Portfolio Diversification Index (PDI) with PCA

Introduction

Diversification is one of the basic tenets of portfolio management. Common approaches to the problem include correlation matrices of historical returns (e.g. Depfel 2003 and Toikka et al 2004) and types of cluster analyses (e.g. Brown and Goetzmann 2003 and Lhabitant 2004). Another method that will be employed here is Principal Components Analysis (PCA). PCA has been used in hedge fund studies and applied to Commodity Trading Advisor Funds. Rudin (2006) considers correlation matrices inefficient due to the calculation intensity as the number of securities in a portfolio increases. A correlation matrix will have a size of $n * (n-1)/2$ given 'n' securities, doubling securities from 10 to 20, will increase the matrix size from 45 to 190, i.e. more than quadruple. PCA enables a user to reduce the diversification problem to a much smaller number of variables. With this technique, the following will be an attempt at replicating Rudin's Portfolio Diversification Index (PDI), a single measure of diversification of a portfolio of assets.

Principal components analysis works by transforming correlated variables into the same number of uncorrelated variables, referred to as principal components. The first principal component captures as much of the variability of the data as possible. The second component counts for as much of the remaining variability in the data that is also uncorrelated with the previous principal component(s). So on and so forth. Ultimately, PCA's objective is to reduce the dimensionality of a dataset such that much of the variation in the data is captured with fewer variables. Along with principal components, PCA also generates what Rudin called a vector of relative strengths of each component. Essentially, the vector are comprised of percentages of total variance that is attributable to a given component, i.e. the percentage of variability of the data captured by each component.

Then, the PDI is given by:

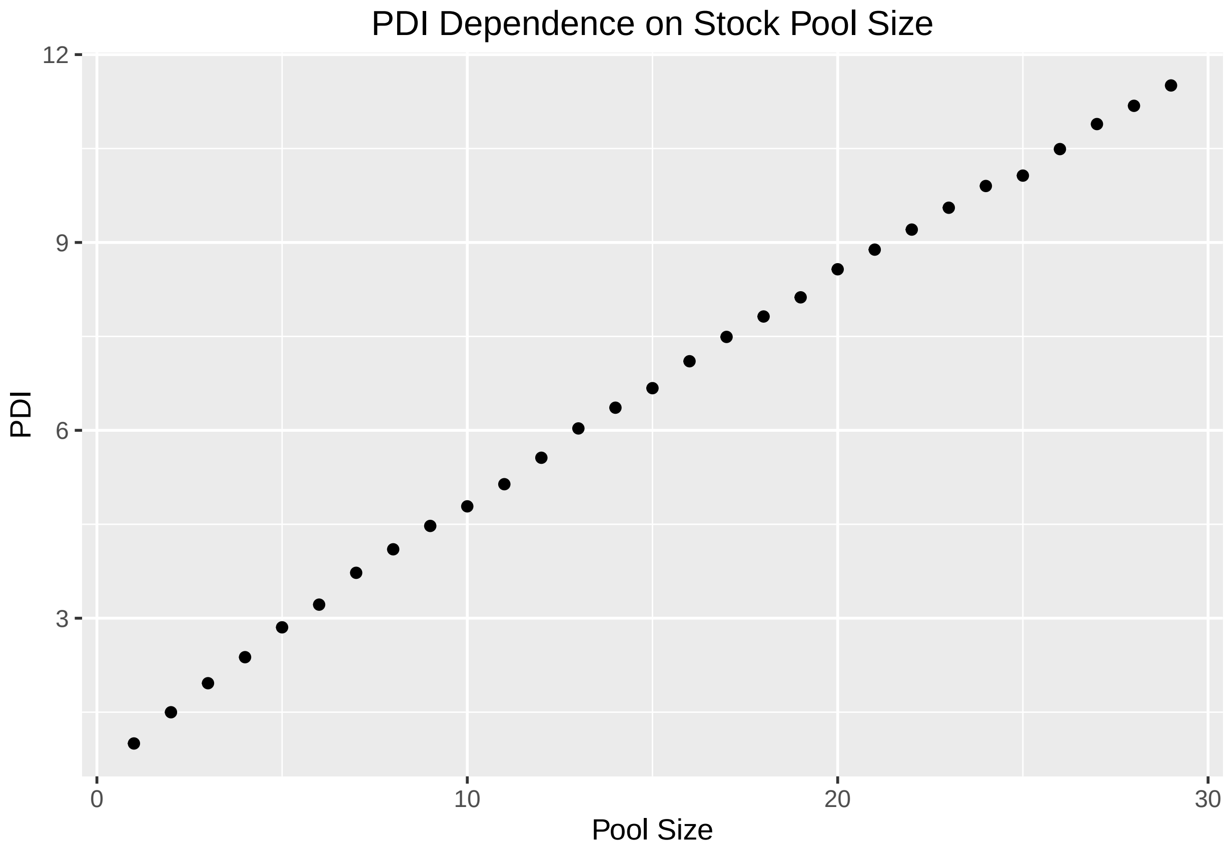
$$(2\sum_{k=1}^{N} kW_k) - 1$$

where 'k' is the kth principal component, and 'N' the number of securities/assets. This measures how front-loaded the vector of relative strengths is. That means that if PDI is close to 1, most of the variability in the data can be explained by the first few components - prices of the portfolio of assets vary in association. Alternatively, if PDI is approaches 'N', then the variability in the data is spread out. To summarize:

