An Application of Principal Component Analysis to Stock Portfolio Management

Libin Yang

Department of Economics and Finance

University of Canterbury

A thesis submitted in partial fulfilment of the requirement for the Degree of

Master of Commerce in Finance

January 2015

Acknowledgements

I would like to express the deepest appreciation to my thesis supervisors Dr. Bill Rea and Dr. Alethea Rea for the continuous support for my master study and research. Without their supervision and constant help this dissertation would not been possible. I would like to thank our librarian Cuiying Mu and Kikai Kobayashi, the University of Canterbury account manager from Sirca Ltd for helping me in data collection. In addition, a special thank to my father and mother for the unconditional love and all the sacrifices that they have made on my behalf. I would also like to thank my friend Karen who inspired me to strive towards my goal and spent sleepless nights with me when I was stuck in writing this thesis.

Abstract

This thesis investigates the application of principal component analysis to the Australian stock market using ASX200 index and its constituents from April 2000 to February 2014. The first ten principal components were retained to present the major risk sources in the stock market. We constructed portfolio based on each of the ten principal components and named these “principal portfolios". Principal portfolio one, which represents the market risk, is essentially a 1/N portfolio on the underlying stocks, and principal portfolio two has the highest price correlation with the ASX200 index. Rebalancing the positions based on the coefficients changes in component two did not bring better performance but more closely followed the ASX200 index. Moreover, allocation strategies applied to the principal portfolios instead of the underlying stocks substantially reduced the risk and would have avoided the significant drop in 2008 financial crisis. The high number principal components were used to identify near linearly correlated stocks and based on this idea we proposed a stock selection procedure that pick stocks according to the correlation structure. We found a portfolio of at most 25 stocks closely resemble the ASX200 index. It is not any combination of stocks can used to present the whole data set like other papers have implicitly suggested. The variance explained by principal component one was used as a measure the level of systemic risk. The market was more concentrated during crisis period and indicates less diversification benefits to exploit. The variance explained by the first principal component can serve as a leading indicator of financial crisis. The KMO measure of sampling adequacy closely followed the variance explained by first principal component and may also be used as a leading indicator of financial crisis.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Markowitz (1952)'s mean-variance theory is the foundation of modern portfolio theory. He introduced the first thorough proof in favor of diversifying a portfolio among a range assets instead of holding a single security alone. Investors had been aware of the benefits of diversification even before the Markowitz (1952)'s mean-variance theory. Lowenfeld (1909) discussed the benefits of diversification and sometimes is considered the first rigorous academic discussion of diversification.

There are some practical drawbacks associated with the mean-variance efficient optimisation. As the input parameters such as the forecast returns and risks are defined, the mean-variance efficient portfolio is determined. However, the allocations in the mean-variance efficient portfolio are very sensitive to the inputs. One would have to estimate or forecast the risk and return, which are usually associated with large estimation errors, especially the errors embedded in estimating returns. Chopra and William (1993) pointed out that the estimation error of expected return is about 10 times higher than the estimation error of variance and about 20 times that of the covariances. A small change in the input can result in a completely different asset allocation (Jorion, 1985). Michaud (1989) studied the limitation of the mean-variance approach and claimed the mean-variance optimizer was an "estimation-error maximizer". Moreover, the mean-variance approach tends to concentrate a portfolio on a few assets, which are the ones with the highest returns if historical data is used or the ones with the highest expected returns if forecasts are used (Bernstein, 2001). This is contrary to the intention of diversification. Allen (2010) argued that the mean-variance approach failed to diversify a portfolio in the 2008 financial crisis.

There is an increasing need from both academic researchers and market practitioners for ways of eligibly building more diversified portfolios. The new paradigm of portfolio allocation is the risk-based allocation strategy, which constructs a portfolio based solely on the variance-covariance of assets. Examples of risk-based allocation strategies are minimum variance (Behr et al., 2008; Clarke et al., 2006; Haugen and Baker, 1991), most diversified portfolio (Choueifaty and Coignard, 2008), risk parity (also known as the equal risk contribution) (Maillard et al., 2010; Qian, 2006) and diversified risk parity (Kind, 2013; Lohre et al., 2012, 2014) [1].

Two salient features of security investment are the uncertainty of security returns and the correlations between security returns. The correlations between securities, particularly low or negative correlations, make diversification possible but it is also the reason the analysis of security investment is complicated. If there are only a few securities involved, for example five stocks, then it is easy and intuitive to simply look at the five variances and 10 correlations or covariances. However, if the number of securities is large, for example 200, it will not be very helpful to simply look at the 200 variance and 19900 correlations or covariances. Principal component analysis (hereafter, PCA) is a statistical method of dimension reduction. It provides us with an alternative approach that is able to reduce the complexity - look for a few derived principal components that are designed to be uncorrelated while preserving most of the variation given by those variances and correlations or covariances (Jolliffe, 1986). Although it has been widely used in many areas, until recently the studies applying PCA to the finance sector were scarce, especially in the context of portfolio management.

Our research focuses on the potential use of PCA in portfolio management. For simplicity, we study a portfolio only with stocks. There are two ways of applying PCA in constructing portfolios. The PCA can reduce the complexity of a stock portfolio by transforming the stocks into a new set of uncorrelated principal components that represent uncor-

[1] The diversified risk parity is an allocation strategy based on principal component analysis. It relates to our research on applying principal component analysis in constructing portfolios, which we will discuss in more detail in Chapter 3

related risk sources. Then, instead of constructing portfolio based on the underlying stocks, one can treat the principal components as individual investment assets and simply choose from them. Hence, the analysis required for portfolio selection is therefore simplified.

The other approach is using a selected subset of stocks to replace the original full data set. This relates to an old question that has been researched for decades - how many stocks are needed to diversify a portfolio? The traditional capital-asset pricing model (CAPM) required the purchase of a market portfolio that contains all risky assets, but essentially the market portfolio can be achieved with a much smaller portfolio. PCA allows us to identify the stocks that can be used as a representative of the whole data set and therefore find the number of stocks that is sufficient to diversify a portfolio.

Beside the use of PCA in portfolio construction, the most important application of PCA in portfolio management is to measure market concentration and the potential for diversification. While there are literatures which applied PCA to research market connectedness and argued that the diversification effects declined when the market was more concentrated, to the best of our knowledge no papers actually compare the degree of diversification to the market connectedness. Moreover, to the best of our knowledge, no papers pointed out that it is crucial to use the correlation matrix rather than the covariance matrix as input to a PCA when studying market connectedness. We research the problems associated with the use of covariance matrix and emphasize the importance of using the correlation matrix in the study.

In many applications of PCA in finance, the high numbered principal components that explained little variance were normally discarded. No useful role was assigned to those principal components. But we found those components are effective in identifying near linearly correlated stocks.

Our research is organized as below: Chapter 2 presents a brief description of principal component analysis. Chapter 3 gives a literature review of relevant topics. Chapter 4 describes the data and methods. Chapter 5 determines the number of principal components required to represent the major risk sources in the stock market. Chapter 6 investigates the use of the high numbered principal components in identifying near linearly correlated stocks. Chapter 7 constructs portfolios based on the principal components retained in Chapter 5 and compares the equally weighted and equal risk contribution allocation strategies. Chapter 8 uses the principal component one to study the evolution of market connectedness over time. Chapter 9 investigates the time evolution of the relative importance of each stock in the principal components retained in Chapter 5. Chapter 10 determines the number of stocks are needed to make a diversified portfolio. Chapter 11 summaries the thesis and discusses potential further research.

Chapter 2

Principal component analysis: a brief review

Principal Component Analysis (PCA) is a statistical method of dimension reduction that is used to reduce the complexity of a data set while minimizing information loss. It transforms a data set in which there are a large number of interrelated variables into a new set of uncorrelated variables, the principal components, and which are ordered sequentially with the first component explaining as much of the variation as it can. Each principal component is a linear combination of the original variables in which the coefficients indicate the relative importance of the variable in the component.

Consider an unrealistic, but simple, case of only two variables and the data are plotted in two dimensions. In Figure 2.1, we display the geometrical interpretation of the two variable principal component analysis. The x_1 and x_2 represent the original variables and axes. The PC_1 and PC_2 are the transformed variables and axes. The direction of the principal axes indicates the principal components. The first principal component looks for a linear combination $\alpha_1'\mathbf{x}$ that explains the most variation, which is

$$\alpha_1'\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 \tag{2.1}$$

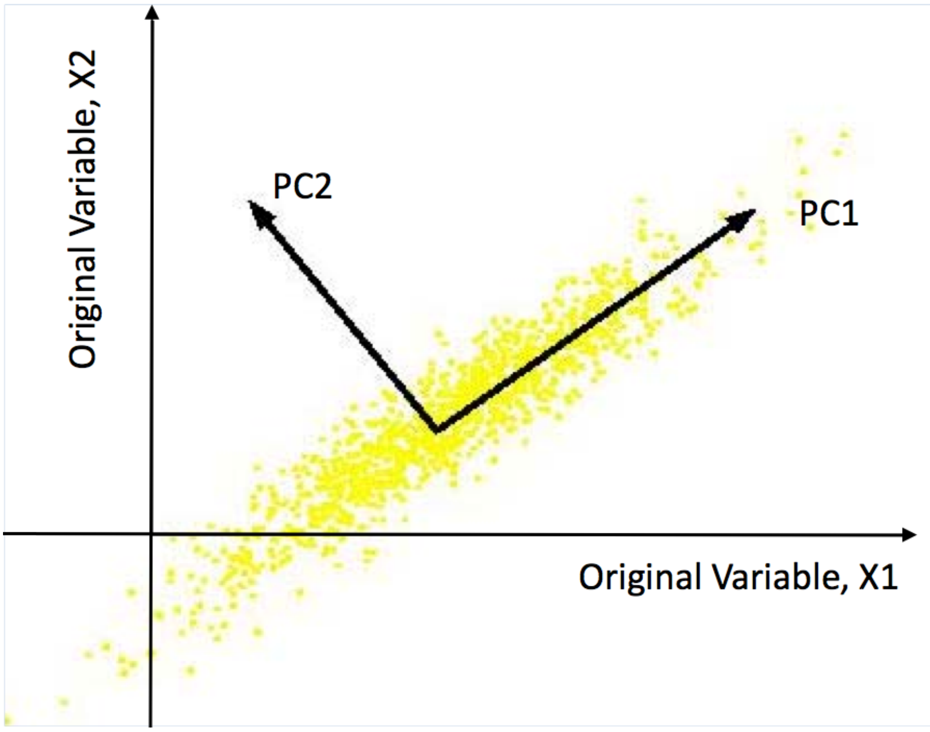
where $\alpha_1'\mathbf{x}$ is the eigenvalue of principle component one, x_1 and x_2 are the two original variables, α_{1i} is the coefficient of stock i in component one. Next, the principal component two looks for a linear combination $\alpha_2'\mathbf{x}$ that explains as much of the remaining variation conditional on it being orthogonal to principal component one. In a two variable case, this is the remaining variation, which is

$$\alpha_2'\mathbf{x} = \alpha_{21}x_1 + \alpha_{22}x_2 \tag{2.2}$$

Clearly, in Figure 2.1 there is more variation in the direction of principal component one than either of the original variables, and leaves very little variation in the direction of principal component two.

PCA has often been treated as one special case of the factor analysis in textbooks. But Jolliffe (1986) stated that "this view is misguided since PCA and factor analysis, as usually defined, are really quite distinct techniques". The PCA seeks to explain the diagonal terms of a covariance matrix or correlation matrix and also does a good job of explaining the off-diagonal terms. The factor analysis, on the other hand, seeks to explain the off-diagonals terms but also explains the diagonal terms well. Moreover, both techniques aim to reduce the dimensionality of a data set (e.g. from a total of p variables) to a much smaller dimension $m < p$. Changing m, the dimensionality of the model, can have much more dramatic effects on factor analysis than it does on PCA. A final difference of these two techniques is the principal component can be calculated exactly from the original variable, but the factors in factor analysis cannot. Here we only provide a brief discussion of the principal component analysis just to help shape the basic understanding of this technique, for more detailed information, refer to Jolliffe (1986).

Fig. 2.1 Geometric interpretation of principal component analysis (Graphic source http://www.cheric.org/ippage/d/ipdata/2014/01/file/d201401-401.pdf).



Chapter 3

Literature Review

PCA is one of the best-known techniques in multivariate analysis. Its range of applications has expanded with the advent of computers and it has been used in a wide variety of areas for the last 50 years (Jolliffe, 1986). The ability of PCA to decompose interrelated variables into uncorrelated components makes it attractive to use in analyzing the complex structure of financial markets. It has been applied to produce market indices (Feeney and Hester, 1967) and to identify common factors in bond returns (Driesson et al., 2003; Pérignon et al., 2007). In more recent years, a growing literature applied PCA to the study of market cross-correlation and systemic risk measurement (Billioand et al., 2012; Kritzman et al., 2011; Zheng et al., 2012). Most works have only discussed the theoretical framework of applying PCA in portfolio management, few have actually looked into its performance. Our research focuses on the practical application of PCA to portfolio management. To the best of our knowledge, no similar work has been done based on the Australian market.

3.1 Market connectedness and systemic risk

Systemic risk is “the risk associated with the whole financial system, as opposed to any individual entity or component. It can be defined as any set of circumstance that threatens the stability of financial system, and so potentially initiates financial crisis" (Zheng et al., 2012). The systemic risk is often misunderstood as the systematic risk. But they are in fact different concepts. Systemic risk indicates the ratio of systematic risk to idiosyncratic risk. An increase in the systemic risk suggests the proportion of systematic risk in the total risk increases. This means the amount of idiosyncratic risk, which is the diversifiable risk, decreases. Conversely, an increase in systematic risk does not necessarily indicate a decrease in diversifiable risk. Consider when the amount systematic risk increases and the idiosyncratic risk also increases while the ratio between them stays the same. So the systemic risk was unchanged. The amount of diversifiable risk in this case increases rather than decreases.

After the financial crisis in 2008, literature relating to systemic risk is substantial. There have been three groups of empirical studies on systemic risk. One of them focused on contagion, spillover effects and joint crashes in financial markets (Adrian, 2007; Billioand et al., 2012; Kritzman et al., 2011; Wang et al., 2011). Those studies were based on the analysis of interconnectedness among market security returns and our analysis on systemic risk follows in their steps. The other two groups of empirical studies on systemic risk include research on the auto-correlation in the number of bank defaults, bank returns, and fund withdrawals (Brandt and Hartmann, 2000; Kenett et al., 2012; Lehar, 2005) and research on bank capital ratios and bank liabilities (Aguirre and Saidi, 2004; Bahansali et al., 2008; Brana and Lahet, 2009) respectively. When the market becomes more connected, the systemic risk is higher in the sense that the negative shocks propagate more quickly and broadly. For this reason, monitoring the time evolution of correlation is critical. Moreover, low correlation between assets is what makes diversification possible, gaining insight into co-movement of securities is important in portfolio management.

Other research has shown that the security correlations change in different time periods. Butler and Joaquin (2001) and Campbell et al. (2002) reported that the market correlation increased in bear markets. Ferreira and Gama (2004), Hong et al. (2007), and Cappiello et al. (2006) reached the same conclusions for global industry returns, individual stock returns and international bond returns.

Instead of comparing different time periods, many recent papers have applied PCA to investigate correlation using a sliding window approach. Fenn et al. (2011) applied PCA to study the evolution of correlation in a diverse range of asset classes: 25 developed market equity indices, three emerging market indices, four corporate bond indices, 20 government bond indices, 15 currencies, nine metals, four fuel commodities, and 18 other commodities. They asserted that increases in the variance explained by the first component implied that there was more common variation in financial market. Moreover, they emphasized that the variance explained by the first component might be the result of increases in the correlations among a few assets or a market-wide correlation. The first case will have less impact on the diversification because one can simply move investments to less correlated assets. In contrast, it becomes much more difficult to reduce risk by diversifying across different assets if it is a market-wide correlation increase. They reported that the variance explained by the first principal component sharply increased when Lehman Brother filed for bankruptcy and Merrill Lynch agreed to be taken over by the Bank of America on 15 Sep 2008 was the case of a market-wide correlation increase.

Kritzman et al. (2011) introduced a measure of systemic risk called the absorption ratio. It is the fraction of variance absorbed by a finite number of principal components. They reported that most global financial crises were coincident with positive shifts of the absorption ratio. These crises include the Asian Financial Crisis in 1997, Russian default and LTCM collapse in 1998, Housing bubble in mid-2006, and Lehman Brothers default in 2008. Another interesting finding in this paper is, in most cases, stock prices changed significantly when the absorption ratio reached its highest or lowest level.

Zheng et al. (2012) not only looked at the absolute value of variance explained by the first component, they also computed the change in the variance explained to capture the systemic risk. They obtained similar findings to Kritzman et al. (2011) that both the absolute value and change of variance explained by principal component one increased during a financial crisis. But they reported that the moving window size and the time length used to calculate the change had an impact on the date of the spike. The spike of absolute value of variance explained by principal component one occurred later when the moving window size was larger and saturated after approximately 20-month time window. When the length of time used to calculate the change was longer, the change became delayed. While many others choose the size of moving window randomly, Zheng et al. (2012) pointed out that the size can impact the result. We believe that what Zheng et al. (2012) suggested is just a matter of sampling adequacy. One should not apply PCA to a data set that does not have enough data points.

