Risk and Portfolio Management with Econometrics, Courant Institute, Spring 2009
http://www.math.nyu.edu/faculty/goodman/teaching/RPME09/index.html
Always check the class bboard on the blackboard site from home.nyu.edu (click on academics, then on the course name) before doing any work on the assignment.

Assignment 6, due March 11

Corrections: (none yet)

1. Consider a one dimensional regression $y = a + bx$. We have n pairs X_k, Y_k and want to find the values $\widehat{\alpha}$ and $\widehat{\beta}$ that solve the minimization problem

$$F(\alpha,\beta) \;=\; \sum_{k=1}^{n} \left(Y_k - (a + bX_k)\right)^2 \;,$$

$$\min_{\alpha,\beta} F(\alpha,\beta) \;.$$

Write formulas for $\partial_\alpha F(\alpha,\beta)$ and $\partial_\beta F(\alpha,\beta)$. Find formulas for $\widehat{\alpha}$ and $\widehat{\beta}$ that come from setting $\partial_\alpha F = 0$ and $\partial_\beta F = 0$. (Many of you will have seen these formulas, and they are in almost every book. They are important enough to bear repeating.)

2. We have studied the general m component linear regression problem

$$\min_b \|Y - Xb\|^2 \;,$$

where $b \in R^m$, $Y \in R^n$, and X is an $n \times m$ data matrix. We saw in class that we can find b using the normal equations

$$Ab = X^t Y \quad , \qquad A = X^t X.$$

Here A is an $m \times m$ matrix which is positive definite whenever X has full rank m. Formulate the linear regression problem from Problem 1 in this general setting, setting $b_1 = \alpha$ and $b_2 = \beta$. What is the $n \times 2$ matrix X in terms of the numbers X_k from Problem 1? What is the 2×2 matrix A? What is A^{-1} (you should start recognizing the quantities that appear in the answer to Problem 1)? Finally, multiply $A^{-1}X^tY$ and show that you get the same answer as in Problem 1.

3. This problem discusses a statistical test for a regression coefficient being significant. The statistical model is that there is a given $n \times m$ matrix X and a $b \in R^m$ with $y = Xb \in R^n$. None of these quantities are random. There are i.i.d. standard normals Z_k that are the components of the Gaussian vector $Z \in R^n$. The observations are $Y_k = y_k + \sigma Z_k$. The least squares estimate of b is $\widehat{b} = X^\dagger Y$, where $X^\dagger$ is the pseudo-inverse $X^\dagger = \left(X^t X\right)^{-1} X^t$. We use $X_k^\dagger$ to denote row k of $X^\dagger$. This is a row vector with n components.

(a) Show that, under the above hypotheses,

$$\left\|Y - X\widehat{b}\right\|^2 \sim \sigma^2 \chi^2_{n-m} \ .$$

Here, $A \sim B$ means that A is a random variable whose distribution is the same as the random variable B, and χ^2_k is the sum of squares of k independent standard normals.

(b) Use this to show that

$$\widehat{\sigma^2} \ = \ \frac{1}{n-m} \left\|Y - X\widehat{b}\right\|^2$$

is an unbiased estimator of σ^2.

(c) Under the above hypotheses, and with the above notation, show that $\widehat{b} - b \sim \mathcal{N}(0, \sigma^2_k)$, with

$$\sigma^2_k \ = \ \sigma^2 \left\|X^\dagger_k\right\|^2 \ .$$

(d) Using the above, show that

$$\frac{\widehat{b}_k - b_k}{\left\|X^\dagger_k\right\| \sqrt{\widehat{\sigma^2}}\sqrt{n-m}} \ \sim \ t_{n-m} \ , \qquad (1)$$

where t_k is the Student t random variable whose distribution is given by $\sqrt{k}Z/\sqrt{\chi^2_k}$, where Z is a standard normal independent of χ^2_k. Of course, part of establishing (1) is showing that the distribution of the left side is independent of σ.

(e) Suppose $n = 100$, $m = 5$, and $b_k = 0$. Define T by

$$T \ = \ \frac{\widehat{b}_k}{\left\|X^\dagger_k\right\| \sqrt{\widehat{\sigma^2}}\sqrt{n-m}} \qquad (2)$$

Show that there is a less than 1% probability that $T > 3$. You may use a t table, or you may use the normal approximation, which probably is reasonable accurate with the present parameters.

