1. The reverse stock split of April 30, 2009 corresponds to \( t = 1382 \) of the data. After adjusting for the stock split, one can use simple calculation in R to solve the problem.

Answer: \( f = (24-6.5)/24 \). The summary statistics are as follows:

<table>
<thead>
<tr>
<th>Type</th>
<th>( \sigma^2_{0,t} )</th>
<th>( \sigma^2_{1,t} )</th>
<th>( \sigma^2_{2,t} )</th>
<th>( \sigma^2_{3,t} )</th>
<th>( \sigma^2_{5,t} )</th>
<th>( \sigma^2_{6,t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>3.175</td>
<td>5.195</td>
<td>6.769</td>
<td>11.88</td>
<td>7.077</td>
<td>12.88</td>
</tr>
<tr>
<td>Maximum</td>
<td>117.15</td>
<td>138.35</td>
<td>78.941</td>
<td>138.22</td>
<td>68.130</td>
<td>122.83</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0.133</td>
<td>0.244</td>
<td>0.157</td>
<td>0.289</td>
</tr>
</tbody>
</table>

2. Use the \( yz.R \) program with nsplit = 1 and split=c(1382,20). Alternatively, you can use adjusted open, high, low and close prices with nsplit = 0. The plot is shown in Figure 1. For modeling, let \( y_t \) be the log volatility series. The fitted model is

\[
(1 - 0.449 + 0.061B^2 - 0.126B^3 - 0.068B^6)(1 - B)y_t = a_t, \quad \sigma_a^2 = 0.0007.
\]

Model checking indicates a large outlier. For simplicity, we do not consider the outlier. The 1-step to 5-step ahead volatility predictions are 0.3188, 0.3181, 0.3178, 0.3177, 0.3177, respectively.

3. The basic conclusion of the problem is that the direction is not predictable from the data used.

(a) All coefficients, but the constant term, of the logistic regression are statistically insignificant at the 5% level.

(b) The fitted NNET model is

\[
\begin{align*}
    h_{1t} &= \frac{\exp(-3.25 - 3.14M_{t-1} - 1.08S_{t-1} + 0.56M_{t-2} + 1.91S_{t-2})}{1 + \exp(-3.25 - 3.14M_{t-1} - 1.08S_{t-1} + 0.56M_{t-2} + 1.91S_{t-2})} \\
    h_{2t} &= \frac{\exp(-3.70 - 3.56S_{t-1} + 2.88M_{t-2} + 4.18S_{t-2})}{1 + \exp(-3.70 - 3.56S_{t-1} + 2.88M_{t-2} + 4.18S_{t-2})} \\
    M_t &= \text{1 iff } 0.31 - 3.54h_{1t} + 3h_{2t} - 0.66M_{t-1} + 0.87S_{t-2} - 0.98M_{t-2} - 0.83S_{t-2} > 0
\end{align*}
\]
Figure 1: Volatility based on Yang and Zhang’s method

(c) The two models fare essentially the same because they are not adequate in predicting the direction $M_t$. The error rate for the logistic regression is 26 (out of 63) whereas that for NNET is 25.

4. Use the R script `hfrtn.R`.
   - The plot of the 5-minute log returns are in Figure 2.
   - Yes, there are serial correlations because $Q(10) = 28.09$ with p-value 0.0017.
   - Realized volatility based on 5-m returns are 0.00781, 0.00725, 0.00733, 0.00874, and 0.00989, respectively.
   - Realized volatility based on 1-m returns are 0.0106, 0.00752, 0.00769, 0.00846, and 0.00951, respectively.

5. Use the R script `hfntra.R`. There exists diurnal pattern. See Figure 3.
Figure 2: 5-m log returns

Figure 3: ACF of the series of numbers of trade within 5 minute interval