3.2 How many stocks make a diversified portfolio?

Markowitz (1952)'s argued that instead of looking at single security alone, one should be concerned with portfolios as a whole. The reason for this is that including securities that have low correlations or even negative correlations could eliminate some risk. The existence of many kinds of index funds provide means for investors to hold a diversified portfolio, but is it necessary to include all constituents within a index fund to obtain the same diversification as the index itself? Conventional wisdom has it that the benefits of diversification are virtually exhausted when a portfolio contains a high enough number of stocks. However, how many stocks are enough remains an open question.

Evans and Archer (1968) reported that approximately 10 randomly chosen stocks would be adequate to diversify a portfolio. They observed that the benefit of diversification decreased as the number of stocks increased. Their conclusion has been cited in many textbooks (Francis, 1986; Gup, 1983; Reilly, 1985; Stevenson and Jennings, 1984). Newbould and Poon (1993, 1996) followed Evans and Archer (1968)'s approach of comparing increasing size portfolio variance and claimed that just 8 to 20 stocks was enough to fully obtain the benefit of diversification. However, Statman (1987) compared the cost and benefit of diversification and reported that the number of randomly chosen stocks that make a well diversified portfolio was at least 30 using the data available in mid-1980s. Statman (2004) used the same approach and more available data, he then concluded that the break even point, where the marginal benefit was equal to the marginal cost, exceeded 300 stocks.

Subsequently, many others have reported different numbers of stocks needed to diversify a portfolio using risk measurements other than variance. Domian et al. (2003) reported that in order to avoid a significant shortfall risk, no less than 60 randomly chosen stocks were required. According to Domian et al. (2007), shortfall risk reduction continues as the number of randomly chosen stocks was increased, even above 100 stocks.

The above researchers were using the number of stocks in a portfolio as a measure of diversification. However, this was problematic. If, in an ideal world, all stocks had same mean, variance and covariance, the number of stocks in a portfolio would be the key variable for estimating the reduction in variance (Frahm and Wiechers, 2011). In reality, such assumptions do not hold. Intuitively, randomly choosing stocks to add to a portfolio, even when they reach the number required, may not result in the promised diversification if the randomly chosen stocks were more highly correlated than expected. The use of PCA deals with the problem associated with randomly choosing stocks. In Chapter 10 we propose a stock selection method that picks stocks based on their correlation structure. The selected stocks are used to describe the original data set and represent the risk sources inherent in the data set. Rudin and Morgan (2006) applied PCA to measure diversification quantitatively and tested equal-weighted portfolios of stocks in the S&P100 index and reported that a pool of 40 randomly selected stocks is approximately as diversified as only 20 truly independent components. PCA provides us with a way to identify uncorrelated risk sources in the market and pick stocks from those different risk sources, the resulting portfolio size is more meaningful from the point of view of diversification. In addition, as we discussed in Section 3.1, the market connectedness does not stay constant over time. Markets become more tightly coupled in volatile periods and even a portfolio with same stocks would be less diversified (Billioand et al., 2012; Fenn et al., 2011; Kritzman et al., 2011; Zheng et al., 2012). Campbell et al. (2001) did point out the number of stocks needed to achieve a certain level of diversification was not the same in the 1963-85 period and the 1986-97 period.

3.3 PCA in portfolio construction

Partovi and Caputo (2004) first proposed the idea of using PCA to analyse the efficient portfolio problem. Their basic idea was based on the fact that if there were no correlations among assets, the complexity in portfolio selection dramatically decreased. They reported that if short sales were allowed, any asset set could be transformed into a set of uncorrelated "principal portfolios". After transforming to a principal portfolio environment, they constructed an efficient frontier based on principal portfolios. Based on the theoretical interpretation of such a transformation, they stated "return-volatility structure of the efficient frontier is more simply related to the principal portfolio environment than the original asset set".

Partovi and Caputo (2004)'s theoretical framework of constructing principal portfolios has inspired many academic researchers as well as market practitioners in portfolio management. Especially after the global financial crisis in 2008, reducing risk became the priority. Meucci (2009) followed the idea of Partovi and Caputo (2004) by transforming a data set in which there were a large number of interrelated assets into a new set of principal portfolios representing uncorrelated risk sources inherent in the original assets. He reported a so-called diversification distribution, a tool to analyze the structure of a portfolio's concentration profile. The diversification distribution is expressed as the ratio of each principal portfolio's variance to its total variance. In the principal portfolio environment, all the principal portfolios are uncorrelated and therefore the variances are additive. The total variance is simply the sum of the variances of all principal portfolios. The ratio of individual principal portfolio variance to the total variance is then in the range of 0 to 1 and sum to 1. These properties are similar to a probability, a non-negative value and always sum to 1. The maximum diversification is achieved where the diversification distribution is close to uniform. In other words, a well-diversified portfolio is the one in which the risk drivers are invested into equally. Meucci (2009) also introduced a diversification index that represented the effective number of uncorrelated bets in a portfolio. The idea is that if the number of uncorrelated bets were small, the risks were rather concentrated in few sources and less diversified.

Lohre et al. (2012) and Lohre et al. (2014) adopted the framework in Meucci (2009)'s paper in order to seek maximum diversification in equity and multi-asset classes respectively. The idea was to evenly distribute investment funds across principal portfolios to well diversify its overall risk. They named such strategy "diversified risk parity". They investigated the diversified risk parity as well as the other three allocation strategies:

- 1/N
- Minimum variance
- Risk parity

Lohre et al. (2012) and Lohre et al. (2014) reported that the diversified risk parity provided better risk-adjusted performance and was the most diversified among the investigated alternatives.

The central idea of risk parity, also known as the equal risk contribution strategy, is achieving an equal risk contribution of each asset in the portfolio. Qian (2011) claimed that risk parity allocation results in better diversification and brings higher returns. He also emphasized that if one believed in the benefit of diversification then one should believe in risk parity. Other researchers such as Qian (2006) and Maillard et al. (2010) also advocate a risk parity strategy.

One similarity of the risk parity and diversified risk parity allocation strategy is they both construct portfolios by allocating risks. But the difference is while the risk parity allocates a risk budget based on underlying assets, the diversified risk parity allocates risk based on uncorrelated principal portfolios. The risk parity allocation strategy can have rather concentrated risk if most assets have high correlations with each other. The diversification of a risk parity strategy is sensitive to the correlation between the assets included. Consider an extreme case with all stocks perfectly positive correlated, then allocating equal risk budget to all stocks is actually the same as holding one stock. The diversified risk parity, on other hand, does not have such a problem since it is allocating its risk budget based on uncorrelated risk sources. Though in the case of perfect correlations all variation would be explained by the first principal component.

Kind (2013) compared risk parity and diversified risk parity with the 1/N on eight generic futures contracts that covered four asset classes. The back test[1] was based on monthly US-Dollar returns over the period February 1990 to January 2013. They stated that no strategy was able to outperform the simple equal weight allocation strategy. DeMiguel

[1]Back-testing seeks to estimate the performance of a strategy if it had been employed during a past period.

et al. (2009) and Lee (2011) also reported the simple 1/N strategies dominated the other allocation strategies available for portfolio selection.

When a data set is transformed into a new set of uncorrelated principal portfolios, all the allocation strategies that can be applied to the underlying assets can also be applied to the principal portfolios. The risk parity and diversified risk parity are good examples of same allocation strategy based on original underlying assets and the transformed principal portfolios respectively. Lohre et al. (2012) and Lohre et al. (2014) reported that the diversified risk parity, which is based on the transformed principal portfolios, was a better strategy than the risk parity that was based on the untransformed data set. We include only two allocation strategies in our research. We applied these two strategies to both the original underlying stocks and the transformed principal portfolios. One of the allocation strategies is the equal risk contribution. When it is applied to the stocks, it is known as the risk parity and when it is applied to the transformed principal portfolios, it is known as the diversified risk parity. The other strategy we study is the 1/N, the simplest strategy and one advocated by many papers (DeMiguel et al., 2009; Kind, 2013; Lee, 2011).

3.4 Research questions

The major research questions of this thesis are:

1. How has the systemic risk of Australian market changed over time?
2. Can principal component analysis be used as leading indicator of financial crisis?
3. How many stocks are needed to diversify a portfolio? Which stocks are they?
4. Can an asset allocation strategy based on PCA provide better performance and diversification?
5. Can the change in the coefficients signal a portfolio manger to trade?

While the research questions are posed in the order based the discussion of the literature review, the design of this thesis differs from this order (see Chapter 1).

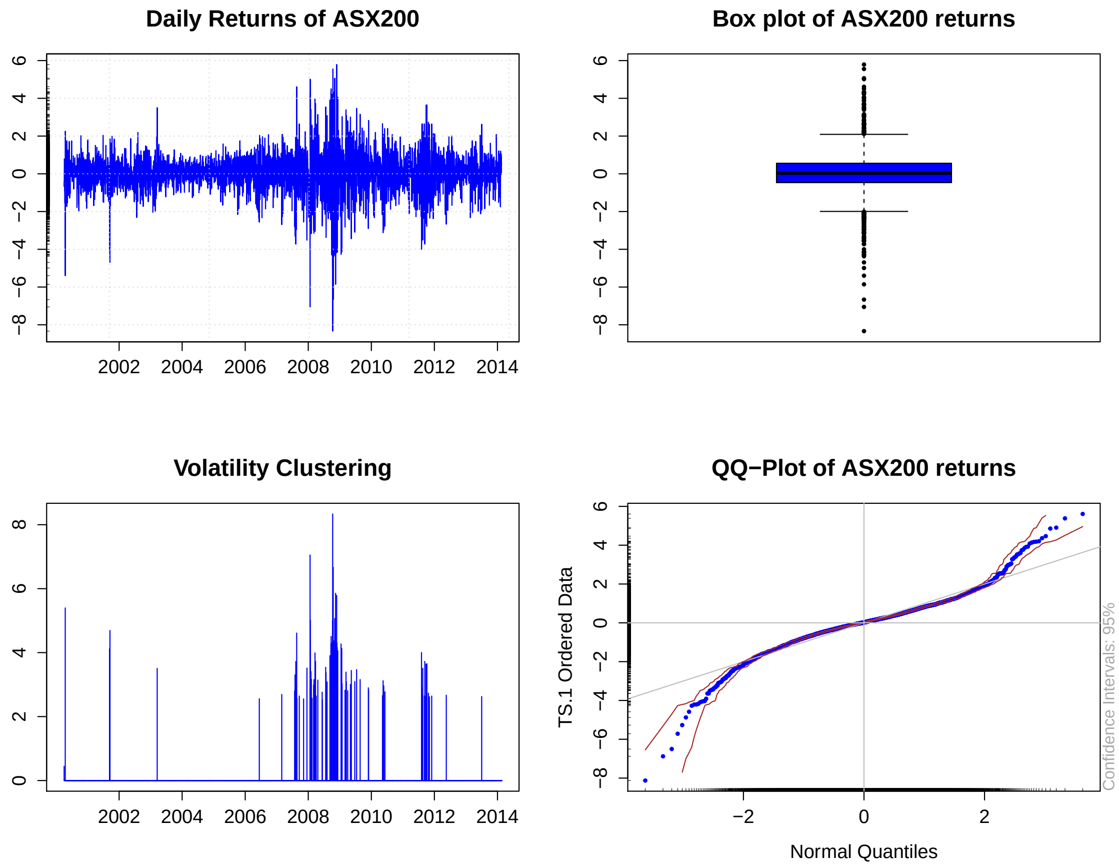
Chapter 4

Data and Methods

4.1 Data and Descriptive Statistics

Our research is based on the Australian market. We investigated the constituents of the ASX200 index from April 2000 to February 2014. The ASX200 index is a market capitalization weighted index of the 200 largest shares by capitalization listed on the Australian Securities Exchange, which starts from 31 March 2000. We note that the ASX200 index returns does not adjust for the dividends but the returns we calculated later for all the constituents were adjusted for dividends paid. In Fig. 4.1, we investigated the characteristics of the ASX200 return data and the time series plot of the ASX200 percentage return, a box plot as well as the 100 largest absolute returns and a Quantile-Quantile plot compared to the normal distribution were produced and are exhibited. We found evidence of volatility clustering: "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963). From the two plots in the left-hand panel, we observed most large absolute returns occurred during the 2008 financial crisis. The ASX200 index level continued to change significantly until the end of 2009. There was also a cluster of large returns at the end of 2011, this is when the Australian stock market was affected as investors responded to America's credit downgrade, the European sovereign debt crisis, and fears over the global economy. Moreover, in the box plot and QQ plot, the ASX200 daily returns are skewed to the left and the heavy tails are evident.

Fig. 4.1 Stylized facts for ASX200.



We further calculated the skewness and kurtosis of the ASX200 daily returns. There were -0.383 and 5.657 respectively, clearly indicating a heavy tail.

There was a high frequency of stocks that were added to or deleted from the index from time to time, so we identified all stocks which had been in the ASX200 for the whole study period. After adjusting for mergers, acquisitions, and name changes we obtained a final data set of 524 unique stocks. We obtained daily closing prices and dividends for each stock from the SIRCA database[1]. All the prices and dividends were adjusted to be based on the AUD. For a more detailed description of data, see Appendix A. The PCA was performed on the correlation matrix of the return series and in which we assumed the dividends were reinvested into stocks which issued them when calculating the return. The

[1]http://www.sirca.org.au/

return was calculated in the following steps:

1. We created a new column in the spreadsheet named Dividend Factor. Then we started with a factor of 1 and every time a dividend was paid we multiplied the Dividend Factor,

$$\text{Daily Dividend Factor}_i(t) = \begin{cases} 1 & \text{if no dividend} \\ 1 + \frac{D_i(t)}{P_i(t)} & \text{if dividend} \end{cases}$$

$$\text{Cumulative Dividend Factor}_i(t) = \prod_{j=1}^{t} (\text{Daily Dividend Factor}_i(t)) \tag{4.1}$$

where $D_i(t)$ is the dividend for stock i in time t, $P_i(t)$ is price for stock i at time t and t is in units of one trading day.

2. We adjusted the price series with the dividend factor, the adjusted price was calculated by

$$\text{PNEW}_i(t) = P_i(t) \times \text{Cumulative Dividend Factor}_i(t). \tag{4.2}$$

3. The return series for a given stock i is

$$\text{R}_i(t) = \frac{\text{PNEW}_i(t+1) - \text{PNEW}_i(t)}{\text{PNEW}_i(t)}. \tag{4.3}$$

4.2 Methods

For most parts of our research, a rolling window approach was applied. We extracted a set of stocks that had complete return information for the whole study period, and there were 156 such stocks. The remaining 368 stocks were either listed after April 2000 or delisted before February 2014.

We used the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970; Kaiser and Rice, 1974) to test the shortest length of sliding window that a PCA could be efficiently applied to. The KMO statistic compares the value of correlations between stocks to those of the partial correlations. If the investigated stocks share more common variation, the KMO will be close to 1. On the other hand, a KMO near 0 indicates the PCA will not extract much useful information.

We calculated the KMO statistic in rolling windows of different size for the 156 stocks that have complete data. A window size one year (252 trading days) had KMO statistics in the range from 0.36 to 0.90 for the whole study period and a quarter of them had KMO values below 0.5, which is the smallest KMO value that is considered acceptable to do a PCA. On the other hand, a window of two year (504 trading days) had better KMO statistics, with a lowest of 0.62 and highest of 0.95 during the whole study period. So we decided to apply PCA in rolling window approach with window size two years.

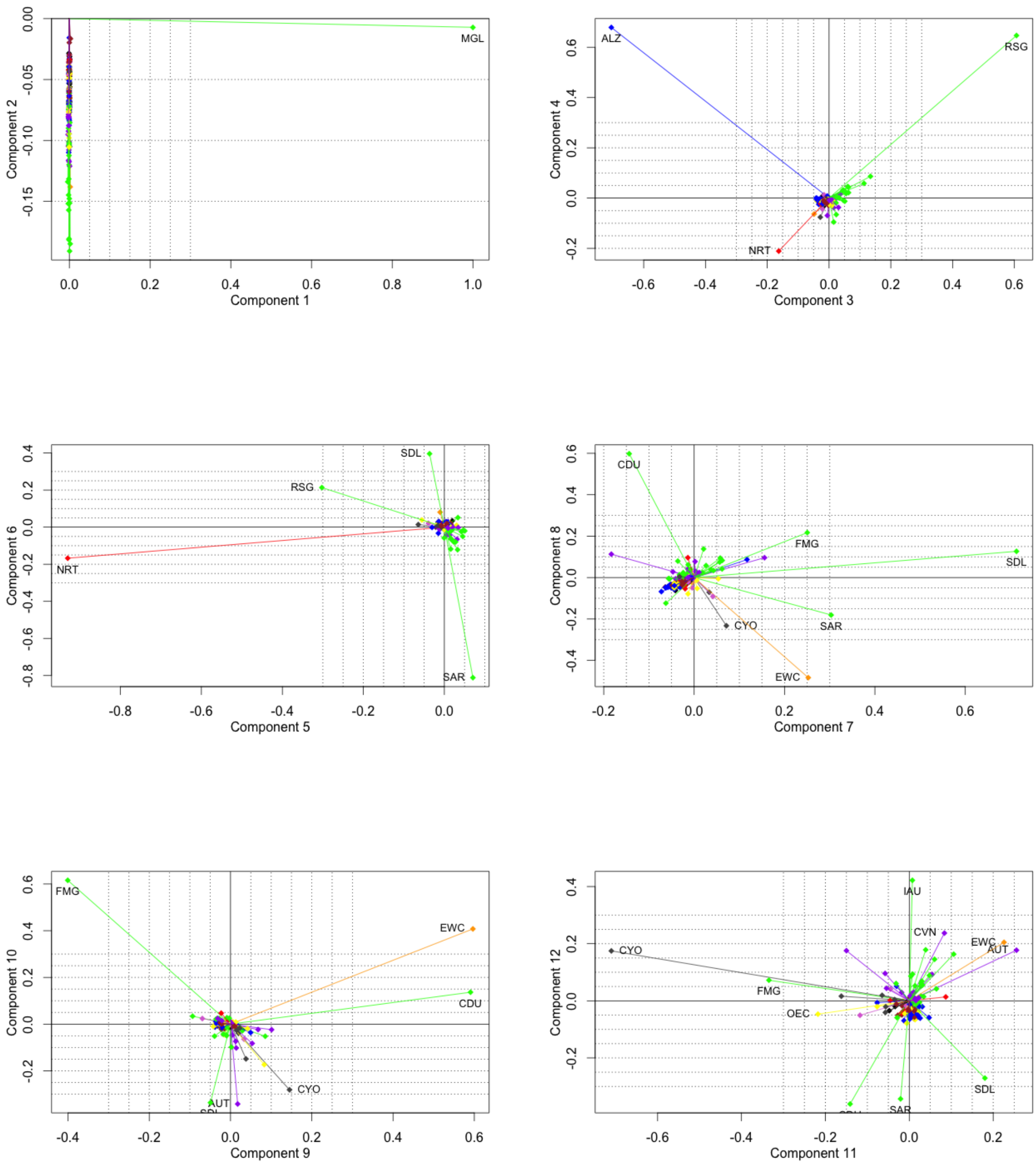
The KMO statistic can also be used to assess the potential for diversification. The KMO statistic measures the degree of common variation among stocks. One can expect less diversification opportunities when stocks have more common variation, which means a high KMO test statistic. We will extend this discussion in Chapter 8, when we examine the market connectedness and systemic risk.

