

Big Data: Homework 7

Will Clark & Matthew DeLio
41201-01

May 28, 2015

1 Foreign Exchange Factor Modeling

2 Principal Components Analysis

At first glance, our data set of monthly foreign exchange rates seems to have a lot of noise and very little signal. The goal of principal components analysis is to see if there are any patterns connecting variables in our data set, and if there are, use these relationships to reduce the dimensionality of our data. To this end, we use `prcomp` to find the principal components of foreign exchange movements. For each observation i , which is a month of foreign exchange rates for 23 currencies, the method estimates:

$$E[x_i] = \varphi_1 v_{i,1} + \varphi_2 v_{i,2} + \dots + \varphi_k v_{i,k} \quad (1)$$

We can now represent the data along the new set of dimensions $v_{i,j}$, which should reveal any latent patterns that were not observable when we were looking at the set of original dimensions $x_{i,j}$.

We can start by looking at the scree plot of our PCA, shown in Figure fig. 1 on the following page. This shows us the sorted eigenvalues of the covariance matrix of the scaled data; the highest eigenvalue is the principal component that explains most of the variation in the data. The steep drop off after the first bar tells us that the first principal component explains a large degree of the variability in our data.

We can look at the rotations on the first principal component and see if there is any obvious interpretation. Figure fig. 2 on the next page shows the rotations of PC1 on each country; countries in red have floating exchange rates and countries in blue have fixed exchange rates (Venezuela, China, Hong Kong, and Sri Lanka). Given that all the pegged exchange rates are on one side and all the floating rates are on the other, we can tentatively conclude that the first principal component is really telling us about the fixed/floating divide. It makes sense that most of the variation in exchange rates would occur between those that are allowed to move freely and those that are not, and since PCA is supposed to find the latent sources of variation, it follows that this is the first dimension on which it chooses to sort the data.

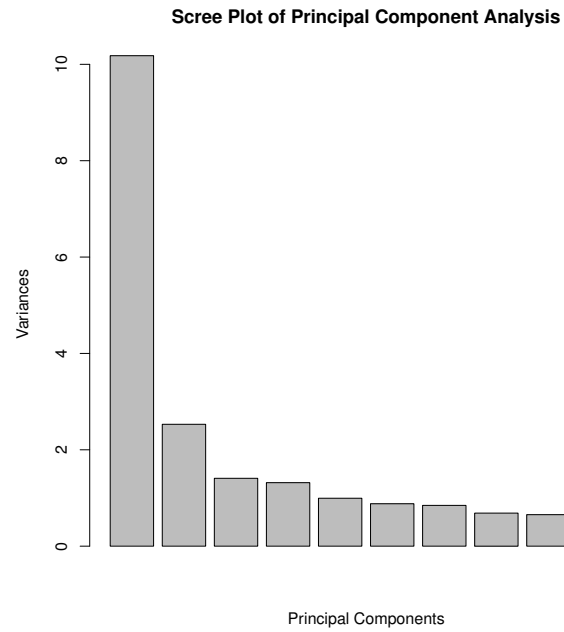


Figure 1: Foreign Exchange PCA Scree Plot

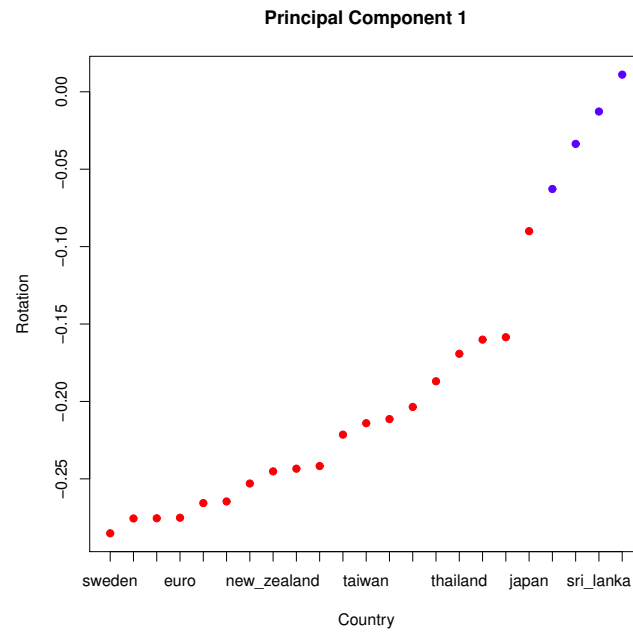


Figure 2: Distribution of First Principal Component

3 S&P 500 Returns on Currency Factors

4 Regression on All Covariates