

Homework 4: An Insufficiently Random Walk Down Wall Street

36-402, Advanced Data Analysis

Due at the start of class, 8 February 2011

In this assignment, you will use the same data set of values for the S& P 500 stock index that was used in the lecture notes for 1 February. You will need to download `SPhistory.short.csv` from the class website.

Problems 2 and 3 are about estimating the first percentile of the return distribution, $Q(0.01)$, under various assumptions. The returns will be larger than this 99% of the time, so $Q(0.01)$ gives an idea of how bad the bad performance will be, which is useful for planning. Note that a calendar year contains about 250 trading days, and so should average two or three days when returns are even worse than $Q(0.01)$.

Include code for all problems as an appendix. Clearly indicate which block of code is for which problem. Comment your code when at all possible.

1. (5 points) Load the data file, take the last column (containing the daily closing price), and calculate the logarithmic returns. Note that the file is in reverse chronological order (newest first). When you are done, if everything worked right, running `summary` on the returns series should give

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.094700	-0.006440	0.000467	-0.000064	0.006310	0.110000

Hint: Read the notes for 1 February.

2. In many applications in finance, it is common to model daily returns as independent Gaussian variables.
 - (a) (5 points) Find the mean and standard deviation of the best-fitting Gaussian, and the $Q(0.01)$ it implies.
 - (b) (5 points) Write a function which simulates a data set of the same size as the real data, using the independent Gaussian model you fit in part 2, and returns a list, with components named `mean` and `sd`, containing the parameter values estimated from the simulation output.

- (c) (5 points) Write a function which takes as arguments a list with components named `mean` and `sd`, and returns the first percentile of the corresponding Gaussian distribution. Check that it works by verifying that when run with mean 5 and sd 2, it returns 0.347. *Hint:* Look at the examples in the notes of parametric bootstrapping.
 - (d) (10 points) Using the code you wrote in (2b) and (2c), find a 95% confidence interval for your estimate of $Q(0.01)$ from (2a). *Hint:* Look at the examples in the notes of parametric bootstrapping.
 - (e) (5 points) What is the first percentile of the data? Is it within the confidence interval you found in (2d)?
3. (a) (5 points) Use `density()`, or any other suitable non-parametric density estimator, to plot the distribution of returns. Also plot, on the same graph, the Gaussian distribution you fit in problem 2. Comment on their differences.
 - (b) (10 points) Write a function to re-sample the returns, and calculate $Q(0.01)$ on each surrogate data set. Use this to find a 95% confidence interval for $Q(0.01)$. *Hint:* Look at the examples in the notes of non-parametric bootstrapping.
4. (15 points) In an **autoregressive** model, the measurement at time t is regressed on the measurement at time $t - 1$, $X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t$. Use `lm` to fit an autoregressive model to the returns. Give the estimates of ϕ_0 , ϕ_1 and $\text{Var}[\epsilon]$, and try to interpret what they mean. Also give the reported standard error for $\widehat{\phi_1}$.
 5. *Hint:* Look at the examples in the notes of re-sampling regression residuals.
 - (a) (5 points) Write a function which re-samples the residuals of the autoregressive model from (4). Check that the mean and standard deviation of its output are close to those of the residuals.
 - (b) (15 points) Write a function which simulates the autoregressive model you fit in (4), with noise provided by the function you wrote for (5b).
 - (c) (5 points) Write a function which takes a time series, fits an autoregressive model, and returns the estimate of ϕ_1 . Check that it works by seeing that when it's give the data, the output matches what you found in (4).
 - (d) (10 points) Using the function you wrote in (5c), and the simulator you wrote in (5b), find the bootstrap standard error for $\widehat{\phi_1}$. Does it match what `lm` reported in (4)?

Note: If you cannot solve (5b), you can get full credit for (5d) using the built-in function `arima.sim` instead, but make sure that the distribution of innovations or noise comes from the function you wrote in (5a). If you cannot solve (5a), you can get full credit for (5b) and (5d) by providing suitable Gaussian noise.