PCA can be applied to either a correlation matrix or a covariance matrix but there are some problems associated with using a covariance matrix. If there are large differences between the variances of variables, then using a covariance matrix will result in low numbered principal components being dominated by variables that have a large variance. This will impede getting useful information for diversification from a PCA in some cases (Jolliffe, 1986).

Before we discuss how using a covariance matrix will impede getting useful information from a PCA, we first look at the principal components arising from a PCA on a covariance matrix of 156 stocks using the full study period. Figure 4.2 presents bi-plots of the coefficients of each stock in principal components 1 to 12 using the Industry Classification Benchmark (ICB). We can see in principal component one, while MGL had a coefficient of 1, all other stocks had zero coefficients. The reason that MGL dominated the first principal component is it had the largest variation and was significantly larger than the others. Appendix B lists the standard deviations of daily returns of the 156 stocks. MGL had a standard deviation of 39.79% while the second largest was 7.83%. In principal component two, all stocks had small negative coefficients and most of them were in the range of 0 to 0.15. We found this interesting because the coefficient structure was similar to the principal component one arising from a correlation matrix (see Figure 5.3), which has been understood as the market component with roughly equal contribution from all stocks (Fenn et al., 2011; Kim and Jeong, 2005; Zheng et al., 2012). However, it is normally the first principal component, which explains the most variation, which is understood as the market component. When using a covariance matrix, the first principal component was dominated by a single large variance stock and so was not necessarily the market component. In Chapter 8, we use the variance explained by principal component one as a measure of systemic risk. Using a covariance matrix to do a PCA will be misleading. Furthermore, principal components three and four picked up the second and third riskiest stocks, ALZ and RSG. Principal component five was dominated by the stock with the fourth largest standard deviation, NRT. Obviously, the principal components were dominated by stocks ordered by standard deviation. Using a covariance matrix in a PCA may only picks stocks with large variances and does not illustrate the correlation structure within the stocks, as it did with the ASX200. In Figure 5.3, we produced the same bi-plots of the coefficients in principal components 1 to 12 arising from a PCA on a correlation matrix using the same data set and study period [2]. Figure 5.3 shows there are structures in the coefficients that were not seen in the bi-plots using a covariance matrix. This clearly demonstrates the point that using a covariance matrix may not result in obtaining useful information for diversification from the principal components. So we decided to use the correlation matrix to do the PCA in our research.

[2] The discussions of the bi-plots for principal components 1 to 12 arising from a PCA on a correlation matrix are in Chapter 5.

Fig. 4.2 Bi-plots of relative weights of each stock in components 1 to 12 arising from a PCA on a covariance matrix from the whole study period, April 2000 to February 2014, using the Industry Classification Benchmark (ICB). The colors correspond to respective ICB sector classifications: Financials are Blue (33 stocks), Health Care are Red (9 stocks), Industrials are Yellow (24 stocks), Consumer Services are Brown (19 stocks), Basic Materials are Green (31 stocks), Oil&Gas are Purple (16 stocks), Utilities are Orange (5 stocks), Consumer Goods are Black (9 stocks), Telecommunications are Orchid (4 stocks), Technology are Grey (6 stocks).



Chapter 5

How many components should be retained?

A data set which contains p variables can be transformed into a new set of p principal components using PCA, in which each principal component is a linear combination of all the original variables. The original large number of variables can be replaced by a much smaller set of principal components that explain most of the variation if the KMO statistic is large [1]. When we apply PCA to a set of stocks, the principal components can be interpreted as uncorrelated risk sources inherent in the original data set. For instance, a portfolio of 156 stocks contains 156 uncorrelated risk sources. The eigenvalues of principal components typically decrease quickly and the higher numbered principal components have relatively small eigenvalues. This suggests the relevance of principal components quickly drops off. In theory, one can construct portfolios based on the principal components to get an exposure to all the risk sources. But it seems unreasonable to allocate any risk budget to the higher principal components that are not major risk sources.

Kim and Jeong (2005) decomposed the correlation matrix into three parts based on the Spectral Decomposition Theorem [2]. The Spectral Decomposition Theorem is about decomposing a correlation or covariance matrix into the principal components and their eigenval-

[1] Larger number of KMO statistic indicates there is more common variation in the data set. This means more variation will be explained by the first few principal components.

[2] More information of Spectral Decomposition Theorem, see Jolliffe (1986, p13).

ues. One can reconstruct an approximation of the matrix using the principal components that have been obtained by the decomposition. Including more principal components will always result in better approximation of the matrix. In the case of including all the principal components, one will get back the matrix started with. Kim and Jeong (2005) decomposed the correlation matrix and posited that the principal components are organized by the market part with the largest eigenvalue, the group part of intermediate discrete eigenvalues, and the random part of small bulk eigenvalues. They argued that this refers to the three kinds of fluctuation in the stock price. We followed Kim and Jeong (2005) and break the principal components into three parts that correspond to the three kinds of fluctuation of stock price changes:

1. The first principal component with the largest eigenvalue represents a market wide effect that influences all stocks.

2. A number of principal components following the market component represent synchronized fluctuations that only happens to a group of stocks.

3. The remaining principal components indicate randomness in the fluctuations.

In this chapter, we aim to determine the threshold for cutting off the random part of stock price fluctuation and preserve the major risk sources in the data set. Note that Kim and Jeong (2005) considered the random part of principal components as the “random noise" which contained no useful information. But we found that the last few principal components successfully identify stocks with near linear relationships and which have important financial implications in portfolio management (see Chapter 6).

There is not a fixed number of principal components to retain. One should vary this number based on the different circumstances (see below). Many rules can be applied to determine the number of principal components to retain. The best way will be a combination of them (Jolliffe, 1986). The data set we used in this section was the 156 stocks with complete price and dividend information for the whole sample period. We applied PCA to the correlation matrix of the 156 stocks and used three rules together with bi-plots of the coefficients in each principal component to determine the threshold for cutting off rather irrelevant principal components [3].

1. The most obvious rule in determining the number of components to retain is deciding the cumulative variance desired. PCA was designed so that the variances of the principal components are in descending order with the first principal component explaining the most variance. In the case of a correlation matrix, the percentage variance explained by the first m components was calculated by

$$\text{Variance}_k = \frac{100}{p} \sum_{k=1}^{m} l_k \tag{5.1}$$

where l_k is the eigenvalue of principal component k, p is the total number of variables. The number of components to retain is then the smallest number, m, which exceeds the desired percentage variance explained. Normally the cut-off is in the range of 70% to 90%. But in our case, to preserve 70% of the variation of the original data set, 82 principal components were required when the total study period was used (see Figure 5.1a). The correlations among the 156 stocks were rather moderate during the last 14 years. If the stocks were highly correlated, we would expect that the first few components would absorb most of the variation and leave less variation in the higher numbered components. This means the slope of the cumulative variance explained plot will be steep in the beginning and flatten off towards the end. However, as we can see in Figure 5.1a, the variation of stocks were spread over the components with the first 50 components explained about half of variation and the following 50 components explained approximately 30% of the variation. This leaves about 20% variance explained by the last 56 components. We also separated the study periods into two sub periods: pre- and post- 2008 financial crisis. The results in Figure 5.1b and Figure 5.1c showed that before the crisis, 82 components were needed to explain 70% variance while after the crisis it only required 69 components to explain the same variance. This suggests that there was more common variation in the stock market after

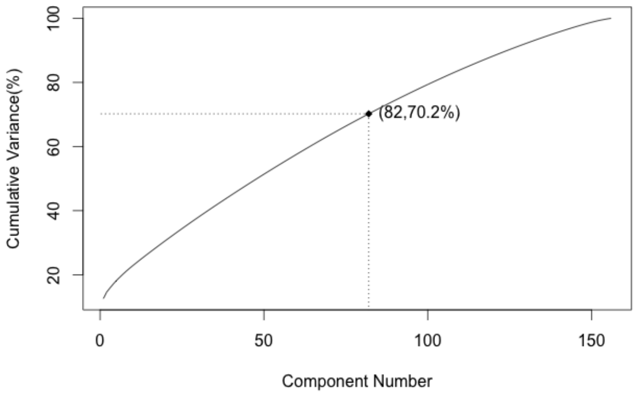
[3]The irrelevant components refer to the random noise in the stock price fluctuations. We note that we do not claim there is not useful information one can extract from those components.

the crisis. When the market was more concentrated, there would be less diversification benefits to exploit. A portfolio that contains the same stocks was less diversified after the crisis compared to before the crisis (see Chapter 8). To summarize, using the cumulative variance to decide the number of components to retain, we would have to choose at least 82 components.

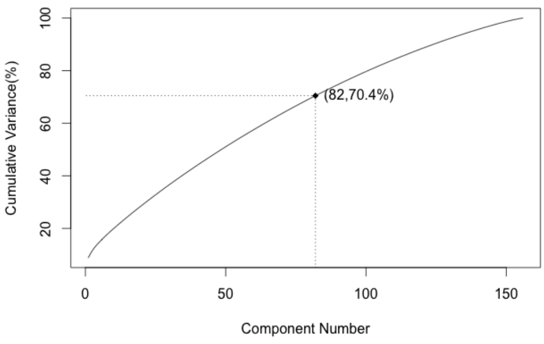
2. The second rule we used is Kaiser's rule (Kaiser, 1960), retaining principal components that have an eigenvalue greater than 1. The idea behind this rule is that if all the stocks were uncorrelated, then the principal components are the same as the original stocks and all have unit variance in the case of a correlation matrix. So any principal components with eigenvalue less than 1 contains less information than one of the original variables and so is not worth retaining. Based on Kaiser's rule, we needed to retain 49 components for the total period, 57 components pre-crisis and 45 components post-crisis. Jolliffe (1986) advised a more conservative cut-off eigenvalue, 0.7, because it would be unwise to delete the components that have eigenvalue close to 1. However, this significantly increased the number of components retained. In this sample, the number of components we needed to be retained increased from 49 to 109, 57 to 107 and 45 to 87 for the three periods respectively. One important thing about Kaiser's rule and the first rule we discussed above is that, when there are a large number of stocks in the sample and the correlations between them are relatively small, a PCA on a correlation matrix will result in most principal components have eigenvalues close to 1 and be retained based on these two rules.

3. The third rule we used is a scree graph (Cattell, 1966) and log-eigenvalue diagram (or LEV diagram) (Farmer, 1971). The scree graph is a plot of eigenvalue against component number. The log-eigenvalue diagram is an alternative to the scree graph and plots the log of the eigenvalue rather than eigenvalue against the component number. The decision is made based on finding the point in the graph where the slopes of lines joining the plotted points are 'steep' to the left and have a linear decay to the right. This is even more subjective than the previous two rules because it involves

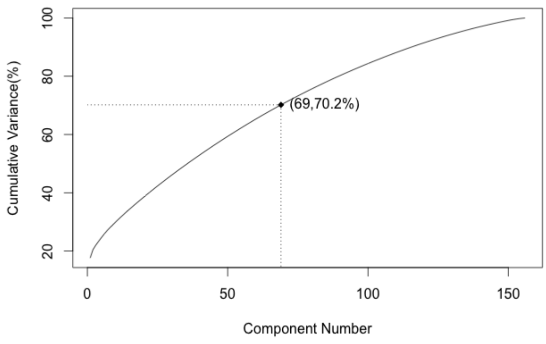
Fig. 5.1 Plots of percentage cumulative variance explained by principal components against component number for total study periods and pre- and post-crisis. The data set used was the 156 stocks which have complete price and dividend information for the complete study period.



(a) Total study period from 04/04/2000 to 17/02/2014.



(b) Before 2008 financial crisis, from 04/04/2000 to 19/09/2007.



(c) After 2008 financial crisis, 20/09/2007 to 17/02/2014.

looking at a plot of eigenvalues or log-eigenvalues against component numbers and deciding at which component the linear decay begins. However, despite the subjectivity, deciding the number of principal components based on the scree graph and log-eigenvalue diagram are used extensively in practice. In Figure 5.2, we present both the scree graph and the LEV diagram for the total study period, the time before and after the crisis. Figure 5.2a, Figure 5.2c and Figure 5.2e show, based on the scree graphs, approximately five components will need to be retained in all three study periods. The LEV diagram, on the other hand, suggests more components are needed to be retained. In Figure 5.2b, the line joining the plotted points became less steep from principal component six. But a linear decay is actually staring from principal component 11. The LEV diagram of pre-crisis period suggests the similar number of principal components to retain, five principal components or a more conservative 10 principal components (see Figure 5.2d). The post-crisis period in Figure 5.2f shows that the slope of the line joining the plotted points stays steep until principal component eight. One can decide conservatively by choosing the point beyond which the scree graph or LEV diagram becomes, approximately, a straight-line. A decision to include a few more components will result in little difference. So based on the scree graphs and LEV diagrams, we will need to retain as many as 11 components.

The three rules above suggested significantly different numbers of components to retain and there is no way that we can tell which one is more accurate. However, if an analysis requires us to preserve as much as possible of the variation, then either deciding based on the cumulative variance or Kaiser's rule is more appropriate. The objective of our analysis is to study the structure of the stock market and preserving more variance is not the priority. We will only retain the components that identify structure within the stock market. In other words, we filter out the random part of principal components and keep the market and group components[4]. In order to do this, we further investigated the contribution of each stock in the principal components for the full periods. We used bi-plots to present the relative

[4]Note that we are excluding the random part of principal components in this chapter, and further investigate the use of high numbered principal components in identifying stocks with near linear relations in Chapter 6.

weights of each stock in the components with respect to ICB industry classification, see Figure 5.3. Each pair of coefficients is connected to the origin by a colored line where the colors correspond to the stocks' ICB industry classification.

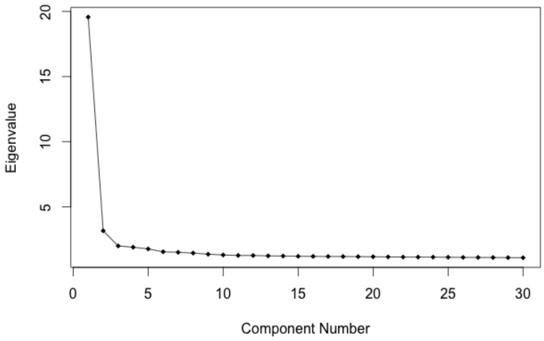
Based on the bi-plots, we can visualize the coefficient structure in the principal components easily. Principal component one has relatively homogenous negative coefficients across all 155 stocks. Note that the principal component one is normally understood as the market component with roughly equal contribution of the underlying stocks and all the coefficients should be positive (Fenn et al., 2011; Kim and Jeong, 2005; Zheng et al., 2012). This is different from our result. Although our results show approximately equal importance of each stock in the principal component one, the sign of the coefficients is opposite. Jolliffe (1986) stated that in circumstances where the first principal component has all its coefficients of the same sign and all the variables are thought of equal importance, it is a measure of size. This is a reasonable interpretation for all negative coefficients but we still cannot explain why our finding is contrary to the positive coefficients that are reported in many other papers. We also discuss the problem of all negative coefficients in Chapter 7 when we constructing portfolios based on each principal component.

Principal component two, on the other hand, has positive coefficients on Basic Materials, Oil & Gas, and Consumer Goods and negative coefficients on Financials, Health Care, and Consumer Services. The Industrials are more ambiguous. Stocks in the Industrials show both positive and negative coefficients in principal component two. In the subsequent principal components it is less straightforward to pinpoint certain sector tilts. Even so, there are some interesting distinctions in some principal components. Principal component three and principal component five showed similar structure but in an opposite way. Health Care, Consumer Services, and Industrials have positive coefficients in component three but have negative in component five. Basic Materials and Financials mostly are negative in component three but positive in component five. Principal component seven has positive Industrials and Consumer Services against negative Financials, Health Care, Utilities, and Technology. In principal component eight, SDL which belongs to the Basic Materials industry has a large positive coefficient above 0.4 in contrast to the Financial company SDG that has same value of coefficient but opposite sign. Likewise, principal component nine is dominated by these two stocks. Unsurprisingly, SDL became negative and SDG became positive.