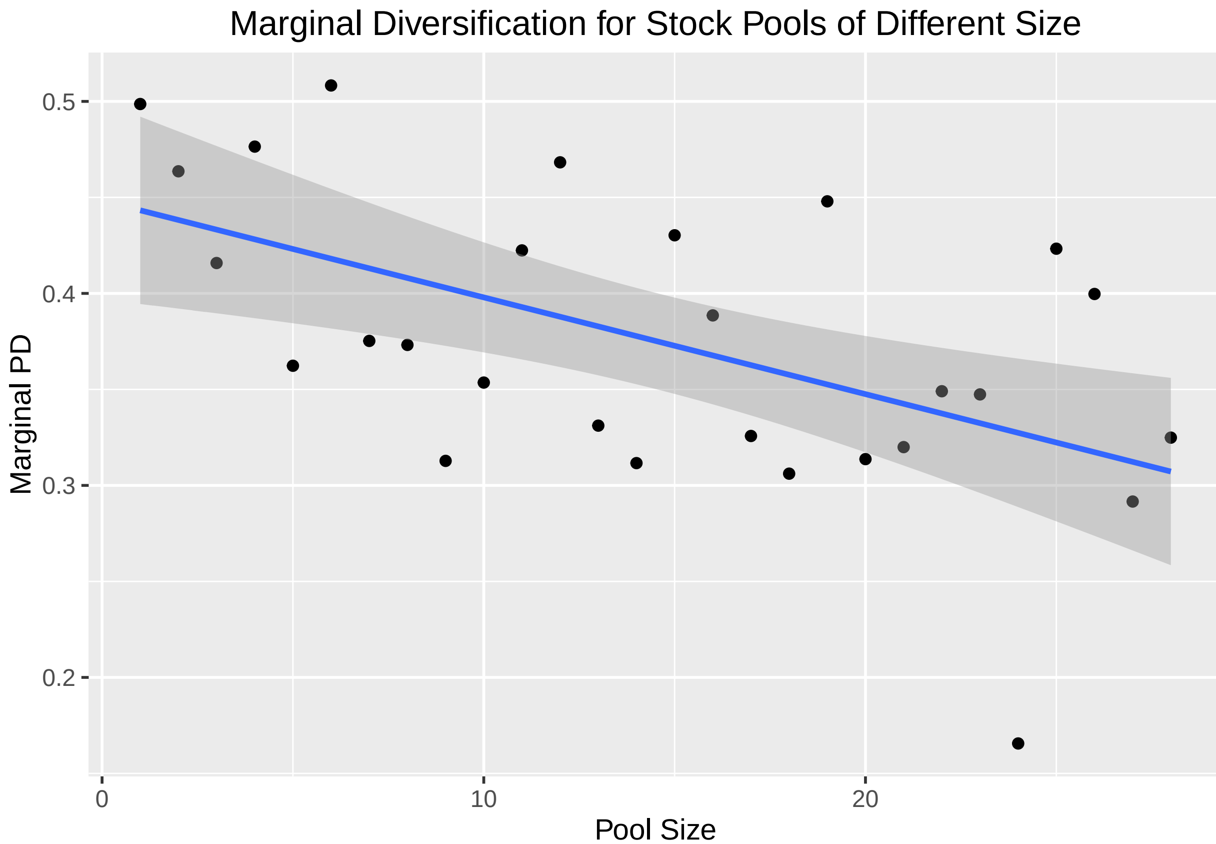
1. For a completely non-diversifited portfolio dominated by a single factor, PDI = 1.
2. An ideally diversifit portfolio will have PDI = N.
3. Any positive change in front-loading, when explanatory power moves from W_k with higher k to W_k with lower k, reduces the index.

