1. Apple stock.

<table>
<thead>
<tr>
<th>Type</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{0,t})</td>
<td>10.72</td>
<td>7.32</td>
<td>0</td>
<td>144.03</td>
</tr>
<tr>
<td>(\sigma_{1,t})</td>
<td>14.17</td>
<td>9.93</td>
<td>0.13</td>
<td>122.31</td>
</tr>
<tr>
<td>(\sigma_{2,t})</td>
<td>10.44</td>
<td>8.51</td>
<td>0.86</td>
<td>80.35</td>
</tr>
<tr>
<td>(\sigma_{3,t})</td>
<td>18.72</td>
<td>15.18</td>
<td>1.60</td>
<td>140.69</td>
</tr>
<tr>
<td>(\sigma_{5,t})</td>
<td>10.41</td>
<td>8.57</td>
<td>1.00</td>
<td>94.01</td>
</tr>
<tr>
<td>(\sigma_{6,t})</td>
<td>19.07</td>
<td>15.62</td>
<td>1.89</td>
<td>169.46</td>
</tr>
</tbody>
</table>

2. The time plot of volatility is in Figure 1. Let \(x_t\) be the log volatility, then the model selected is

\[
(1 - 0.078B - 0.056B^2)(1 - B)x_t = a_t, \quad \sigma^2 = 8.18 \times 10^{-4}
\]

where both the AR coefficients are significant. Model checking indicates the presence of possible outliers. The 1-step to 5-step predictions are \(-1.6350, -1.6362, -1.6370, -1.6372\) and \(-1.6372\).


(a) The fitted logistic regression model is

\[
\logit(p_t) = 0.26 + 0.016M_{t-1} - 0.161S_{t-1} + 0.294M_{t-2} - 0.037S_{t-2},
\]

where \(p_t = P(M_t = 1)\). The \(p\)-value of the coefficient estimate indicates that only the coefficient of \(M_{t-2}\) is significant at the 5% level. Thus, the model is marginally informative.

(b) The fitted network model is

\[ h(o_t) = \begin{cases} 1 & \text{if } o_t > 0 \\ 0 & \text{if } o_t \leq 0 \end{cases} \]

\[ o_t = -1.19 - 3.47h_{1t} + 2.56h_{2t} - 0.11M_{t-1} - 0.07S_{t-2} + 1.56M_{t-2} + 0.21S_{t-2}, \]

where

\[ h_{1t} = \frac{\exp(-3.96 - 4.25M_{t-1} - 2.64S_{t-1} - 5.24M_{t-2} + 2.65S_{t-2})}{1 + \exp(-3.96 - 4.25M_{t-1} - 2.64S_{t-1} - 5.24M_{t-2} + 2.65S_{t-2})} \]

\[ h_{2t} = \frac{\exp(0.70 + 0.28M_{t-1} - 0.61S_{t-1} - 4.93M_{t-2} - 0.49S_{t-2})}{1 + \exp(0.70 + 0.28M_{t-1} - 0.61S_{t-1} - 4.93M_{t-2} - 0.49S_{t-2})} \]
(c) In term of out-of-sample prediction, the sum of square of forecast errors for the logistic regression is 37 and that of the network is 34. The two approaches provide similar result with the neural network fares slightly better.

   
   • The 5-m log returns are shown in Figure 2.
   • No, there are no serial correlations in the 5-m log returns as we have $Q(10) = 11.23$ with $p$-value 0.34.
   • The RV is given below:
     
     ```r
     > v1
     [1] 0.008395630 0.006789461 0.006008014 0.009072835 0.008442985 
         0.011154396 0.009659880 0.008314613
     [9] 0.007924621 0.007276704 0.013619415
     ```

   • The RV is given below:
     
     ```r
     > v2
     [1] 0.010152794 0.009275675 0.007086352 0.010412172 0.009648235 
         0.009969376 0.009610830 0.009312424
     [9] 0.009315167 0.008163358 0.012399506
     ```

5. The tick-by-tick trade data of Google stock in May 2013.
   
   • See Figure 3(a).
   • Yes, there is a clear diurnal pattern in the trading intensity. See Figure 3(b).
Figure 1: Time plot of volatility via Yang-Zhang method

Figure 2: Time plot of intraday 5-m log returns of Google stock in May 2013.
Figure 3: Time plot and ACF for the numbers of trade in 5 minute interval of Googlestock in May 2013.