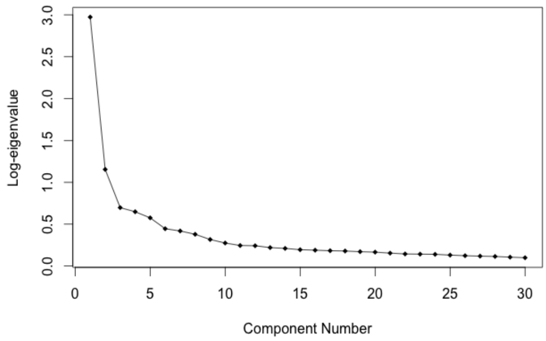
Principal component 10 onwards, bi-plots showed a "star-like" graph, the distinction is less clear-cut and structure was hardly seen. Thus, retaining 82 components as the first rule suggested or about 49 components based on Kaiser's rule will be too conservative. The scree graph and LEV diagrams indicated a more reasonable number of components to retain in our case. Combining the investigation of stock coefficients in principal components with the scree graph and LEV diagram in Figure 5.2, 10 principal components were retained. This means there were 10 major risk sources in the Australian stock market. In Chapter 7, we will construct portfolios based on each of the retained principal components. So each portfolio represents a risk source that is uncorrelated to the others. One can diversify a portfolio by holding all the ten "principal component mimicking portfolios", in which one gains exposure to all the major risk sources in the stock market (for more discussion see Chapter 7).

Table 5.1 presents the eigenvalues and cumulative variances of first 10 principal components for the full periods, before the 2008 financial crisis and after the 2008 financial crisis. In previous discussion, the number of principal components required to retain about 70% variation was 82 components pre-crisis and post-crisis it only took 69 components. Table 5.1 showed that the reason for this was that the first principal component absorbed a lot more variation post crisis. The variance explained by the first component was 8.73% before crisis and doubled after the crisis, to 17.61%. The variance explained by component two also increased significantly, a 44% rise from 2% to 2.88%. The variance explained by subsequent principal components showed no distinction before and after the crisis. The intuition behind this is that after the financial crisis in 2008, the market risk, which is the undiversifiable risk, increased significantly compared to the idiosyncratic risk. In Chapter 8, we will discuss using the variance explained by principal component one as a measure of systemic risk, which is a ratio of systematic risk and idiosyncratic risk.

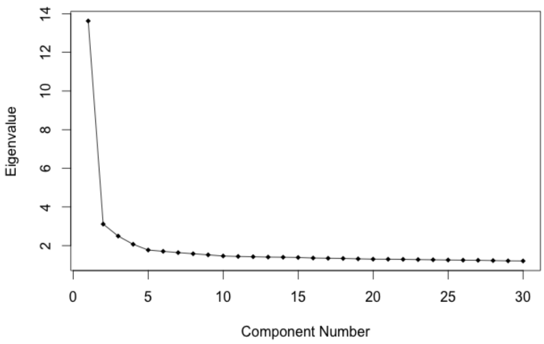
Fig. 5.2 Scree graph and Log-eigenvalue diagram using complete study periods, pre- and post-crisis study periods arising from PCA on a correlation matrix of 156 stocks (only the first 30 principal components are presented).



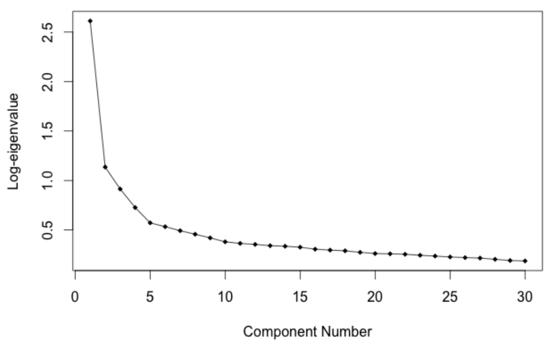
(a) Scree graph for correlation matrix of 156 stocks, for the complete study period from 04/04/2000 to 17/02/2014.



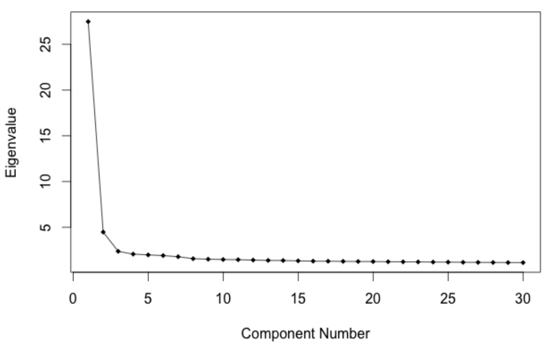
(b) Log-eigenvalue (or LEV) diagram for correlation matrix of 156 stocks, for the complete study period from 04/04/2000 to 17/02/2014.



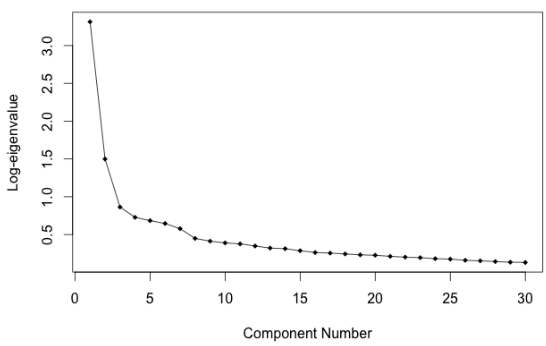
(c) Scree graph for correlation matrix of 156 stocks, pre-crisis.



(d) Log-eigenvalue (or LEV) diagram for correlation matrix of 156 stocks, pre-crisis.

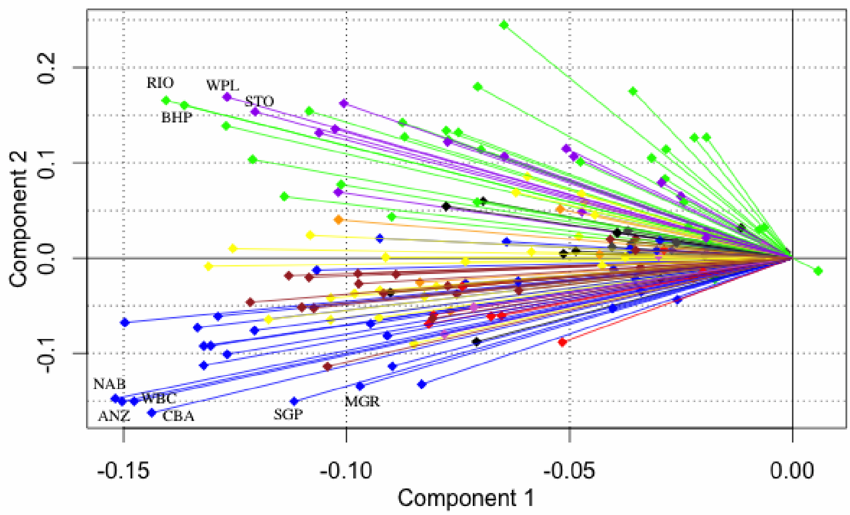


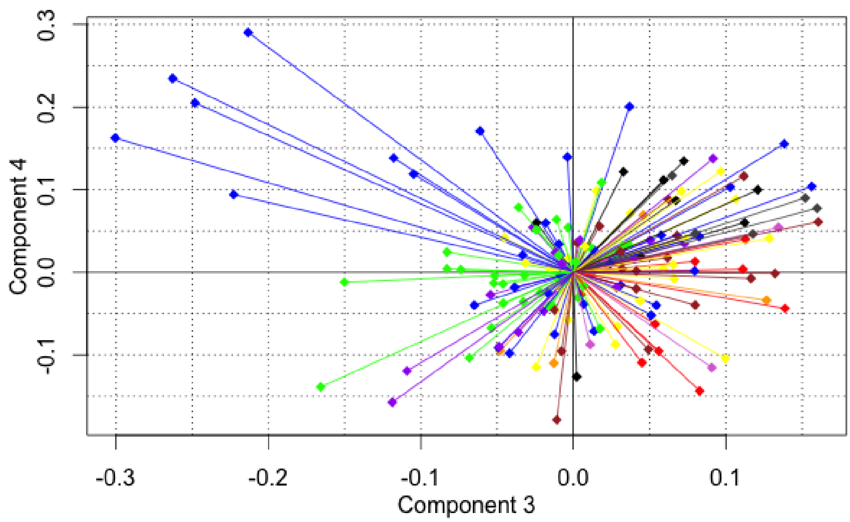
(e) Scree graph for correlation matrix of 156 stocks, post-crisis.

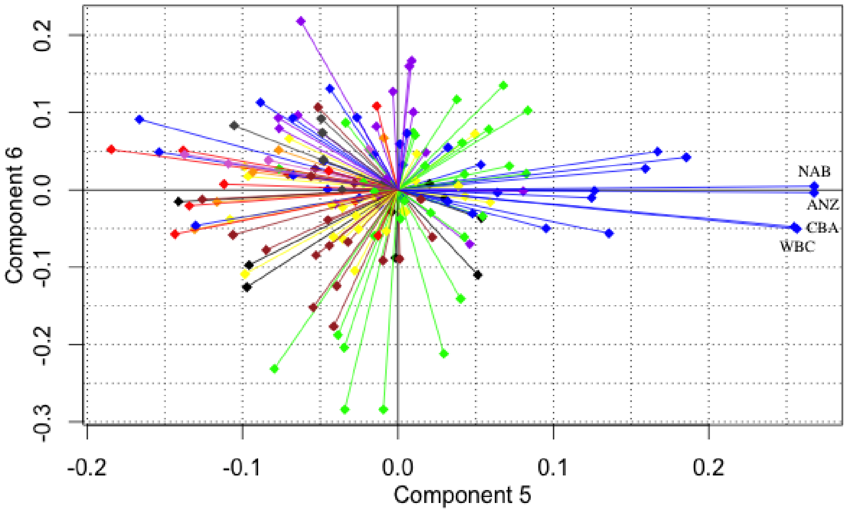


(f) Log-eigenvalue (or LEV) diagram for correlation matrix of 156 stocks, post-crisis.

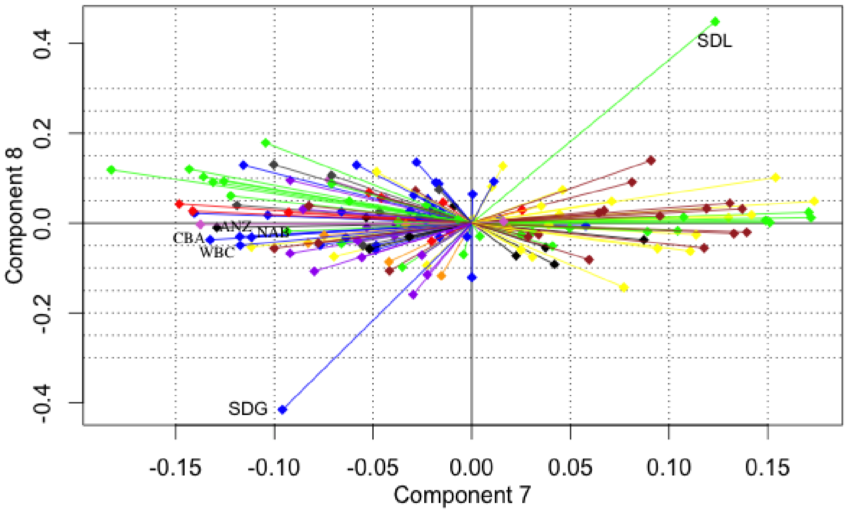
Fig. 5.3 Bi-plots of relative weights of each stock in components 1 to 12 arising from a PCA on a correlation matrix from the whole study period, April 2000 to February 2014, using the Industry Classification Benchmark. The colors correspond to respective ICB sector classification: Financials are Blue (33 stocks), Health Care are Red (9 stocks), Industrials are Yellow (24 stocks), Consumer Services are Brown (19 stocks), Basic Materials are Green (31 stocks), Oil&Gas are Purple (16 stocks), Utilities are Orange (5 stocks), Consumer Goods are Black (9 stocks), Telecommunications are Orchid (4 stocks), Technology are Grey (6 stocks).



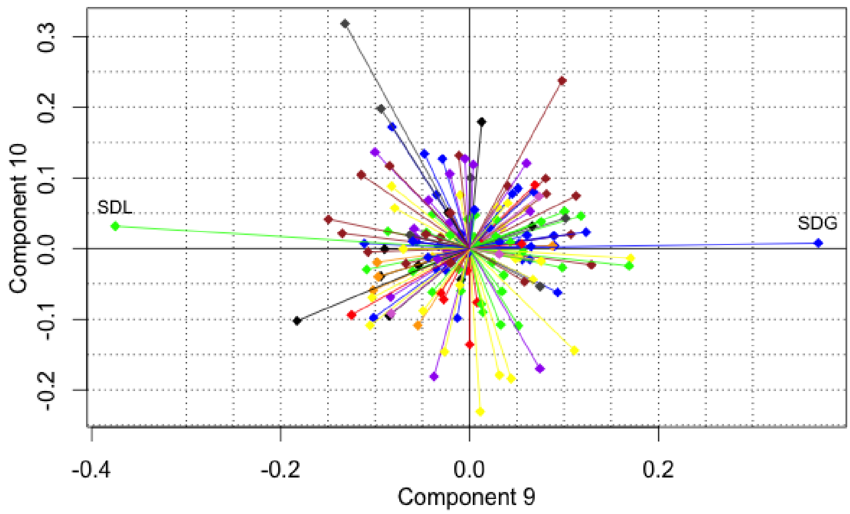




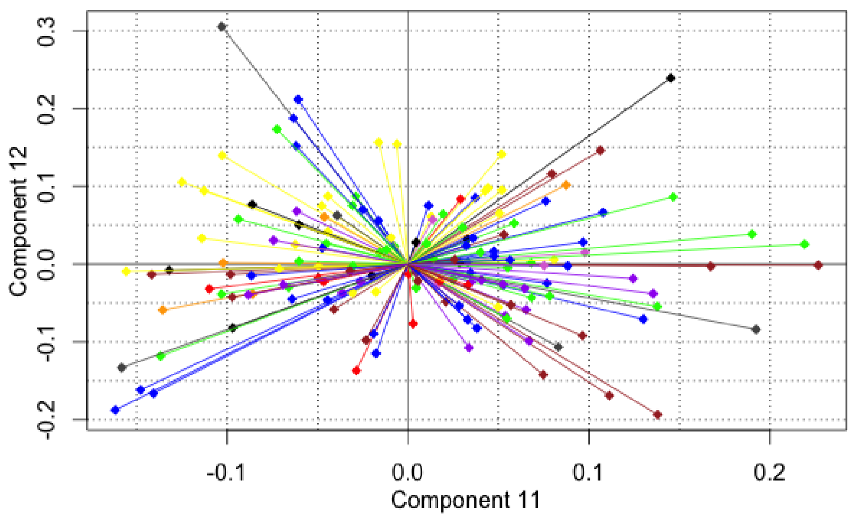
NAB
ANZ
CBA
WBC
Component 6
Component 5



SDL
SDG
CBA
ANZ
NAB
WBC
Component 8
Component 7



SDL
SDG
Component 10
Component 9
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.4
-0.2
0.0
0.2



Component 12
Component 11
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.1
0.0
0.1
0.2

Table 5.1 Eigenvalues and variance explained by PC1 to PC10 using the full study periods, pre- and post-crisis study periods arising from a PCA on a correlation matrix of 156 stocks.

	Total periods		**Pre-Crisis**		**Post-Crisis**	
No.	Eigenvalue	Cumulative Variance(%)	Eigenvalue	Cumulative Variance(%)	Eigenvalue	Cumulative Variance(%)
1	19.56	12.54	13.63	8.73	27.48	17.61
2	3.17	14.57	3.11	10.73	4.48	20.49
3	2.01	15.86	2.49	12.33	2.37	22.01
4	1.91	17.09	2.07	13.65	2.07	23.34
5	1.78	18.22	1.77	14.79	1.98	24.61
6	1.56	19.23	1.70	15.89	1.91	25.83
7	1.52	20.20	1.61	16.93	1.78	26.97
8	1.46	21.14	1.58	17.94	1.57	27.98
9	1.37	22.02	1.52	18.92	1.51	28.95
10	1.32	22.02	1.46	19.85	1.48	29.89

Chapter 6

Identifying near-linear relationships between stocks

In the last chapter, the potential use of bi-plots in visualizing the coefficients within the principal components was briefly discussed. We have used the bi-plots to identify the boundary where the coefficients in the principal components become star like, that is, there is no clear grouping of stocks. In this chapter, we present the bi-plots of the last six principal components which successfully pick up stocks with near linear relationships. Notably, if a set of variables has substantial correlations among them, the low variance principal components will ensure any near linear relationship is detected (Jolliffe, 1986).

In Figure 6.1, principal components 151 and 152 pick up six stocks. The first pair Mirvac Group (MGR) and Stockland (SGP) are two large diversified property groups in Australia. Their price correlation coefficient in the last 14 years was 0.71 (see Figure 6.2a). In the beginning of 2000, they had roughly the same stock price[1] level and have diverged since then. However, despite the price level being different, they have moved closely together in the study period. Especially in the first ten years, the prices moved up and down together.