(f) (*No action items here, just comments*) The "t test" from elementary statistics may be used to test the hypothesis that a regression coefficient is equal to zero. It is common to report *t scores* for regression coefficients, which are the numbers (2). We discussed ill-conditioned regression problems that have very large regression coefficients. We see in (1) that $\left\|X^\dagger_k\right\|$ is in the denominator. This means that if the $X^\dagger_k$ is very large, it is hard for even a large $\widehat{b}_k$ to be statistically significant. If we estimate m regression coefficients, it makes sense to set to zero any coefficient that is not statistically significant, say with a t-score below 2.

4. Download the file `yc.xls`. This is interest rate data from the Fed going back a little more than a year. For example, if the 6 month (= 1/2year) rate is 3.9%, that means that if you invest \$1 on that date, you may elect to receive $\$1 \times e^{3.9\times\frac{1}{2}}$ in six months.

   (a) Upload this data into Matlab using the command
   `yc = xlsload('yc.xls');`
   The result should an array with 12 columns and 292 rows. Check this using `size(yc)`. The first column consists of dates and the last 11 are interest rates. Remove the first column, leaving an 11×292 array, also called `yc`. Each row of `yc` consists of 11 interest rates, for periods ranging one month to thirty years. This is a partial *yield curve* for that date. (It is common to create a more detailed yield curves using other interest rate information, but this is not the subject of this class.)

   (b) The Matlab command `plot( yc( t,1:11) );` produces a plot of the yield curve on date t. Create a length 11 vector x with $x(k) = k$. Then `plot( x, yc( 10,1:11),x, yc( 54,1:11) );` produces, on the same plot, the yield curve on dates 10 and 54. Experiment with this and create one plot that has several yield curves that illustrate the different ways the yield curve looked during this period. The curve changes by having different overall height, different slopes, and different degrees of convexity. Try to find these in the data.

   (c) Use the command `[U,S,V] = svd(yc);` to compute the singular value decomposition – the principal component analysis – of yc. Comment on the sizes of the singular values. Verify that $yc = U\Sigma V^t$ by doing the multiplication on the right in Matlab.

   (d) For an $n\times m$ matrix X, let $\|X\|_{HS}^2$ be the *Hilbert Schmidt* or *Frobenius* norm

   $$\|X\|_{HS}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}^2 .$$

   This is the sum of the squares of the matrix elements. We saw on previous assignments that if $X = U\Sigma V^t$, then $\|X\|_{HS}^2 = \sum_{i=1}^{m}\sigma_i^2$ (assuming $m \leq n$ all the time). We also showed that if Σ_k is the matrix containing only the k largest singular values, then

   $$\left\|X - U\Sigma_k V^t\right\|_{HS}^2 = \sum_{j>k}\sigma_j^2 .$$

   Verify this for $k = 0, 1, 2, 3$ for the SVD of yc.

   (e) Let v_k be the row vector that is row k of V^t. (Warning, in Matlab, this is not exactly the same thing as saying that v_k is column k of V.) Plot v_1, v_2, and v_3 on the same plot. Verify using Matlab that they are orthonormal ($\|v_k\| = 1$, $v_j v_k^t = 0$ for $j \neq k$. $v_j v_k^t$ is a number because v_k is a row vector.) Do the v_k look orthogonal in the plot?

Comment on the shapes of the three curves. When I did it, I got v_1 with all negative components. We could fix it by multiplying v_1 and some entries of U by -1. This is cosmetic, and does not effect any of the calculations here. There is no reason for you to do it. These first three principal components are called *level*, *slope*, and *convexity* respectively.

(f) We can fit the yield curve on a given date, t using some number of principal components. Let y_t be row t of yc. This is a row vector with 11 components. Find formulas for the best least squares coefficients a_t, b_t, and c_t to minimize the sum of squares $\|y_t - (a_t v_1 + b_t v_2 + c_t v_3)\|^2$. For $t = 11$, plot the four curves y_t, $a_t v_1$, $a_t v_1 + b_t v_2$, and $a_t v_1 + b_t v_2 + c_t v_3$. *Put all four curves on the same plot rather than making four separate plots.* Describe the improvement in the fit as you add more principal components.

(g) Experiment with different t values in part (f) to see how exceptional $t = 11$ was. It will be a great time saver to use Matlab tools such as m-files. Otherwise, you will be typing the same commands over and over.

(h) Show that the row vectors $a_t v_1 + b_t v_2 + c_t v_3$ are the same as the rows of the matrix $U \Sigma_3 V^t$ described above. This makes parts (f) and (g) much simpler.