Homework 8 – Due Noon 5/22

Volatility Estimation Via Hidden Markov Chain

The file `XOM5YrsDaily.txt` contains daily returns in percent for Exxon Mobil over the five year (1258 trading day) period 1/3/2012 – 12/30/2016. In this project you will estimate historical XOM price volatility over this period using a hidden Markov chain approach. The underlying (imperfectly observed) Markov chain is XOM's time-dependent daily volatility in percent, σ_t. We assume it is undergoing an exponentiated random walk: $\sigma_t = \sigma \exp(\alpha W_t)$, where σ and α are model parameters. Here $0 \le t \le 1258$ is elapsed time in days and $(W_t : 0 \le t \le 1258)$ is a simple symmetric random walk on $\mathbb{Z}$ starting at $W_0 = 0$. The volatility process is not observed but we do observe a realization of XOM's daily percent returns $R_t \sim \sigma_t N_t$ where $(N_t : 0 \le t < 1258)$ are iid standard normals that are independent of the W process. Call this realization $(r_t : 0 \le t < 1258)$; r_t is the realized return from time t to time $t+1$. Here t is in (trading) days.

Preliminaries. Let us label the states of the MC according to the value of the random walk, so the state space is $S = \{k : -1258 \le k \le 1258\}$. At time t the only states with positive probability are

$$S_t \;=\; \{k : -t \le k \le t \text{ in steps of } 2\},$$

e.g., $S_0 = \{0\}$, $S_1 = \{-1, 1\}$, $S_2 = \{-2, 0, 2\}$, and so forth. For $k \in S$, let $s_k = \sigma \exp(\alpha k)$, the daily volatility in percent when the MC is in state k. To estimate σ_t at time t — the volatility of R_t — the only data we may legitimately use is $r_0, \ldots, r_{t-1}$. So put

$$D_t \;=\; \{R_0 = r_0, R_1 = r_1, \ldots, R_{t-1} = r_{t-1}\}, \quad \text{so also} \quad D_{t+1} \;=\; D_t \cap \{R_t = r_t\}.$$

For any event E, let $P_0[E] = P[E]$ (no data yet, so conditional probability = unconditional probability) and, for $t \ge 1$, let $P_t[E] = P[E \,|\, D_t]$. We note that $P_{t+1}[E] = P_t[E \,|\, R_t = r_t]$ (updating for D_{t+1} = updating for D_t and then for $\{R_t = r_t\}$).

Estimating Volatility. Suppose that, with our Bayesian prowess, we have calculated the numbers $p_k = P_t[W_t = k]$ for $k \in S_t$. We have this for $t = 0$: $p_0 = P_0[W_0 = 0] = P[W_0 = 0] = 1$. Using MAP estimation, put $k^* = \operatorname{argmax}_k \left(P_t[W_t = k] : k \in S_t\right)$ and take $\widehat{\sigma}_t = s_{k^*}$, so, e.g., $\widehat{\sigma}_0 = s_0 = \sigma$.

Bayesian Updating. Next we seek to update the p_ks, which reflect data through r_{t-1}, to reflect the new data $\{R_t = r_t\}$. To this end, for $k \in S_{t+1}$, put

$$\widetilde{p}_k \;=\; P_{t+1}[W_{t+1} = k] \;=\; P_t[W_{t+1} = k \,|\, R_t = r_t] \;=\; \frac{P_t[R_t = r_t, W_{t+1} = k]}{P_t[R_t = r_t]}.$$

Since the walk steps up or down by one each day, if $W_{t+1} = k$ we must also have either $W_t = k+1$ or $W_t = k-1$. We partition the numerator into those two cases. Given $\{W_t = j\}$, $R_t \sim \text{Normal}\,(0, s_j^2)$, so

$$\begin{aligned} P_t[R_t = r_t, W_{t+1} = k, W_t = k+1] \;&=\; P_t[R_t = r_t, W_{t+1} = k \,|\, W_t = k+1]\, P_t[W_t = k+1] \\ &=\; f(r_t, s_{k+1}) \cdot \frac{1}{2} \cdot p_{k+1}, \end{aligned}$$

where $f(r, s)$ is the density function of a Normal $(0, s^2)$ evaluated at r: $f(r,s) = \frac{1}{\sqrt{2\pi}s} e^{-r^2/2s^2}$. Similarly,

$$P_t[R_t = r_t, W_{t+1} = k, W_t = k-1] \;=\; f(r_t, s_{k-1}) \cdot \frac{1}{2} \cdot p_{k-1}.$$

It follows that

$$P_t[R_t = r_t, W_{t+1} = k] \;=\; \frac{1}{2} f(r_t, s_{k-1}) p_{k-1} + \frac{1}{2} f(r_t, s_{k+1}) p_{k+1}. \tag{1}$$

To complete the calculation of $\widetilde{p}_k$ for $k \in S_{t+1}$, we compute the denominator:

$$P_t[R_t = r_t] \;=\; \sum_{j \in S_t} P_t[R_t = r_t \,|\, W_t = j]\, P_t[W_t = j] \;=\; \sum_{j \in S_t} f(r_t, s_j) p_j, \tag{2}$$

yielding

$$\widetilde{p}_k \;=\; \frac{f(r_t, s_{k-1})p_{k-1} + f(r_t, s_{k+1})p_{k+1}}{2\sum_{j \in S_t} f(r_t, s_j)p_j}. \tag{3}$$

In (1), (2), and (3), $f(r, s)$ is a density rather than a probability, but it gives the right answer; why?

The numbers $\widetilde{p}_k$ then become the new p_k when we advance to updating for the next new data, namely $\{R_{t+1} = r_{t+1}\}$. In this manner we may compute the numbers $(p_k : k \in S_t)$ for $t = 1$, then 2, then 3, and so forth to 1258 — and compute the corresponding $\widehat{\sigma}_t$.

Your mission is to implement this model. The model has two free parameters, σ and α, which must be estimated using the data. I have done that for you! Take $\sigma = 1.11$ and $\alpha = 0.038$. The code `HiddenMC.cpp` will get you started. As it stands, this code reads in the time series $(r_t : 0 \le t < 1258)$ from the data file `XOM5YrsDaily.txt`. It also generates some output files as discussed below.

- Implement this model and generate the sequence $(\widehat{\sigma}_t : 0 \le t \le 1258)$. Do not report it, but ...
- Graphically present the annualized time series $(\sqrt{252}\widehat{\sigma}_t : 51 \le t \le 1258)$ and compare this to the same data that EWMA produced in class (this data is in the file `EWMA.txt`). You may wish to use the TeX file `HistoricalVols.tex` to do this, in which case you must put the hidden MC data in a file called `HMC.txt` in the same format as `EWMA.txt`. (The `Report` function in the starter code does this for you!)
- Generate a scatter plot of the points $((x_t, y_t) : 51 \le t \le 1258)$ where x_t is the EWMA annualized daily volatility in percent and $y_t = \sqrt{252}\widehat{\sigma}_t$ is the HMC annualized daily volatility in percent. You may wish to use the TeX file `ScatterPlot.tex` to do this, in which case you must put the data in a file called `ScatterData.txt`. (The `Report` function in the starter code does this for you!)
- Generate a (normal) histogram of the standardized returns $(r_t/\widehat{\sigma}_t : 51 \le t < 1258)$. You may wish to use the `NormalHistogram` function in `Functions.h` to do this. Use 20 "buckets". (The `Report` function in the starter code does this for you!)
- Graphically present the time series of standardized returns. You may wish to use the TeX file `ReturnTS.tex` to do this. (The required data file `StandardizedXOM.txt` is generated in the `Report` function!)