The second pair of stocks are Santos Limited (STO) and Woodside Petroleum Limited (WPL), which are in the Oil & Gas industry. Both companies explore for and produce oil and gas from onshore and offshore wells. They had a even higher correlation coefficient,

[1]For the price of stock here and after, we are referring to the dividend adjusted price.

Table 6.1 Price correlation coefficients of the four big banks.

	ANZ	WBC	CBA	NAB
ANZ	1	0.97	0.96	0.85
WBC	0.97	1	0.98	0.76
CBA	0.96	0.98	1	0.73
NAB	0.85	0.76	0.73	1

0.95 (see Figure 6.2b). We observed the stock prices of STO and WPL were both relatively stable in the first five years and they became volatile about the same time.

The last pair of stocks are not in the same industry but had a high correlation coefficient, 0.91 (Figure 6.2c). These two stocks are BHP Billiton Ltd (BHP) in Basic Materials and CFS Retail Property Trust Group (CFX) in the Financial industry. While the correlation coefficient of the two stocks was 0.91, they tended to move in the opposite direction after the 2008 financial crisis. Figure 6.2b shows in some points of time (e.g. 7 Feb 2012), BHP reached its peak and CFX in its bottom or the other way around.

The last four components all picked up the four big banks in Australia, ANZ, WBC, CBA and NAB. Table 6.1 shows the price correlations between the four big banks. Except for NAB, the other three banks had price correlations higher than 0.95. NAB had the strongest price correlation with ANZ and the price correlations with the other two banks were also high (all above 0.7). These relationships are easily visualized in Figure 6.3a. To help visualizing the price co-movement of the four big banks, we use a different scaling for CBA. Its price changed from \$20 in the beginning of our study period to approximately \$150 at the end of study period while the other three banks had price levels that ranged from \$10 to \$70. Obviously, NAB was least correlated with others among the four banks. But after the 2008 financial crisis, all four banks converged to move very similarly

In the bi-plot of principal components 155 and 156, Australia's two biggest mining firms also were picked up, BHP Billiton Ltd (BHP) and Rio Tinto Ltd (RIO) (see Figure 6.3b). At the beginning of our study period, the price of RIO was approximately 1.4 times of BHP. Before the price collapsed in 2008, both two stocks increased significantly and RIO increased even more. At the end of 2007, the price of RIO was about 2.5 times of BHP.

Table 6.2 Eigenvalues and variances explained by the last six principal components.

	Eigenvalue	**Variance explained(%)**
PC151	0.40	0.25%
PC152	0.38	0.24%
PC153	0.32	0.21%
PC154	0.30	0.19%
PC155	0.29	0.18%
PC156	0.24	0.16%

However, during the 2008 financial crisis, RIO also declined more than BHP. At the end of our study period, the price of RIO was about 1.5 times of BHP, which is almost the same as it was at the beginning of our study period.

The low variance principal components effectively detected stocks with high correlations. The coefficients of stocks in the principal components in theory can be an approximation for connectedness between stocks. Recall that each principal component is a linear combination of all the variables (see Chapter 2). The eigenvalues of the last few principal components were small. In some applications the eigenvalues are very close to zero. In our case the eigenvalues were clearly different from zero, nevertheless, they still picked up near linear relationships between some stocks (see Table 6.2). The eigenvalue of each principal component is a linear function of all variables (Jolliffe, 1986), which can be rewritten as

$$\alpha_k'\mathbf{x} = \alpha_{k1}x_1 + \alpha_{k2}x_2 + \sum_{i=3}^{p} \alpha_{ki}x_i \tag{6.1}$$

where $\alpha_k'\mathbf{x}$ is the eigenvalue of component k, α_{ki} is the coefficient of stock i in component k. Assume the eigenvalue of principal component k is small and very close to zero, if x_1 and x_2 are the two highly correlated variables being detected in component k, and with much larger coefficients while the rest of the variables have near zero coefficients, the function above is then

$$0 \approx \alpha_{k1}x_1 + \alpha_{k2}x_2 + 0. \tag{6.2}$$

As a consequence, the closer α_{k1} and α_{k2} are in magnitude, the more correlated of the x_1 and x_2. If x_1 and x_2 are highly positively correlated then α_{k1} and α_{k2} will have opposite signs.

In Figure 6.1, we have seen symmetrical lines in the bi-plots, which suggested a similar magnitudes of coefficient levels with opposite signs and therefore high correlations.

Notably, the highly correlated stocks tend to have similar coefficients in the first few components. We noticed that the four big banks showed up together in the same position in the bi-plot of components one and two, components five and six, and components seven and eight (see Figure 5.3). In the bi-plot of components one and two, other pairs of high correlated stocks also coexist: RIO and BHP, WPL and STO, SGP and MGR.

Kim and Jeong (2005) reported that only the market component and its subsequent group components contains useful information. They assigned no useful role to the high principal components. However, our results illustrate the use of the last few principal components in identifying stocks with near linear correlations. To reduce risk, one should avoid including stocks with high correlation with each other in a portfolio. One hundred stocks whose returns increase and decrease together provide little more protection than the uncertain return of a single stock. In Chapter 10, we propose a method of selecting stocks to describe the full data set based on PCA. The idea of this method is based on the fact that the last few principal components pick highly correlated stocks. By retaining one of the highly correlated stocks and excluding the others, little information will be lost. Chapter 10 shows this stock selecting method is effective and a portfolio of approximately 20 stocks will closely resemble the ASX200 index, that includes 200 stocks, in terms of the fluctuation in portfolio value.

Fig. 6.1 Bi-plots of relative weights of each stock in components 151 to 156 arising from a PCA on a correlation matrix from the whole study period, April 2000 to February 2014, using the Industry Classification Benchmark industry classification. The colors correspond to respective ICB sector classification: Financials are Blue (33 stocks), Health Care are Red (9 stocks), Industrials are Yellow (24 stocks), Consumer Services are Brown (19 stocks), Basic Materials are Green (31 stocks), Oil&Gas are Purple (16 stocks), Utilities are orange (5 stocks), Consumer Goods are Black (9 stocks), Telecommunications are Orchid (4 stocks), Technology are Grey (6 stocks).

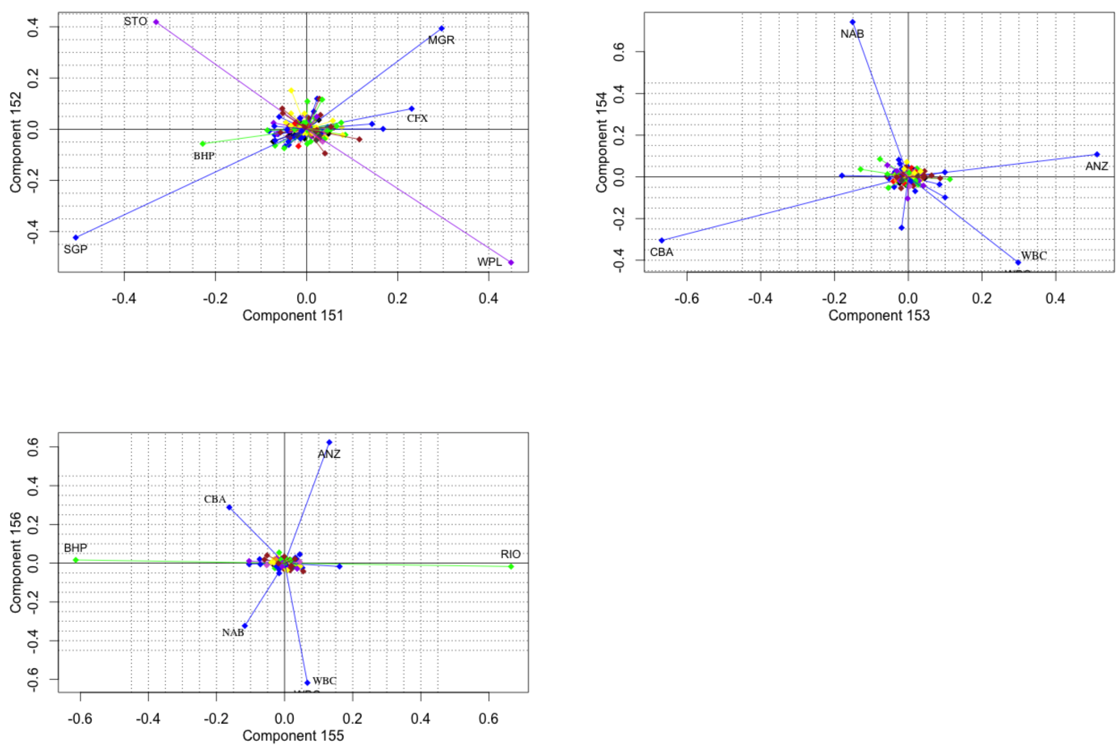
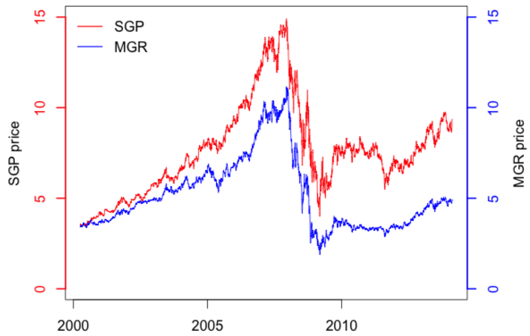
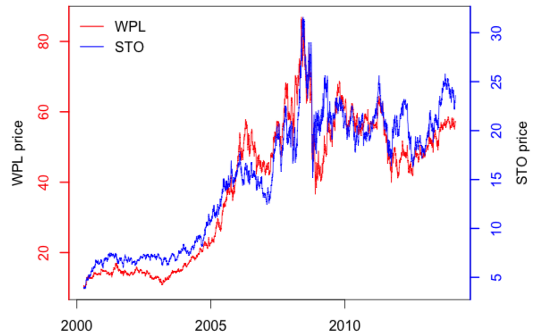


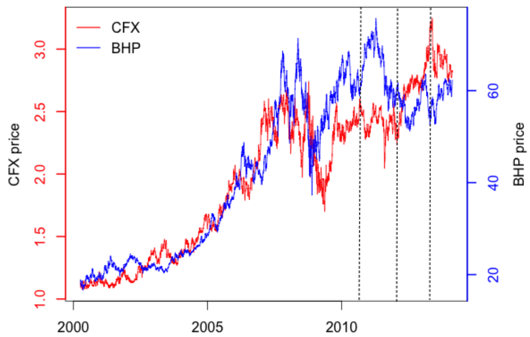
Fig. 6.2 Time series plots of near linear correlated stocks identified in principal components 151 and 152.



(a) Time series plot of two stocks in Financial industry: SGP and MGR. The correlation coefficient is 0.71.

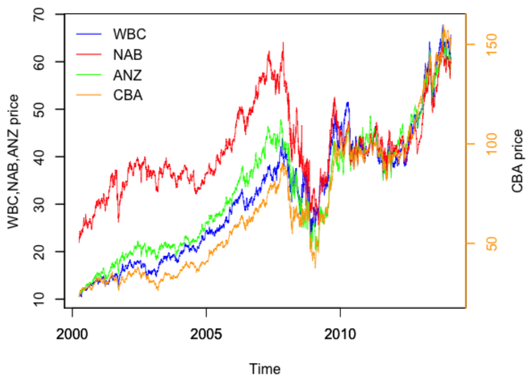


(b) Time series plot of two stocks in Oil & Gas industry: STO and WPL. The correlation coefficient is 0.95.

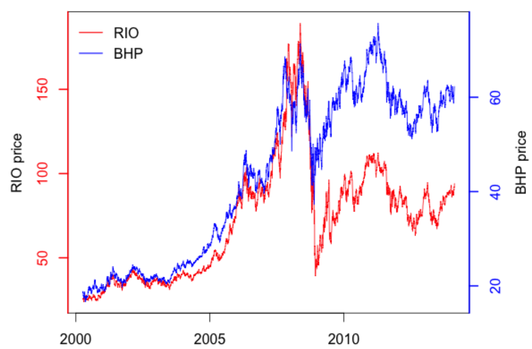


(c) Time series plot of two stocks: CFX (Financial industry) and BHP (Basic Materials industry). The correlation coefficient is 0.91.

Fig. 6.3 Time series plots of near linear correlated stocks identified in principal components 153 to 156.



(a) Time series plot of the four big banks in Australia.



(b) Time series plot of two stocks in Basic Materials: RIO and BHP. The correlation coefficient is 0.83.

Chapter 7

Principal Portfolios

Recall that a principal component analysis can be used to extract uncorrelated synthetic portfolios which represent uncorrelated risk sources in the stock market (Partovi and Caputo, 2004). In Chapter 5, we filtered out the random part of the stock risks and retained the first 10 principal components that represent the market risk and each risk group. The original set of 156 stocks was transformed to a principal system which included 156 uncorrelated principal components in which the first 10 principal components identified the major risk drivers of stock returns. Essentially, the portfolios constructed based on the principal components were treated as individual investment assets with no correlations. We followed Partovi and Caputo (2004) and called the portfolios constructed based on principal components the "principal portfolios" (PPs).

The purpose of doing this is straightforward. A single risk exposure becomes feasible. Investors can choose to hold any principal portfolio to get exposure to a single risk source that is uncorrelated with the other risks in the market. The performances of principal portfolios also provide means of monitoring single risk exposures. The investment universe is simplified in the sense that the choices are among assets with uncorrelated risks. One can decide whether to include an asset solely based on its variance and return without concern about its co-movements with the others in the portfolio.

In the principal system, all the asset allocation strategies that can be applied to individual stocks can also be applied to PPs. The only difference is instead of using stocks returns as input, we are using the returns of PPs. Meucci (2009) stated that maximum diversification is achieved when a portfolio has equal exposure to all uncorrelated risk sources. This concept coincides with allocating equal risk budgets to all PPs. Conversely, holding a single risk portfolio is considered under-diversified. So in theory, in order to achieve maximum diversification, we should include all 156 PPs. However, it is unreasonable to allocate equal risk budget to both major risk sources and the random part of the stock price fluctuations. So of all the PPs, we refer to the 10 PPs that represent the major risk sources. The allocation strategy - that of budgeting equal risk on investment assets - is known as the equal risk contribution (ERC).

Moreover, we also applied the naive allocation strategy, which is equal investment in 10 PPs. The difference between these two strategies is the naive allocation strategy budgets equal dollar investment on the PPs while the ERC budgets equal risk. In this chapter, we begin with the investigation of the individual PPs. We will discuss the construction of the PPs and the performance of each PP relative to the ASX200 index over time. We next test the naive and ERC allocation strategies on PPs then compare them to portfolios constructed of the same allocation strategies based on the stocks.

7.1 Constructing principal portfolios

The PPs were constructed in the following steps:

1. Apply a PCA and get the coefficients of the principal components.

2. A positive coefficient indicates a long position while a negative coefficient indicates a short position. The weights of investment in each stock is the stock coefficient divided by the sum of all positive coefficients (if it is positive) or divided the absolute value of sum of all negative coefficients (if it is negative). This gives a set of weights in which both the long positions sum to 1 and the short positions sum to -1 respectively. The portion of short positions is the ratio of the sum of all negative coefficients to the sum of all positive coefficients. The funds obtained from the short positions are assumed to

be invested in an average risk free rate[1] (Australian Negotiable Certificates of Deposit-90 days) over the last 14 years. Because we were using daily returns, the annual risk free rate was converted into a daily rate.

3. The PPs returns are then the sum of the weighted returns of each stock plus the product of risk free rate and the ratio of the short position.

The back-testing of the PPs was carried out on a rolling window basis for the whole sample period. The PP returns were calculated with respect to a window size of two years (504 observations) using the procedure described above and the portfolio weights were re-balanced daily. So daily returns from 3 April 2002 to 17 February 2014 were computed for each principal component one to ten.

It is worth noting that PP1 is constructed slightly differently than the others. We have mentioned in Chapter 5 that the sign of the coefficients in principal component one in our result was opposite to the findings that were reported in many other papers (Fenn et al., 2011; Kim and Jeong, 2005; Zheng et al., 2012). Principal component one is understood to be the market component that has approximately equal contribution of all stocks, but sampling variability ensures that they are never exactly the same (Jolliffe, 1986). Normally the sign of all coefficients in principal component one would be positive. However, our results show principal component one contains all negative coefficients, except for MGL which has nearly half of the zero return observations. We suspected that MGL is different from the others due to its large number of zero returns. We decided to manually change the coefficient sign of MGL to be the same as others. Because its coefficient in principal component one is small and near zero, we believe changing its sign to negative will not affect the performance of the PP1. The negative coefficients suggest that the market component is shorting all the stocks. This is inconsistent with principal component one as a "market component". When a negative shock affects all stocks and they all have negative returns, PP1 will perform better than the market. PCA by design is such that the direction of the principal components has no effect on their variances (the eigenvalues). As a consequence,

[1]The average 90 days NCDs from April 2000 to February 2014 is 4.97% annually, and the daily rate is 0.02%

we decided to change the all the coefficients to be positive, which is equivalent to rotating the principal component one by 180 degrees. After changing the sign of the coefficients, the PP1 returns were calculated using the same procedure as the other PPs.

