2 · *Fitting equations*

Due Monday, February 8, 2016

(1) The CAPM and a stock's "beta"

If you've taken a finance class that covered anything about asset-pricing theory, you may know that the Capital Asset Pricing Model (CAPM) assumes that the rate of return on an individual stock is linearly related to the rate of return on the overall stock market. Roughly speaking, this means that each stock's returns follows a linear regression model That is,

$$Y_t^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} X_t + e_t^{(k)},$$

where $Y_t^{(k)}$ is the return of an individual stock (k) in some given time period t; X_t is the return of the entire stock market in that same time period; and $e_t^{(k)}$ is the residual for stock k in that time period. The superscript (k)'s here are simply denoting the different stocks, while the subscript t's are denoting the different time periods. The market (X_t) is a predictor common to all stocks.

The data set "marketmodel.csv" contains information on the daily percentage returns for the S&P 500 stock index, along with 6 individual stocks: Apple, Google, Merck, Johnson and Johnson, Wal-Mart, and Target. The data are from February 1, 2011 through January 30, 2012. The entries are interpretable as percentages—for example, looking at the first row, we see that the S&P 500 gained 1.595% in value on February 1, 2011, while Target lost 0.335% in value on that same day.

(A) Regress the returns for each of the 6 stocks individually on the return of S&P 500 (which is like X_t, the market return, in the equation above). Make a table that shows the ticker symbol, intercept, slope, and residual standard deviation for each of the 6 regressions. Which stock seems to be the most tightly coupled to the movements of the wider market?

(B) What do you notice about the intercepts? Are they mostly small, or mostly large? Interpret these intercepts in terms of whether any of the individual stocks appear to be outperforming the market.[1]

(C) Does your estimate of the slope for Wal-Mart versus the S&P 500 agree (roughly) with the "beta" reported by Yahoo Finance?[2] If you notice a discrepancy, offer a possible explanation.

(D) Assess the evidence in the data for the following claim: "*Even after adjusting for their shared dependence on the broader market*, we should

[1] Note: these intercepts are often referred to as "alpha" rather than β_0 in the finance community, e.g.`http://en.wikipedia.org/wiki/Alpha_(investment)`.

[2] Check for yourself on: `http://finance.yahoo.com/q/ks?s=WMT`. The beta is in the right-most column, under "Trading Information." If you don't quite know how to interpret a stock's beta, do a bit of background reading, e.g. `http://en.wikipedia.org/wiki/Beta_(finance)`.

expect Wal-Mart's stock market returns to be most closely related to Target's returns than with any of the other four firms, because they are both large retailers." (Hint: Consider the 6 models you fit in Part (A). Each model leads to a set of residuals for one particular stock. Which set of residuals has the largest correlation with the residuals for the model having Wal-Mart as the response variable? Why do you think this is so?)

Note to those who might enjoy playing around with stock-market data: if you install the R package "fImport" and use the command `yahooSeries`, you can get data on any stock, for any time period, that you want! For example, I got this data using the following commands:

```
library(fImport)
mytickers=c("SPY", "AAPL", "GOOG", "MRK", "JNJ", "WMT", "TGT")
X = yahooSeries(mytickers, from = "2011-01-31", to = "2012-01-31", frequency="daily")
```

I then constructed the percentage returns from each "Adjusted Closing Price" column.

(2) Demand curves

The data in "milk.csv" contains a random sample of daily sales figures for a small neighborhood grocery store of cartons of chocolate milk. (Think of something like Fresh Plus in West Campus or Hyde Park.) The "price" column gives the price at which the chocolate milk was sold that day; the "sales" column says how many units were sold that day.

Let's say that the store's wholesale cost of milk is c dollars per carton. If you were the merchant and wanted to maximize profit, how much would you charge for a carton of chocolate milk? Explain your thought process carefully, and express your final answer in terms of c. Also, calculate how much profit you'd expect to make if the cost per carton were $c = \$1$.

Some points to think about here. . . . The store can choose what to charge for chocolate milk in order to maximize profit. Can you write an equation for profit in terms of the price charged? How can the data be used to help you write this equation? What tools should you use in order to figure out the choice of price that maximizes profit?

(3) Should we aggregate or not?

For this question, you will need the "TenMileRace" data set from the `mosaic` package in R, which you will load using the command

`data(TenMileRace)` after having loaded the `mosaic` package at the beginning of your R session. Quoting the data set description: "The Cherry Blossom 10 Mile Run is a road race held in Washington, D.C. in April each year. The name comes from the famous cherry trees that are in bloom in April in Washington.... This data frame contains the results from the 2005 race." If you type the command `?TenMileRace`, you will get a description of each variable in the data set.

(A) Fit a regression model to quantify the relationship between a runner's net finishing time (in seconds) and his or her age in years. What seems to be the effect of one additional year of age on finishing time?

(B) Now fit two separate linear models for finishing time versus age: one for men alone, and one for women alone. Within each subset, what seems to be the effect of one additional year of age on finishing time? Is this consistent with what you found in Part A? Describe what you think is going on here. Remember from the software walkthroughs: you can create a new data set from a subset of the original one using the subset command. For example:

```
women = subset(TenMileRace, sex=="F")
```

Notice the quotation marks and the double-equals sign, which is how we test for whether a variable takes a specific value.