Statistics Citadel

Jeffrey R. Russell

Homework 9

1. Fit an AR(1) model to the Tbill data from last homework.
    a. from the last observation in the data forecast 15 periods ahead.
    b. Find the forecast error variance out through k periods ahead.
    c. Plot the forecast and a 95% predictive interval out through 15 periods. Again, the forecast should start from the last observation in your sample.

2. Use the dataset gdp.xls. This file contains real quarterly gdp measured in 2000 dollars and is seasonally adjusted. Take the natural log of the series.
    a. Assume that the log gdp series follows a random walk. Estimate the drift by calculating the sample average of the growth rate (i.e. changes in log gdp)
    b. You have data out through the last quarter of 2007. Forecast the quarterly gdp out through the last quarter of 2008.
    c. Construct a 95% predictive interval for each of the 4 forecasts.

3. Use your S&P500 data from yesterday. Construct the absolute value of the S&P500 daily returns.
    a. Plot the absolute value of the s&P500 returns.
    b. Examine the autocorrelation function for this series. What do these autocorrelations imply about volatility? Explain.
    c. What is the kurtosis of the returns (not absolute value). What does the kurtosis tell you about the distribution of the returns relative to a Normal distribution?

4. For the S&P500 data, construct a moving average of the squared returns. Use the k most recent observations to get an estimate of volatility at time t (you can't do this for the first k observations).
    a. Do this for k=100 and plot the annualized daily volatility series.
    b. Do this for k=252 and plot the annualized daily volatility series.

5. Construct the Riskmetrics exponential smoother for the daily S&P500 returns. Initiate the first variance ($\pm_1^2$) to the unconditional sample variance. Then construct the series for each period after the first.
    a. Do this for λ =.95 and plot the annualized daily volatility series.
    b. Do this for λ=.8 and plot the annualized daily volatility series.
    c. Do this for λ=.1 and plot the annualized daily volatility series.

6. Consider the GARCH(1,1) model for annual returns where

$h_t = .00045 + .04 r_{t-1} + .94 h_t$.

a. What is conditional variance at time *t* if r_{t-1}=.04 and h_{t-1}=.03?
b. What is the conditional variance at time t if r_{t-1}=.01 and h_{t-1}=.03?
c. What is the unconditional variance of r_t?

7. Use the S&P500 data of daily returns. The "likelihood" associated with an estimated model gives the log of the probability of observing the data given the estimated model. Hence larger values of the likelihood are consistent with models that fit the data better.

   a. Estimate an ARCH(1) model and report the Likelihood
   b. Estimate an ARCH(9) model and report the Likelihood
   c. Estimate a GARCH(1,1) model and report the Likelihood
   d. Estimate a GARCH(2,2) model and report the Likelihood
   e. Which model has the highest likelihood?

8. For your preferred GARCH model from the previous problem:

   a. Calculate the sequence of conditional daily variance for each day in the sample.
   b. Using the conditional standard deviation ($\sqrt{h_t}$), build a 2 standard deviation interval for the return in period t. Calculate the fraction of the daily returns fall into these intervals.
   c. Calculate the time series of annualized volatilities and plot them.

9. For this problem you will download recent price data from Yahoo Finance. There are better data sources available such as CRSP but this data source is not as current as Yahoo Finance. The Yahoo Finance web page is not terribly user friendly. Here is one way to get to the historical data. Click on this link http://finance.yahoo.com/q/hp?s=^GSPC . From this page you can enter beginning and ending dates for the data, as well as a ticker symbol in the upper right corner of the page. After you have selected dates and a ticker symbol, go to the bottom of the page where you can click on "download to Spreadsheet". Pick your favorite index (you can get ticker symbols for additional indices here: http://finance.yahoo.com/indices?e=dow_jones). Download 20 years worth of data for your favorite index. Pick your favorite stock that has been around for at least 20 years. Download the data for a single stock over the last 20 years.

   a. Report the skewness and the kurtosis for both the index and the stock. What do these numbers tell you about the distribution of the returns? Report the autocorrelations through lag 20 for the *squared* returns. What does this tell you about the time varying volatility in each series?

   b. Fit a GARCH(1,1) model to the index and the stock you chose. Report the results. Plot the sequence of one-day ahead annualized volatilities. How does the volatility of the index compare to that of your stock? Do you see any striking differences?

c. What is the implied unconditional variance for each model? How close is this to the sample variance of the returns?

d. In class we saw that we could write a GARCH model as $r_t = \sqrt{h_t} z_t$ where z_t is iid, mean zero and variance 1 so if the returns are conditionally normal then the z_t should be N(0,1). If we divide both sides of the equation by $\sqrt{h_t}$ then we get $\frac{r_t}{\sqrt{h_t}} = z_t$. This means that if we divide each in-sample return by its conditional variance these "standardized returns" should look like iid mean zero variance 1 random variables. If in fact the returns are conditionally normally distributed (i.e. that $f\left(r_t \middle| F_{t-1}\right) \sim N\left(0, h_t\right)$) then the z_t should look iid normal. Create this sequence of standardized returns for your two series. Present the histogram. Present the skewness and Kurtosis for the standardized series. Are these statistics (and the historgram) consistent with the conditional normality assumption? If not, how does the distribution differ from the Normal?