Figure 7.1 presents the trajectory of PPs against ASX200 index value and their relative performance. All the portfolios are assumed to have an initial wealth position equal to $1 million. The relative performances are shown in the right panel of Figure 7.1, which are the differences between the principal portfolio value and index value over the index value:

$$\frac{PP_n(t) - ASX200(t)}{ASX200(t)} \tag{7.1}$$

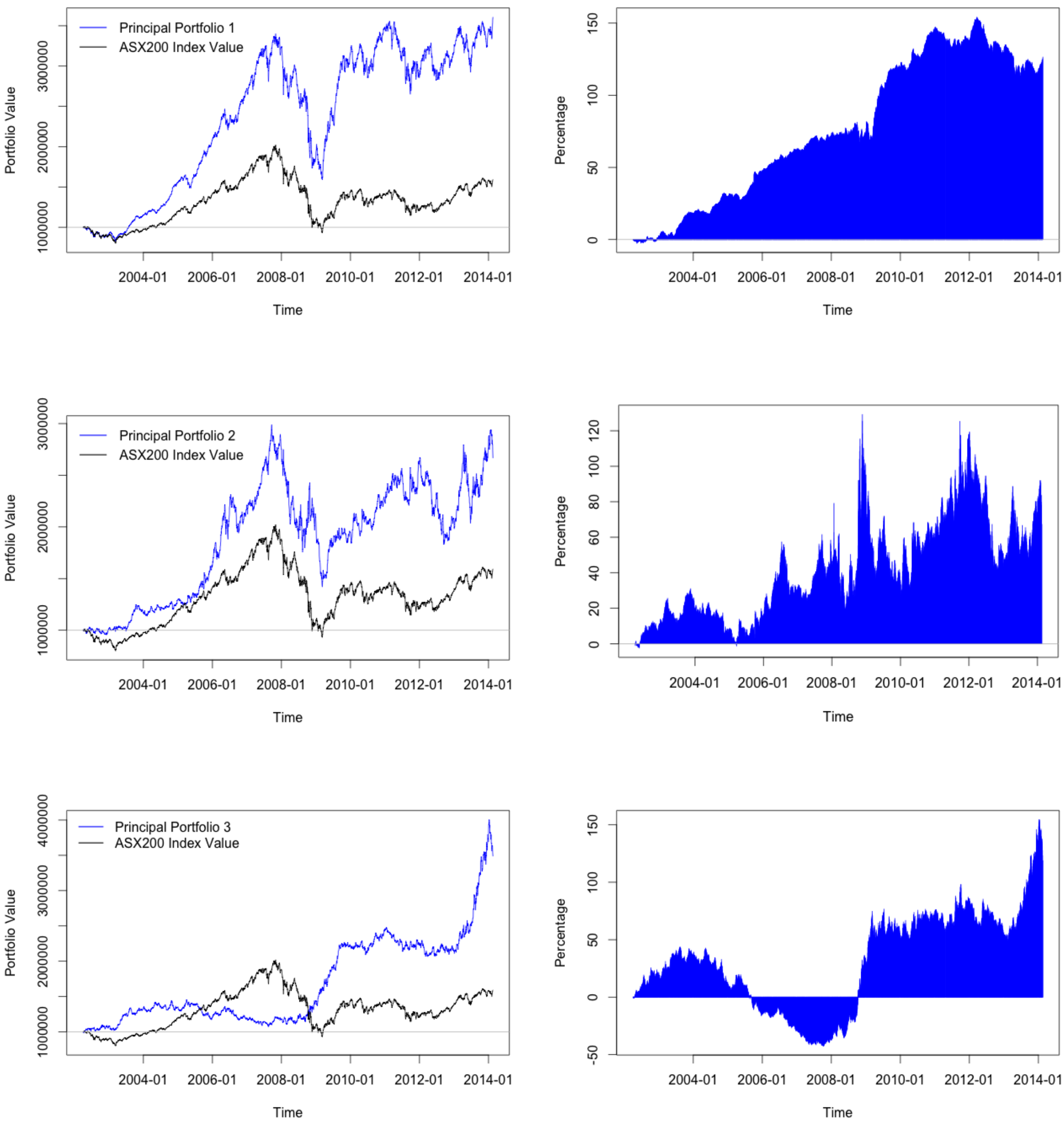
where $PP_n(t)$ is the principal portfolio n value at time t, $ASX200(t)$ is the index value at time t and t is in units of one trading day.

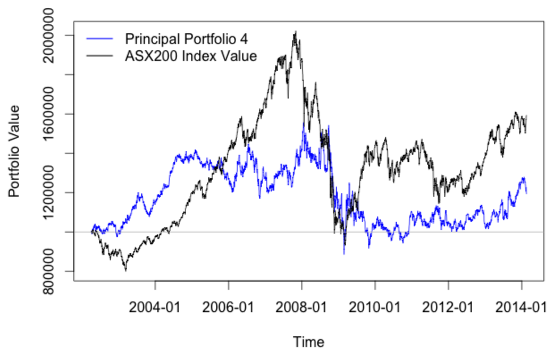
The graph shows PP1 and PP2 were closely related ASX200 index especially before beginning of 2009. PP1, PP2 and the ASX200 all increased significantly from 2002 and peaked on 1 November 2007. The price collapsed started from November 2007. This was coincident with the time the US S&P500 index started to drop. The 2008 financial crisis spread from the US to Australia quickly and the Australian stock market reacted almost immediately. The Australian market began to recover from March 2009 and the ASX200 index price level had recovered only half of the price drop caused by the crisis by February 2014. However, PP1 and PP2 increased sharply to a portfolio value even higher than their peaks at the end of 2007. Both PP1 and PP2 were far more volatile than the index and look like an amplified version of the ASX200 index. Table 7.1 presents the price correlations between the PPs and ASX200 index and equally-weighted portfolio of 156 stocks. The table shows PP1 and PP2 were more correlated than the other eight PPs with the ASX200 index. Interestingly, the price correlation between the PP1 and ASX200 index was 0.78 and slightly lower than the correlation between PP2 and ASX200 index. PP1 has its special role among all the other PPs. It is a market component that has approximately equal contribution of all stocks and typically all the stocks have same sign of coefficient. This suggests PP1 is essentially an equal weighted portfolio of the underlying stocks. We can see that PP1 had a price correlation of 0.99 with 1/N portfolio (see Table 7.1). They have the roughly the same portfolio value trajectory over the full study period (for 1/N portfolio, see Figure 7.3a). PP1 represents a market wide effect, and it is not necessarily the most risky portfolio of all PPs, but it is certainly the one which has the most systematic risk. This means the 1/N portfolio is the easiest way to get a portfolio that has market risk rather than idiosyncratic risks. We note that PP1 was the only portfolio that had high daily return correlations with the 1/N portfolio and ASX200 index while all other nine PPs were near zero (see Table 7.2). This again reaffirmed its role as a market component. PP2 had the highest price correlation with the ASX200. It also highly correlated with the 1/N portfolio over the full study period. Although the price correlations between PP2 to the ASX200 and 1/N were high, their daily return correlations were about zero (see Table 7.2). PP2 followed the ASX200 index in the long term, but its daily movement was uncorrelated with the index. This was also the case for PP7. While its price correlation with the index was 0.75, their daily return correlation was only 0.01. Moreover, Figure 7.1 also shows PP1 and PP2 were the only portfolios among all PPs which consistently outperformed the ASX200 over the full study period. The outperformance of PP1 increased gradually until 2008 and decreased slightly during the 2008 financial crisis. When the market started to recover from 2009, the outperformance of PP1 rose suddenly and the cumulative percentage outperformance was over 100%. Conversely, the percentage outperformance of PP2 was more volatile than PP1 over the study period. It increased in the first two years and dropped to zero around 2005. This decrease of the outperformance was due to a relatively slow increase in the portfolio value in PP2 from 2004 to 2005. We note that when the outperformance of PP1 dropped slightly in 2008, the outperformance of PP2 increased dramatically to the highest point in the full study period. One explanation for this is PP2 was more sensitive to the positive news during the 2008 financial crisis. When the market recovered only a little, PP2 reacted in a dramatic way. The percentage outperformance of PP2 also increased after the 2008 financial crisis and declined from 2012.

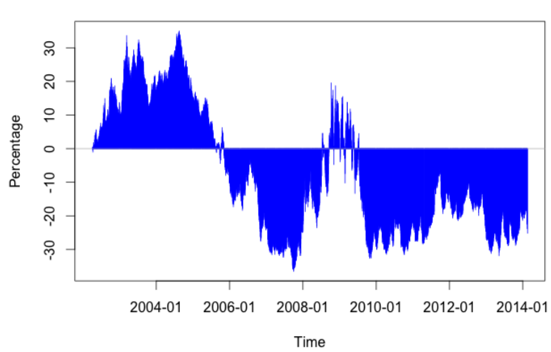
While PP1, PP2, and PP7 had high price correlations with the 1/N portfolio and ASX200 index, PP3, PP8, PP9, and PP10 had high price correlations with the 1/N portfolio and low price correlations with ASX200 index. Figure 7.1 shows that all four PPs were not affected by the 2008 financial crisis, their portfolio values were all increasing in the falling market period. Our investigations into the relative performance of PP3 and ASX200 index showed there were two periods of sharp increase in the cumulative percentage outperformance. One was around the end of the 2008 financial crisis, when the ASX200 continued decreasing, PP3 increased significantly and reached its first peak at the end of 2009. The other one was at the end of our study period, while the index value was relatively stable and increased a little at the end of the period, PP3 jumped from roughly \$2 million to about \$4 million. PP8 increased in the first four years. The two years before the 2008 financial crisis, the portfolio value stopped increasing and the value stayed in the range between \$1.2 million and \$1.4 million. When the market started falling, PP8 continued to increase and reached its peak at the end of 2011. We note that PP8 was affected by the meltdown which happen in August 2011, the contagion of European sovereign debt crisis, American credit down grade and fears about the global economy. PP9 increased all the way from 2002 to 2014 and the portfolio value tripled. PP10 moved similarly to PP3 except its increase was more gradually than PP3 in the last two years.

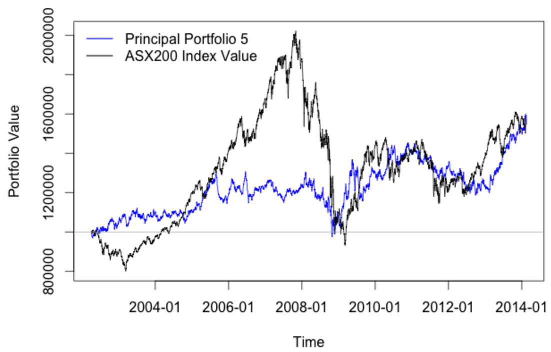
PP4 had a negative price correlation with the 1/N portfolio and a low positive price correlation with ASX200. The portfolio value trajectory of PP4 was similar to PP8 until 2008. When PP8 moved in the opposite with the ASX200 during the 2008 financial crisis, while PP4 tended to follow the market trend. Moreover, PP4 has mostly underperformed the ASX200 from 2006. PP5 had a price correlation 0.84 with 1/N portfolio and a rather moderate correlation with the ASX200 index, which is 0.53. We can see that PP5 was relatively stable until 2009 and has moved closely with the ASX200 index since then. The relative performance plot of PP5 illustrates that the differences between PP5 portfolio value and ASX200 index value were small in the second half study period. PP6 had low price correlation with both 1/N and the ASX200 index. We can see that its portfolio was the most stable but mostly underperformed the ASX200 index in the full study period.

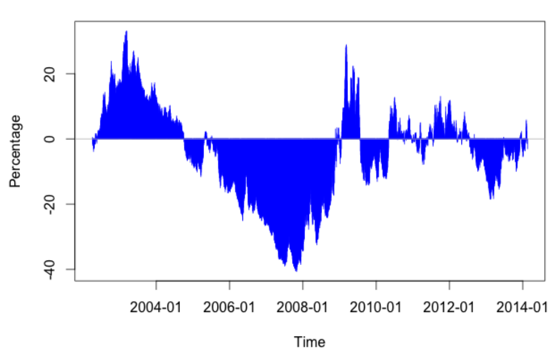
Fig. 7.1 Plots of principal portfolios 1 to 10 with the ASX200 index. The data set used is the 156 stocks for the whole study period. The backtesting of each PP is constructed on a rolling window of two years and rebalanced daily. All plots start from April 2002 and end February 2014. The right panel shows the relative performance of PPs and the index and is calculated using $100 * (PP_n - ASX200)/ASX200$.

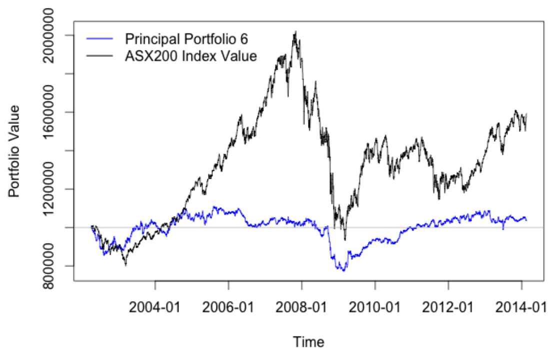


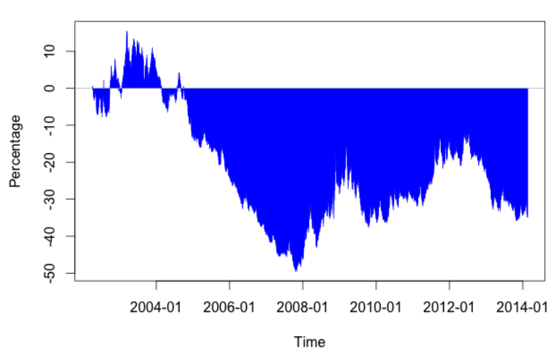

Principal Portfolio 4
ASX200 Index Value
Portfolio Value
800000
1200000
1600000
2000000
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time

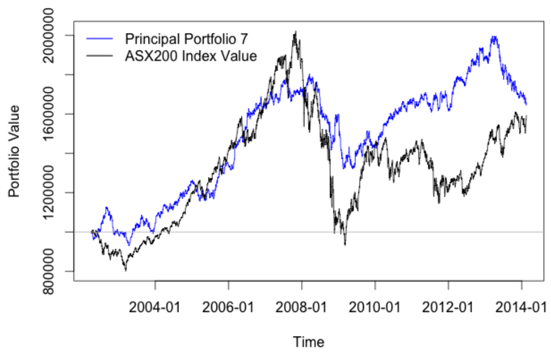

Percentage
-30
-20
-10
0
10
20
30
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time

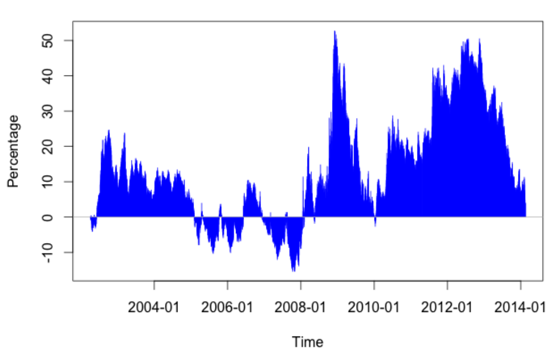

Principal Portfolio 5
ASX200 Index Value
Portfolio Value
800000
1200000
1600000
2000000
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time


Percentage
-40
-20
0
20
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time


Principal Portfolio 6
ASX200 Index Value
Portfolio Value
800000
1200000
1600000
2000000
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time


Percentage
-50
-40
-30
-20
-10
0
10
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time


Principal Portfolio 7
ASX200 Index Value
Portfolio Value
800000
1200000
1600000
2000000
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time


Percentage
-10
0
10
20
30
40
50
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01
Time

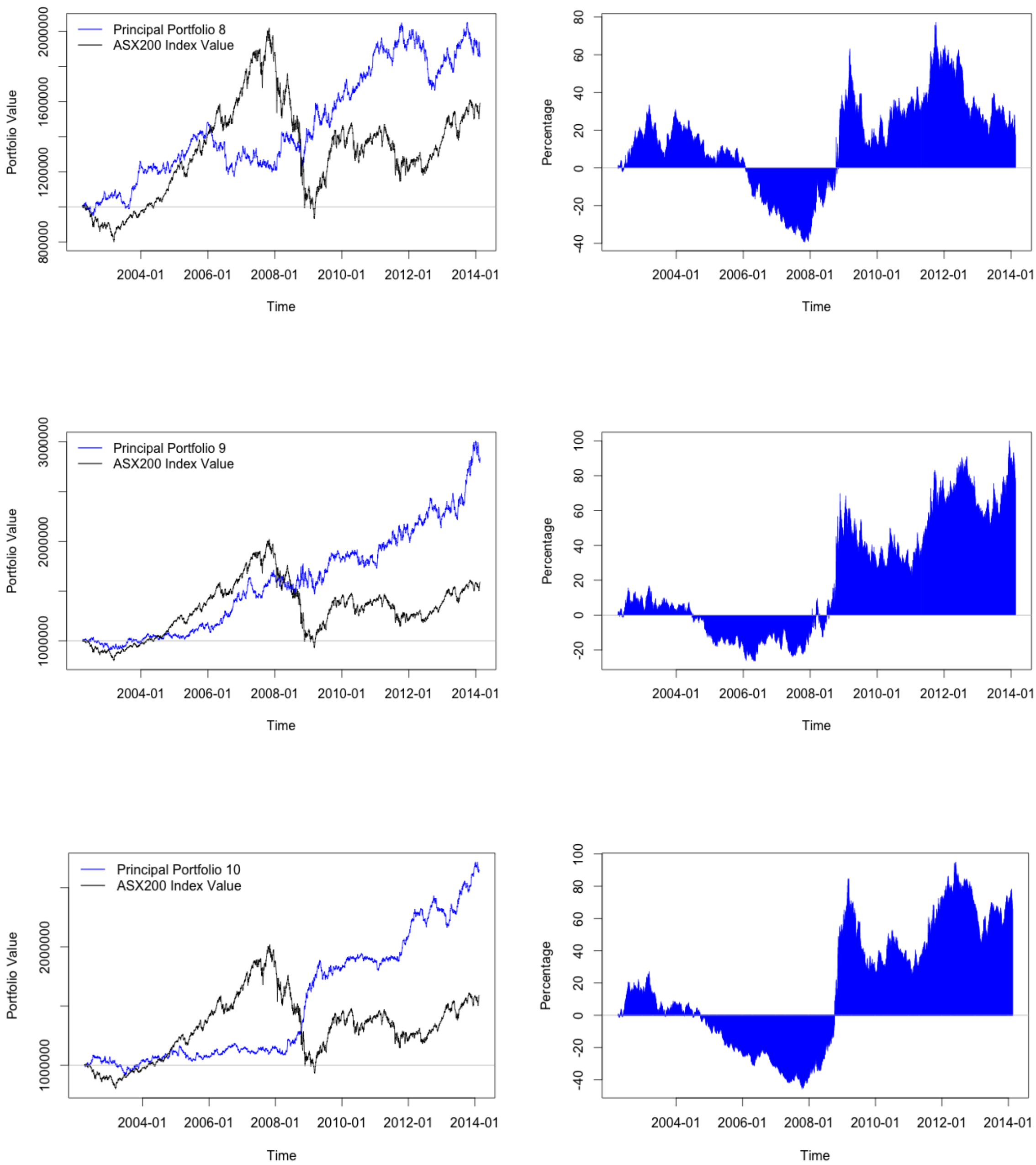

Principal Portfolio 8
ASX200 Index Value
Portfolio Value
Percentage
Principal Portfolio 9
ASX200 Index Value
Principal Portfolio 10
ASX200 Index Value
Time
2004-01
2006-01
2008-01
2010-01
2012-01
2014-01

Table 7.1 Price correlations of each PP to the 1/N and ASX200 index.