PDI of the Dow Jones Index

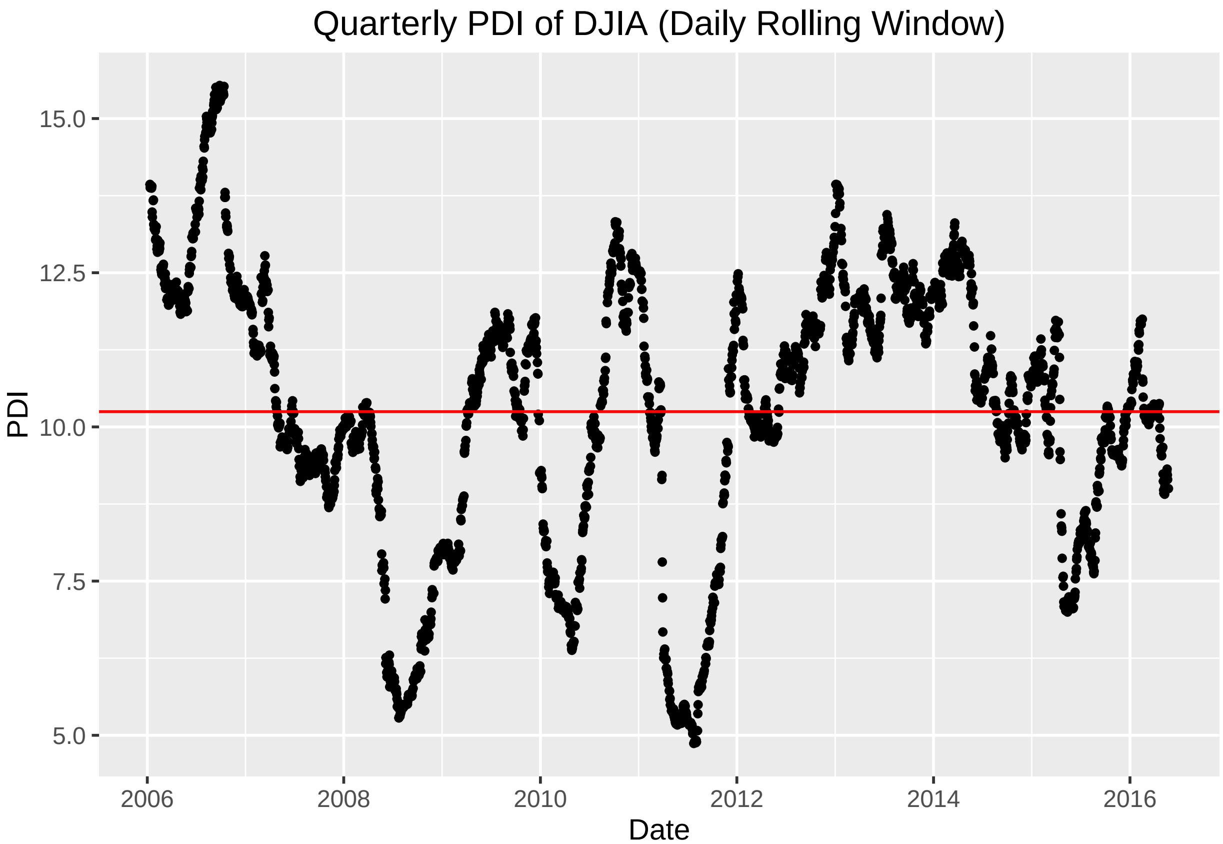
To illustrate the concept of PDI, we will examine the constituents of the Dow Jones Industrial Index. Of the 30 constituents, one was removed as its data does not cover the full sample period. Nevertheless, that should not detract from our illustrative purposes. Daily historical closing prices for the past 10 years is used. Returns are calculated and then normalized to have the same mean of 0. For a specified pool size K, we randomly draw K different stocks (20 trials) from our sample to determine the sample mean of PDIs for each pool size. It is assumed that each constituent takes an equal weight.



The above graph confirms that as more securities are added, the PDI increases as expected. Nevertheless, there is a good deal of correlation among the DJIA constituents. For example, a pool of 20 stocks is approximately equivalent to about 8 truly independent components. Ideally, the graph should also show, in theory, diminishing marginal returns to diversifcation. However, due to the small number of constituents of the index, the visual effect is not as pronounced.



The above graph plots the marginal portfolio diversification as pool size increases. Ideally, the points should roughly follow a concave function. A linear regression line is plotted to show the decreasing marginal increases to PDI for each addition of a security.



One might also be interested in how correlated the DJIA has been through time. The following graph plots the quarterly PDI with a daily rolling window. Clear dips can be seen, where a familiar one has been the 2008-09 crisis. The PDI peaked late 2006 and continued to fall to a bottom at the middle of 2008. The red horizontal line depicts the sample mean.

Other Applications of PCA

Kritzman et al (2010) applied a similar idea of the aforementioned PDI to the economy to measure how tightly-associated assets are as a proxy for systemic risks. It is an indication of underlying market fragility, although not necessarily market downturn. Apart from constructing diversification indices, PCA is also very applicable to classification problems. An example would be doing PCA on the various funds of a bank to sort them into distinct categories or onto a multi-dimensional spectrum. This is done by Commodity Trading Advisor funds, for instance.