Statistics 240
Problem Set 3
Due: Thursday March 2 in class

The new data for this problem are interest rates, and foreign exchange rates, in the course web page.

There will be one more problem set, due on Thursday March 16 at 5pm. This is the day that would have been used for an in class exam, had there been one. That final problem set will be worth slightly more than the others.

Your final grade will be a weighted average of four problem set grades. Your single worst problem set will be downweighted by a factor of 0.5. [Algorithm: for each problem set consider replacing it by 0.5 times its score plus 0.5 times the average of the other three scores. Whichever replacement produces the greatest final average is the one that will count.]

1. **Yield curves**

Construct a set of monthly data on Swedish interest rates. Keep all 5 maturities and as many common (non-NaN) observations as possible.

(a) What range of dates were available at all maturities?
(b) Plot the Swedish interest rates versus time.
(c) Were there any dates on which the Swedish 3 month rate was higher than the Swedish 10 year rate? If so, name one.
(d) Print the covariance (not correlation) matrix for these interest rates.
(e) What fraction of the variance is explained by the largest principal component?
(f) What fraction of the variance is explained by the largest two principal components?
(g) Plot the eigenvectors corresponding to the largest two principal components versus the corresponding maturities.
(h) Plot $\mu \pm \sqrt{\lambda_1} e_1$ versus $\tau$, where $\tau$ has components $(0.25, 1, 2, 5, 10)$ corresponding to the 5 maturities in years, $\mu$ is a comparable vector of mean interest rates (over the time points with full data), $\lambda_1$ is the largest eigenvalue, and $e_1$ is the corresponding eigenvector.
(i) Repeat that plot for the second largest eigenvalue $\lambda_2$. 

1
Sometimes principal components are interpretable, sometimes not. Can you interpret the first two principal components for this data?

Plot the trajectory taken by the first two principal components of Swedish interest rates over time, as described below. Suppose that the interest rates are in a matrix $X$ with one row for each time point and one column for each maturity. The matrix with rows $X - \mu$ records at each time how far above or below the mean the rates are. The vector $\sqrt{\lambda_1}(X - \mu)e_1$ gives the number of standard deviations away from the mean for the largest principal component at each time point. (Here $e_1$ is a column vector.) The vector $\sqrt{\lambda_2}(X - \mu)e_2$ is the analogous quantity for the second principal component. Plot these quantities, with line segments joining consecutive times. Label some or all of the points on this curve with their dates. (Using matlab functions TEXT and NUM2STR, or otherwise.)

Based on this plot, what is happening to Swedish interest rates in early 1996?

Can you say when that trend started, and when it ended?

2. Exchange rates

The data for this problem has the number of units of each foreign currency that can be purchased for one US dollar. There are 40 currencies including the US dollar. The data is monthly from January 1977 to January 2000. The data have been cleaned for you.

For this data it makes sense to consider principal components of the correlation matrix not the covariance matrix, because the values are in different units.

(a) Explain why the correlation matrix is only reasonable for the first 39 variables in the data.

(b) How many principal components are required to capture 90% of the variance (in the correlation scale)? How many for 95% and how many for 99%?

(c) What is the smallest eigenvalue? Comment on it.

(d) Construct the monthly log returns for each currency in dollars

$$-\log(x_{t,i}/x_{t-1,i}), \quad t > 1$$
and answer the questions in part (b) using the covariance scale for log returns. Only use the first 39 currencies.

(e) Find one interpretable principal component in either analysis and give an interpretation of it.