	1/N	**ASX200**
PP1	0.99	0.78
PP2	0.91	0.80
PP3	0.75	0.18
PP4	-0.34	0.31
PP5	0.84	0.53
PP6	0.21	0.40
PP7	0.93	0.75
PP8	0.85	0.30
PP9	0.92	0.44
PP10	0.81	0.19

Table 7.2 Daily return correlations of each PP to the 1/N and ASX200 index.

	1/N	**ASX200**
PP1	0.96	0.94
PP2	0.00	-0.03
PP3	0.13	0.16
PP4	0.07	0.05
PP5	0.01	-0.01
PP6	-0.03	-0.02
PP7	0.01	0.01
PP8	0.02	0.00
PP9	0.01	0.01
PP10	-0.02	-0.01

While the plot of the PPs helps us to understand the behavior of the single risk sources we further investigated the performance statistics of individual PPs. We calculated the mean return, three measures of risk and the Sharpe Ratio [2] for all three risk measures. In Chapter 5, we tested the properties of the ASX200 index daily returns and found that the returns were skewed to the left and heavy tailed. Therefore, a risk measure that is based on the assumption of normality could be misleading. Pfaff (2013) pointed out that time series data of returns, particularly daily return series, are in general not independent and identically distributed (i.i.d.). For this reason, using the standard deviation will be inappropriate in assessing the

[2] Sharpe ratio examines the performance of an investment by adjusting the risk. It is a ratio measuring the excess return per unit risk (Sharpe, 1963).

risks. But we still report the standard deviation of each portfolio to enable comparisons. We used value at risk (VaR) (RiskMetrics Group, 1994) and expected shortfall (ES) (Artzner, 1999; Artzner et al., 1997) at the 95% confidence level as the measurement of portfolio risks. A definition of VaR is it is the smallest loss, in absolute value, such that

$$P(L > VaR) \leq 1 - \alpha \tag{7.2}$$

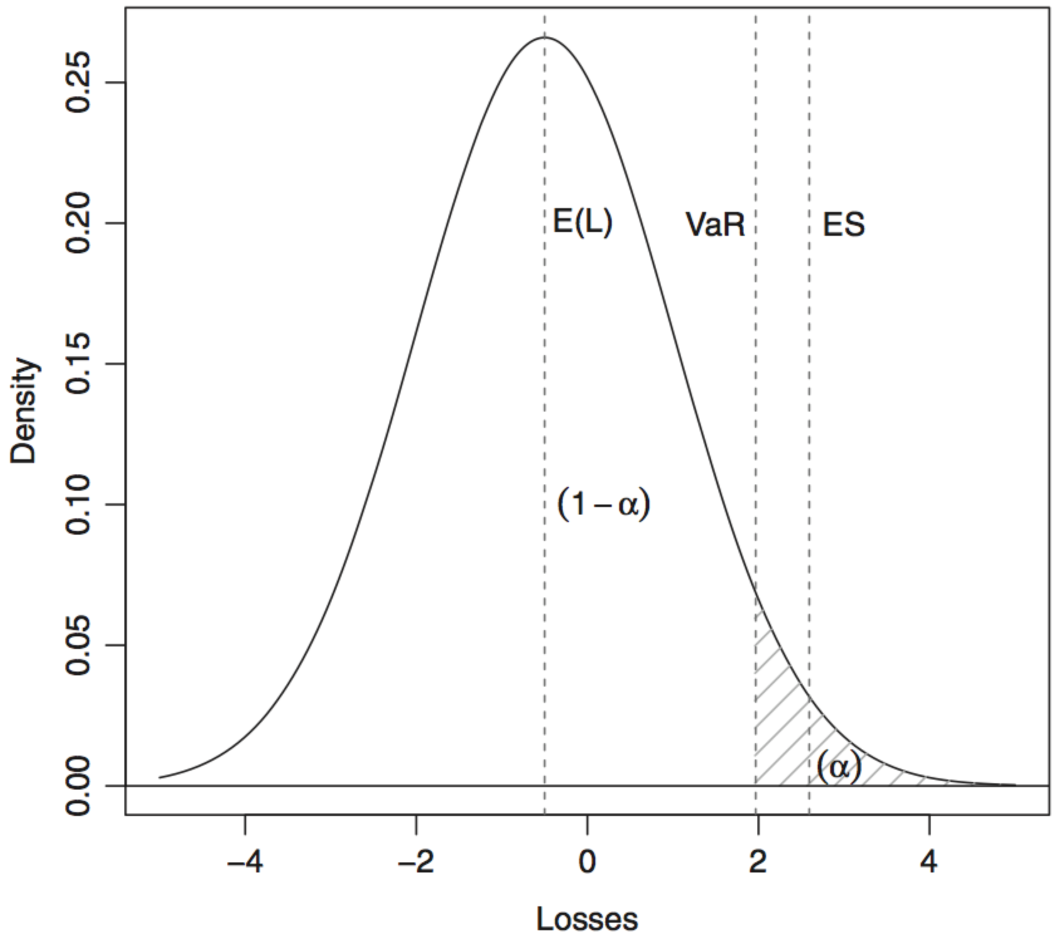
where α is the confidence level and L is the loss, measured as a positive number (Jorion, 2007). For example, if the one day VaR is 1% [3] at a confidence level of 95%, this suggests there is a 5% chance the loss will be greater than 1% in a day. ES is a complementary risk measurement to VaR that provides hindsight about the average size of loss when VaR has been violated for a given level of confidence (Pfaff, 2013). To help understand the concepts of VaR and ES we discussed above, we include an example of density of losses with VaR and ES as well as the expected loss in Figure 7.2.

The loss distribution is critical in computing VaR and ES. Recall that the daily return series are generally skewed and fat-tailed, and then risk measures that derived from the normal distribution assumption will underestimate the riskiness. So we used the modified VaR (mVaR) (Zangari, 1996) mES (Boudt et al., 2008), which directly deals with the non-normal returns.

In Table 7.3, the riskiness of PPs based on mVaR is in following order: PP2, PP1, PP4, PP3, PP9, PP8, PP5, PP10, PP7, PP6. Conversely, the riskiness of PPs based on mES is in following order: PP1, PP4, PP2, PP3, PP9, PP8, PP6, PP10, PP7, PP5. One should notice that riskiness of PPs is not in descending component order. This is because the eigenvalue of principal components does not necessary represent the riskiness of the portfolio constructed if the PCA was done on the correlation matrix. If we were based on a covariance matrix, then it would yield monotonically decreasing riskiness. There are some similarities of the riskiness order based on two risk measures. PP1, PP2 and PP4 were the three most risky portfolios among all 10 PPs. The order of PP3, PP9, PP8, PP10 and PP7 were the same

[3] VaR is reported as a positive number.

Fig. 7.2 Density of losses with VaR and ES.



on both risk measures. However, while the mVaR suggests PP2 was the most risky PP at 95% level, mES indicates PP1 was the most risky PP. The mVaR was 1.801% and 1.700% for PP2 and PP1 respectively. This means the chance for PP2 to have a loss over 1.700% (mVaR of PP1) was higher than 5%, the chance for PP1. Although PP2 has higher mVaR than PP1, it had lower mES, which suggests if the 5% chance happened and the portfolio loss extended beyond the mVaR, PP2 will have a smaller loss than PP1. Moreover, the PP6 and PP5 were the least risky PP based on mVaR and mES respectively.

We next compared the PPs with the 1/N and ASX200 index. Even though the PP1 had high price and return correlations with the 1/N, it was more risky than the 1/N portfolio on all risk measures. The mean return was also lower than the 1/N portfolio. On the other hand, PP1 was approximately as risky as the ASX200 index on all risk measures. Notably that PP1 was the only one which had a higher risk than the ASX200 index. All PPs were more risky than the 1/N portfolio measured by mES except for PP5. If based on mVaR, there were still half of the PPs more risky than the 1/N.

7.2 Allocation strategies comparison

After the behaviour and risk-return characteristics of the individual PPs were studied, we further investigated two allocation strategies, 1/N and ERC, based on the 10 PPs and 156 stocks respectively. We first compare the 1/N and ERC portfolios on the 156 stocks with the ASX200 index. In Table 7.3, the average daily return of 1/N and ERC were all about three times of the ASX200 index. However, the high return of the 1/N and ERC may be the result of the accumulated dividends paid. Recall that the returns of all the stocks were adjusted for dividends paid but ASX200 was not. The 1/N and ERC portfolios also had lower risks on all measures compared to the ASX200. This resulted in significantly higher Sharpe Ratio of the 1/N portfolio and the ERC portfolio. Moreover, the 1/N has performed better than the ERC based on mVaR and mES, in which it had a higher mean return and lower risk. The ERC strategy, which allocates risk based on the individual stocks, was not as diversified as the 1/N portfolio.

The ERC allocation strategy based on the underlying stocks can still have concentrated risk if most stocks have high correlations with each other. The performance of ERC portfolio is sensitive to the correlation between the stocks included. Recall the extreme case with all stocks perfectly positive correlated, then allocating an equal risk budget to all stocks is actually the same as holding one stock. Allocating a risk budget based on risk sources that are uncorrelated directly deals with this issue. We can see that when we applied the ERC to the 10 PPs, the risk of PPERC[4] was reduced significantly on all measures and the average return also dropped. But the Sharpe Ratio of PPERC was still higher than ERC. This suggests allocating risk budget based on uncorrelated risk is a better strategy than allocating risk based on individual assets from the point of view of diversification. Interestingly, an equal weighted portfolio based on 10 PPs performed better than the ERC on PPs. The Sharpe Ratio of PPEqual[5] was higher than PPERC on all three risk measures. PPEqual has a slightly higher mean return than PPERC and a lower risk according to mVaR. But the risk of PPEqual was higher than PPERC if based on mES.

When we compared the 1/N portfolio on stocks (1/N) with the 1/N portfolio on PPs (PPEqual), we found applying the 1/N allocation strategy on PPs effectively reduced the risk on all risk measures. Recall that the PP1, which represents the market risk, is essentially a 1/N portfolio on the underlying stocks. Unsurprisingly, 1/N had a risk higher than the PPEqual that allocated risk over 10 uncorrelated risk sources including the market risk. The Sharpe Ratio of PPEqual was higher based on mVaR and lower based on mES compared to the 1/N portfolio.

We further plot the performance trajectory for all portfolios over the last 12 years. Figure 7.3a presents the portfolio value of 1/N, ERC, PPEqual, PPERC and ASX200. Figure 7.3b shows only the portfolios constructed based on PPs and the ASX200. Figure 7.3a illustrates that two allocation strategies based on the underlying stocks performed better than those based on PPs. They moved closely relative to the ASX200 and look like an amplified version of the ASX200. The 1/N portfolio consistently outperformed the ERC. This is consistent with many other papers, which compared the allocation strategies available to

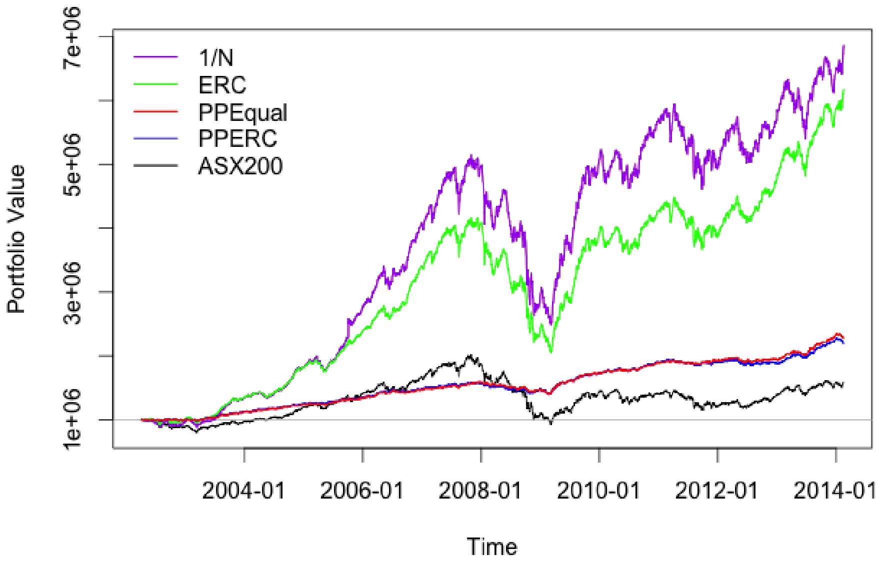
[4] A portfolio budgets equal risk to each 10 PPs.

[5] A portfolio has equal investment in each of 10 PPs.

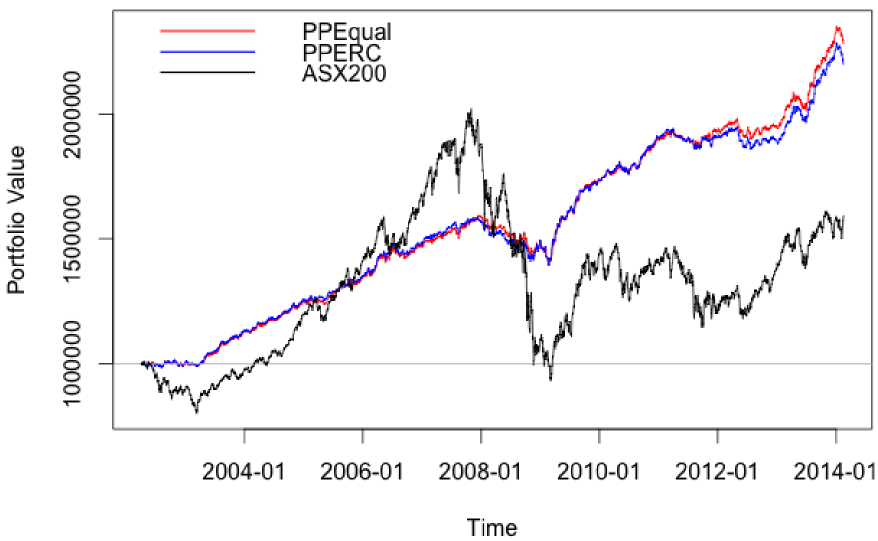
the portfolio selection based on individual stocks and reported the 1/N strategy dominated the others (DeMiguel et al., 2009; Kind, 2013; Lee, 2011). PPEqual and PPERC moved pretty much in sync until 2011 and started to diverge when PPEqual began to outperform the PPERC. The 1/N strategy not only outperformed the ERC based on stocks, it also had a better performance based on PPs.

Our results show constructing portfolios based on PPs are more diversified than portfolios based on the underlying stocks. In addition, the risk adjusted returns such as the Sharpe Ratio mostly increased but the average returns were substantially lower. The higher diversification did not result in better performance. This raises the question whether maximum diversification is desirable. Figure 7.3a shows portfolios constructed based on stocks resulted in portfolio values about three times of the portfolios based on PPs at the end of the study period. If one cares more about the portfolio value and returns, then constructing portfolios based on PPs was certainly not a better way to implement portfolio selection. But it worth noting that if one constructed portfolio based on the PPs, the risk would be decreased dramatically, the value increased gradually and one would less affected by the 2008 financial crisis than other portfolios considered (see Figure 7.3b) .

Fig. 7.3 The performance trajectory of 1/N, ERC, PPEqual, PPERC and the ASX200 index. The 1/N and ERC are portfolios of equal weighted and equal risk contribution of the 156 stocks. The PPEqual and PPERC are portfolios of equal weighted and equal risk contribution of the 10 PPs.



(a) Equal weighted and ERC portfolios on stocks and PPs.



(b) Equal weighted and ERC portfolios on PPs.

Table 7.3 The performance statistics for ASX200 index, 1/N (equally investment in 156 stocks), ERC (equal risk contribution in 10 PPs), PPEqual (equally investment in each 10 PPs), PPERC (a portfolio budgeting equal risk to each 10 PPs) and PP_n (principal component mimicking portfolio). All measures are in daily basis.