e. If the z_t's don't look normal (which they probably don't), what kinds of mistakes would you make if you assumed conditional normality i.e. $f\left(r_t \middle| F_{t-1}\right) \sim N\left(0, h_t\right)$ to get probabilities for the returns (i.e. what types of returns be predicted with too high of a probability)?

f. From the last day in the sample, build the sequence of daily forecasts from one-day ahead through 250 days. Plot the forecasts for both the index and the stock.

g. Cumulate the forecasts to get the forecasted variance for the year for both the index and the stock. How do these forecasts compare to the unconditional variance? (just give a general interpretation)

h. We can also use the standardize residuals to check if our GARCH model fits the data well. From part d, we have $\frac{r_t}{\sqrt{h_t}} = z_t$ so that $\left(\frac{r_t}{\sqrt{h_t}}\right)^2 = z_t^2$. Since z_t is iid, there should be no correlation in the z_t's but also, there should be no correlation in the squared z_t's. That's because if h_t is the correct conditional variance then dividing each return by its conditional standard deviation $\sqrt{h_t}$ should make result in a series with constant variance (constant variance of 1). Only if we divide by the correct conditional standard deviation will this be true. Hence, we can examine the autocorrelation in the squared z_t's to check to see if our fitted GARCH model fits the data well. Examine the autocorrelations for the *squared* standardized residuals out through lag 20 for each series. Has the time varying volatility been adequately captured by your GARCH model? If not, try a higher order GARCH model, and examine the autocorrelation in the (squared) standardized residuals again.

10. Use the S&P500 data for this problem. A mean variance investment strategy seeks to find a set of portfolio weights that delivers high returns with minimal risk. Consider choosing a portfolio of the risk free rate and the daily S&P500. The goal is to choose portfolio weight w where w denotes the fraction of wealth invested in S&P500 and (1-w) denotes the fraction of wealth invested in the risk free asset. The weight is chosen to maximize the utility to the investor. For this problem we assume a quadratic utility function. $U\left(w\right) = «_{w} - \frac{a}{2} ±_{w}^{2}$ where $«_w$ denotes the mean return of the portfolio and $±_{w}^{2}$ denotes the variance of the portfolio. Clearly the mean and variance depend on the chosen weight w.

    a. Let r_f denote the risk free rate and let $«_{SP}$ denote the mean S&P500 returns respectively. Let $±_{sp}^{2}$ denote the true variance the S&P 500 returns. Write out $«_w$ and $±_{w}^{2}$ as a function of the mean return of the S&P500, the risk free rate, and the variance of the S&P500 returns.
    b. Using the expressions of the mean and variance of the portfolio as a function of the portfolio weights from part a, take the derivative of U(w) with respect to w to solve for the optimal portfolio weight as a function of w. For risk aversion parameter λ=5, find the optimal portfolio weight and the corresponding utility.
    c. Assuming a risk free annualize rate of .02 set the daily risk free return equal to .02/252. Use the sample mean return of the S&P500 series (the daily data available from the homework web page) as an estimate $«_{SP}$, use the sample variance of the S&P500 returns as an estimate for $±_{sp}^{2}$. Find the portfolio weight w for these (fixed) values of mean and variance using a risk aversion parameter of λ=5.
    d. Find the utility obtained from this strategy (ie find U(w)).

11. We know that the volatility of the S&P500 is not constant, but is changing over time. Is it possible to time the market getting out of the S&P500 (and into the risk free) when volatility is high and getting back in when the market is less volatile? Let's extend the analysis in part 6 to account for time varying volatility thereby creating a dynamic portfolio allocation using portfolio weights that vary over time as the volatility changes. In the presence of time varying means variances the conditional utility at time period t can be written as $U_{t}\left(w\right) = «_{w,t} - \frac{a}{2} ±_{w,t}^{2}$. Clearly the optimal portfolio weights here are the same as those obtained in problem 2, part b. The only difference is that the portfolio weights must be chosen for each period t. For each period t use the variance from your favorite GARCH model to get the optimal weight for that period.

a. For each period t, find the optimal portfolio weights. Call the optimal weight for year t w_t. Plot the weights and the h_t's. Do the plots make sense?

b. For each period, calculate the returns realized from the optimal portfolio weights. That is, calculate $r_{w_t} = w_t r_{SP500,t} + (1-w_t)\left(.02/252\right)$

c. We evaluate the performance of this strategy by examining average utility obtained by this investment strategy. We therefore explicitly are accounting for the risk of the portfolio using the utility and not the average returns. Specifically, construct $\overline{U} = \overline{r} - \frac{a}{2}\frac{1}{T}\sum_{t=1}^{T}\left(r_{w_t} - \overline{r}\right)^2$ where λ=5 and $\overline{r} = \frac{1}{T}\sum_{t=1}^{T} r_{w_t}$.

d. Assuming the utility in each period $u_t = r_{w_t} - \frac{a}{2}\left(r_{w_t} - \overline{r}\right)^2$ is iid, calculate the standard error of the estimated utility.

e. $\overline{U}$ is an estimate of the actual utility obtained by your dynamic portfolio investment strategy. Test the null hypothesis that the utility from this strategy is equal to the value for the utility obtained from the fixed weights calculated in the last problem, part d.