

Project 1: Quantitative Analysis of the Correlation of Stock

Returns in the Brazilian Equity Market

Objective: Build a risk factor model for Brazilian equities, which will form the basis for a ``portfolio risk margin'' calculator.

Background: Brazil, Latin America's largest economy and the 6th economy in the world by GDP, has a relatively thriving stock market. There are 600+ listed companies, and there are 6 ETFs. We would like to use PCA to build a reasonable risk model and apply it to portfolio risk-calculations. One of the main issues that we have to address is how much correlation/decorrelation exists in the market and how we can model it using factors. The Excel sheet that contains the data also has a breakdown of all listed stocks into sectors and sub-sectors. You can take a look at the sheet to see how the Brazilian companies can be classified in terms of activity. PCA is supposed to look at this diversity using the lens of eigenvalues-eigenvectors.

The data: We have closing prices from Jan 1, 2010 to August 10 2012 on all listed names. We also have a list of the largest companies by market capitalization. I separated and cleaned the price history of 137 companies which have more than 1 billion BRL (Brazilian Reals) in market cap and for which complete (or almost complete) data is available during the period. The price data was downloaded from Bloomberg. It is contained in the file *BrazilPriceData.csv*. [The closing price in the original dataset is given for every day of the year. With a bit of cleaning, we have removed weekends and holidays. The number of returns dates, i.e. intervals for which there are non-trivial price changes, is 649.]

Question 1: Using the entire dataset (649 days), compute the eigenvalues and eigenvectors of the correlation matrix for the returns of stocks of large cap companies (137 names).

It is often the case in Brazil that there is more than one issue per company (e.g. PETR3 and PETR4), because some shares have voting rights and others do not. In the clean data for large caps, we chose the series that is the most traded by looking at another file (not shown) with volume data.

Using random-matrix theory, find a minimal set of eigenvalues/eigenvectors for which the corresponding residuals can be assumed to be white.

Question 2: Once you have determined the significant eigenvectors, sort the entries by decreasing values, from the largest to the smallest. Do the same thing with the weights of the corresponding ``eigenportfolios'' (i.e. coefficient/volatility). Verify that most entries in the 1st eigenportfolio have the same sign and identify companies with negative sign. Can you say why they have negative sign? [Hint: Use the page with the industry sector description]. Perform a similar analysis on the second eigenportfolio: can you observe any pattern in terms of signs. Identify any sector which participates with positive or negative sign across companies.

Question 3: Perform a time-dependent PCA in which you calculate the eigenvalues of the correlation matrix using 252-day windows and 126-day windows. Notice that the theoretical spectral cutoff $\lambda^+ = (1 + \sqrt{N/T})^2$ changes. What are the consequences of this? Explain the variability of the MP cutoff in terms of ``significant modes'' in the time-dependent setting.

Question 4: The downgrade of the U.S. Government debt by the ratings agency Standard and Poor's took place in the later summer of 2011. Is this event reflected in the behavior of the density of states? (With 126 or 252 days).

Question 5: Based on all of the above, what do you think is a reasonable choice for a set of mathematical risk factors (using T=252) for the Brazilian equity market?

Question 6: Using the last year of data (252 days), and returns of the large capitalization stocks, determine significant eigenvectors and the corresponding series of eigenportfolio returns. For simplicity normalize the returns so that the variance is 1 .This data should be a matrix of **252 rows and m columns**, where m is the number of significant eigenvectors.

Question 7: Consider all stocks in the Brazilian big universe which have complete price data for the last year in the dataset. For each of these stocks, compute the corresponding regression coefficients of the stocks with the standardized eigenportfolio returns (the factor loadings for each stock). Compute also the variance of the residual – the ``unexplained variance'' for each stock.

Question 8: Consider the risk-model

$$R = \sum_{k=1}^m \beta_k \xi_k + \left(\sigma^2 - \sum_{k=1}^m \beta_k^2 \right)^{1/2} \zeta_R,$$

Here (ξ_k) are independent random variables with Student-t distribution with 4 degrees of freedom, with mean zero and variance 1. These r.v. s represent common factors driving all

stocks. The variable ζ_R is also student-t with mean zero and variance 1, but represents the idiosyncratic risk for each stock.

Based on this model, calculate **99.5% VaR** for a portfolio consisting of a portfolio consisting of long BRL 10,000,000 in BOVA11 (an ETF), short BRL 7,000,000 in PETR4 (Petrobras), short 6,000,000 in VALE5 (Vale do Rio Doce), long 3,000,000 in BRFS3 (Brazil Foods). The 99.5% VaR is the portfolio loss that would occur with probability no greater than 0.5% according to the model.