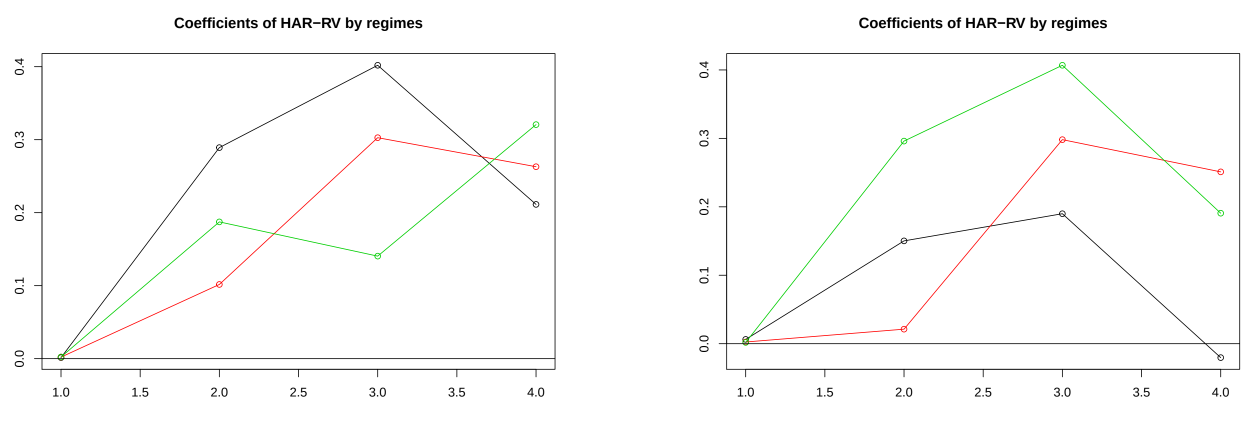


Figure 3 — Sequence of log-returns of $S\&P500$ by 1 days from 19.02.2001 to 27.08.2020 (grey plot) and sequence of realized volatility $\{RV_t\}$ (blue plot)



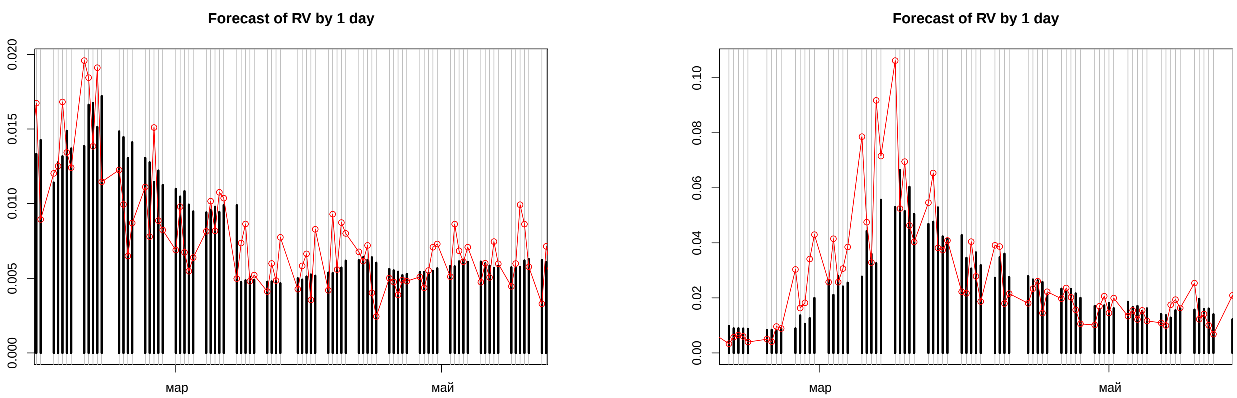
a) b)

Figure 4 — Dividing of $\{RV_t\}$ by 3 regimes depending on different initial values of transition matrix and emission matrix in Baum–Welch algorithm (from 19.02.2001 to 27.08.2020)



a) b)

Figure 5 — Estimations of parameters $\hat{\beta}^{(i)}$ depending on 3 regimes (black, green, red lines). At points 1,2,3,4 on x-axis the value of the estimation of the parameters $\hat{\beta}_{0,i}, \hat{\beta}_{1,i}, \hat{\beta}_{2,i}, \hat{\beta}_{3,i}$ respectively. Colour of line corresponds to regime. Case a) correspond to Figure 4.a), case b) correspond to Figure 4.b) (colours of regimes on Figure 4 and Figure 5 are related)



a) b)

Figure 6 — Forecasts of realized volatilities by 1 day (red line – realized volatility, black vertical lines – value of forecast): a) from 01.02.2016 to 20.05.2016, b) from 12.02.2020 to 20.05.2020