Statistics(%)	**ASX200**	**1/N**	**ERC**	**PPEqual**	**PPERC**	**PP1**	**PP2**	**PP3**	**PP4**	**PP5**	**PP6**	**PP7**	**PP8**	**PP9**	**PP10**
Mean	0.021	0.069	0.064	0.028	0.027	0.048	0.040	0.046	0.012	0.017	0.003	0.019	0.024	0.039	0.034
Std. Dev.	1.061	0.983	0.794	0.284	0.293	1.048	1.193	0.935	1.065	0.740	0.518	0.616	0.736	0.940	0.604
mVaR 95%	1.706	1.195	1.285	0.398	0.600	1.700	1.801	1.397	1.688	1.009	0.857	0.894	1.188	1.364	0.896
mES 95%	3.060	1.195	2.600	0.558	0.418	3.176	2.720	2.024	2.751	1.055	1.526	1.211	1.806	1.830	1.225
Sharpe Ratio (Rf=0%, p=95%)															
Std.Dev.	1.993	7.017	8.030	9.777	9.060	4.599	3.330	4.913	1.083	2.326	0.500	3.033	3.207	4.169	5.656
mVaR	1.240	5.768	4.959	6.974	6.358	2.838	2.206	3.286	0.698	1.707	0.302	2.090	1.987	2.873	3.812
mES	0.691	5.768	2.452	4.980	4.431	1.518	1.460	2.269	0.419	1.632	0.170	1.544	1.307	2.142	2.787

Chapter 8

A study of principal component one

The financial markets have become more integrated during market crashes (Billioand et al., 2012; Fenn et al., 2011; Kritzman et al., 2011; Zheng et al., 2012). Securitization and more complex markets, have prevented us from directly observing the many linkages within the financial markets (Kritzman et al., 2011; Maclean and Nocera, 2010). As a consequence of the financial crisis in 2008, it was urgent to develop tools to monitor systemic risk, the risk that is associated with the whole financial system, for the use of both investors and regulators. We note that systemic risk is not systematic risk (see Chapter 3). Systemic risk indicates ratio of systematic risk to idiosyncratic risk. If the systemic risk increases, there is a higher proportion of risk in the total risk that is systematic risk. This suggests the amount of diversifiable risk decreases.

The PCA was one of the most commonly used methods to develop measures of systemic risks. The idea behind the use of PCA is simple. A PCA transforms the original interrelated assets into uncorrelated principal components with the first component explaining most of variation with the subsequent components explaining as much as possible of the remaining variation. If a set of assets are uncorrelated, the principal components would be exactly equivalent to the original data set. The eigenvalues will be the same and equal to 1 in the case of a correlation matrix. Conversely, when a set of assets are more connected, the first few principal components will have higher eigenvalues and therefore capture more variation. Therefore, the variance explained by the first few principal components can be used as a measure of the level of systemic risk. The choice is how many principal components should be used. Billioand et al. (2012) used the first two eigenvalues to detect the systemic risk in the financial industry. Kritzman et al. (2011) reported the number of principal components they used was approximately 1/5th of the number of assets and in their case this was 10 principal components. Fenn et al. (2011) and Zheng et al. (2012), on the other hand, used only the first principal component as an indicator of systemic risk.

We followed Fenn et al. (2011) and Zheng et al. (2012) and used the variance explained by the first principal component as a measure of the level of systemic risk. The first principal component has been understood as the market component in a stock market, the subsequent principal components are representative of group risks. It is reasonable to use the first principal component alone because the systemic risk is associated with the whole financial system rather than with some part of the financial system. Note that we only study the stock market, and we assume the stock market is a good proxy for the financial system as a whole.

One problem of using more principal components is that the increase in variance explained by principal component 1 can be offset by subsequent principal components explaining less variation or the other way around. Fenn et al. (2011) analyzed the variance explained by the first five principal components individually for the period 2001 to 2010 on 98 financial products, including 25 developed market equity indices, three emerging market indices, four corporate bond indices, 20 government bond indices, 15 currencies, nine metals, four fuel commodities, and 18 other commodities. They reported that the variance explained by first principal component decreased when the variance explained by second and third component increased. Likewise, they also found that when the variance explained by the first principal component increased, the variance explained by principal component two and three both decreased. If the variance explained by first three principal components was used to measure the level of systemic risk, the change in the variance explained will not disappear but will definitely be less obvious. As a consequence, using the variance explained by more than one principal component can be misleading.

A rolling window approach was applied in our estimation process. We performed a PCA on a window size of two years (equivalent to 504 trading days) at weekly intervals.

This resulted in 602 data points of variation explained by principal component one over the whole study period. We compared the variance explained by principal component one to the ASX200 returns which were also calculated using a rolling window approach with window size two years, the ASX200 index value, the level of diversification and KMO measure of sampling adequacy (more information about KMO measure of sampling adequacy, see Chapter 4). We separate our discussion of variance explained by principal component one into two sections. Section 8.1 compares the variance explained by principal component one to the ASX200 index value and returns. Section 8.2 discusses the variance explained by principal component one with the diversification ratio and KMO measure of sampling adequacy.

8.1 Systemic risk VS. ASX200 index value and return

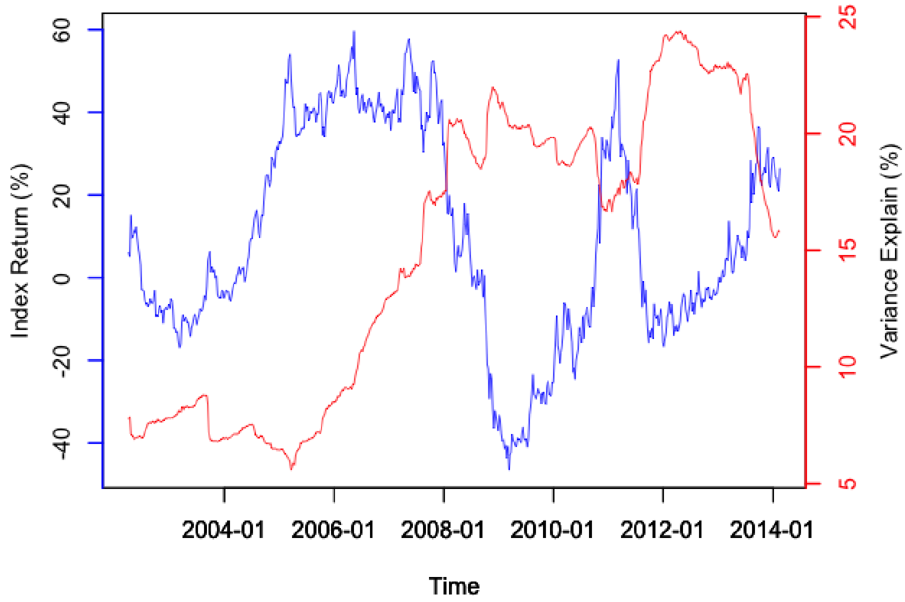
Kritzman et al. (2011) estimated the variance explained by the first 10 principal components, to measure the level of systemic risk, in a rolling window of 500 days based on the returns of the 51 US industries in the MSCI USA index. They called this measure the absorption ratio. They reported a coincident relationship between the absorption ratio and MSCA USA price index. We have a similar finding of a coincident relationship between the systemic risk and ASX200 index (both index value and index return), for the Australian market.

In Figure 8.1a, the variance explained by principal component one together with the ASX200 index return is presented. The percentage variance explained by principal component one has its lowest level around 2005 and at the same time the index return reached its first peak. After this, the percentage variance explained climbed almost monotonically (about 300%) for three years and reached a peak in 2008. This is when the global financial crisis affected the financial markets intensely. The high variance explained by first principal component implies that there was a large amount of common variation in the stock market. This suggests that long before the financial crisis happened, the market had become more closely connected. When the market was tightly coupled towards the end of 2007, a trigger caused a catastrophe. The influences of the crisis have been more serious than one could have expected. Alan Greenspan, who was the chairman of the Federal Reserve until 2006, admitted that he had put too much faith in the self-correcting power of free markets and had failed to anticipate the meltdown in US financial market. He called the financial crisis in 2008 a “once-in-a-century credit tsunami" and said it had “turned out to be much broader than anything I could have imagined" [1]. However, this may not have been entirely unexpected if one had recognized that the market had become more fragile than it had been historically and negative shocks would propagate more quickly and broadly. The index returns stayed high when the variance explained increased from 2005 to 2008. Once the variances explained reach the first peak, the index returns started to drop significantly. The market remained tightly integrated until the beginning of 2010, even when index returns started to recover from the beginning of 2009. This suggests that the Australian stock market was still extremely fragile and therefore vulnerable to negative shocks. There was a small drop of variance explained at the end of 2010 and the index returns reached its second peak. The variance explained stared to rise again at the end of 2011 and reached highest point during the study period at beginning of 2012. The reason for this increase of variance explained was the European sovereign debt crisis, and fears over the global economy. The worries about the global economy have made the Australian stock market even more fragile than it was in 2008. Perhaps the investors psychologically lacked confidence after they had gone through one crisis in 2008. At the end of our sample period, the variance explained by principal component one has decreased significantly and the index returns have partially recovered from the drop in late 2011.

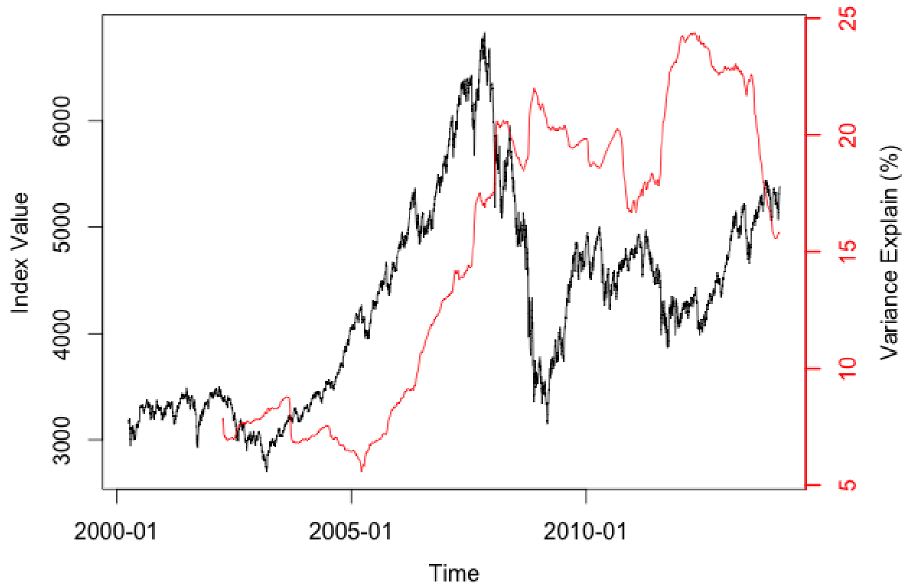
Figure 8.1b plots the variance explained by principal component one alongside the ASX200 index value. As we can see, the formation of more connected market initially did not affect the index value. But when the variance explained reached its high level in 2008, the index value dropped dramatically. Interestingly, the peaks of variance explained and the lowest point of the index value happened about the same time. Our results in the Australian market supports the observations in many papers (Fenn et al., 2011; Kritzman et al., 2011; Zheng et al., 2012) that systemic risk increased steadily in the years before 2008. We also

[1] http://www.theguardian.com/business/2008/oct/24/economics-creditcrunch-federal-reserve-greenspan

Fig. 8.1 Variance explained by principal component one and the ASX200 index price and returns. Both the variance explained by principal component one and ASX200 index returns were calculated weekly using a rolling window size of two years (equivalent to 504 trading days). The index returns (two years return) were calculated using $R_{ASX200}(t) = \frac{P_{ASX200}(t+504) - P_{ASX200}(t)}{P_{ASX200}(t)}$.



(a) Variance explained by PC1 with the ASX200 index returns. Period: April 2002 to February 2014.



(b) Variance explained by PC1 with the ASX200 index value. Period: April 2000 to February 2014.

found an increase of systemic risk in Australian stock market around the end of 2011, which relates to the European sovereign debt crisis. This was consistent with the study of systemic risk in the European market Zheng et al. (2012). Moreover, Kritzman et al. (2011) studied the global market for the period from February 1995 to December 2009 and reported that the absorption ratio (variance explained by first 10 principal components in their case) increased significantly in October 1997 (Asian financial crisis), August 1998 (the Russian and LTCM collapses), mid-2006 (housing bubble) and September 2008 (Lehman Brother default). Our research has only included two market drawdowns. That is not sufficient to draw any conclusion of how the variance explained by principal component one was related to the market drawdowns. However, based on the similar testing framework to many other papers, we are able to draw a conclusion from our findings together with others, the variance explained by principal component one is a leading indicator of the financial crisis.

8.2 Systemic risk VS. Diversification

We also plotted the variance explained by principal component one against the diversification ratio. The diversification ratio is a measure of the degree of diversification for a long only portfolio introduced by Choueifaty and Coignard (2008). The diversification ratio for a portfolio is defined as

$$DR_{\omega \in \Omega} = \frac{\omega' \sigma}{\sqrt{\omega' \Sigma \omega}} \tag{8.1}$$

where Σ is the variance-covariance matrix of the returns for N assets, σ is the vector of asset volatilities measured by their respective standard deviations. ω is weight vector of the portfolio. The numerator of the diversification ratio is then the weighted average volatility of the individual stocks and the denominator is the portfolio standard deviation. By this definition, the higher the diversification ratio, the better the degree of diversification is. If a portfolio is completely non-diversified, in the case of single- asset portfolio, the diversification will achieved its lower bound of 1.

Many researchers have reported that markets offer less diversification in a falling market (Billioand et al., 2012; Cappiello et al., 2006; Ferreira and Gama, 2004). Conventional wisdom has it that when the market is more connected, there are less diversification benefits to exploit. Instead of comparing different periods in history, which may implicitly assume the diversification effects are stationary within the same period, we applied a rolling window approach to study how the degree of diversification has changed throughout the study period. Because the purpose in this part of research is to study how the potential for diversification has changed over time with the systemic risk, not to compare how different allocation strategies result in different degrees of diversification, we only present the portfolio with the most simple allocation strategy, 1/N. We note that the other allocation strategies have followed the same trend as the diversification ratio of 1/N portfolio over time (see Appendix C).

The financial crisis in 2008 made investors wonder what went wrong with their portfolios that they believed to be diversified (Lee, 2011). The explanation for this can be easily seen in Figure 8.2. The stock market had become more connected long before the bursting of the financial crisis. The diversification ratio almost monotonically decreased with the increasing in the market connectedness. The variance explained by principal component one reached its first peak followed by the diversification ratio dropping to its lowest point in our study period. This suggests that even if you were holding the same portfolio, it would not be as diversified as it was at other times. The higher level of variance explained by principal component 1 indicates more systemic risk, which means a higher ratio of systematic risk to the idiosyncratic risk. When the amount of non-diversifiable risk reached its first peak during the 2008 financial crisis, a portfolio which held the same stocks had the least diversification benefits available to exploit than it had any other time in our study period.

Between the financial crisis in 2008 and the market drawdown in late 2011, the variance explained dropped a little and the market diversification went up to 3.75, but this was still 40% lower than it was during 2002 to 2005. If we go back to Figure 8.1, we found that the index return and value recovered at the same time as the diversification ratio went up. Moreover, when the variance explained by principal component one rose again at the end of 2011, the diversification ratio dropped. It is interesting that even at the end of 2011, the variance explained by principal component one rose to its second peak and was higher than it was in 2008, the diversification ratio, on the other hand, was not lower than in 2008.

This higher diversification ratio than in 2009 immediately after the 2008 financial crisis was coincident with the higher index returns and value in the late 2011 drawdown than after the 2008 financial crisis. This again raises the question that was posed in the last chapter whether the higher diversification brings better performance. Comparing the two market drawdowns which happened within our study period we found that the degree of loss in ASX200 index was more related to the level of potential diversification than to the level of systemic risk. A higher diversification ratio in the 2011 market meltdown compared to the 2008 financial crisis resulted in a relatively smaller loss. Conversely, the higher level of systemic risk in late 2011 compare to what it was in 2008 was not consistent with the relatively smaller loss.

In Chapter 4, we discussed the use of the KMO measure of sampling adequacy to test the degree of common variation among stocks. This gives us another tool to assess the market connectedness and therefore the potential for diversification. We also applied a rolling window approach to estimate the KMO statistic over time. Figure 8.3 presents KMO statistics from 2002 to 2014 together with the variance explained by principal component one. Higher level of KMO statistic indicates more common variation among stocks and higher level of systemic risk. This suggests less potential of diversification. We can see that KMO statistics and the variance explained by principal component one evolved closely over time. The KMO measure of sampling adequacy relates to the ASX200 index and diversification ratio in the same way as the variance explained by principal component one. We conclude that KMO measure of sampling adequacy is also an effective measurement of the level of systemic risk.

The variance explained by principal component one has effectively assessed the level of systemic risk. We have seen that it is a leading indicator of a financial crisis. The KMO measure of sampling adequacy has shown a close relationship to the variance explained by principal component one over time. Using either KMO measure of sampling adequacy or variance explained by principal component one to monitoring systemic risk would have the same insightful results.

Fig. 8.2 Variance explained by principal component one and the diversification ratio. The variance explained by principal component one was calculated weekly using a rolling window size of two years (equivalent to 504 trading days). The diversification ratio was calculated using $DR_{\omega \in \Omega} = \frac{\omega' \sigma}{\sqrt{\omega' \Sigma \omega}}$.

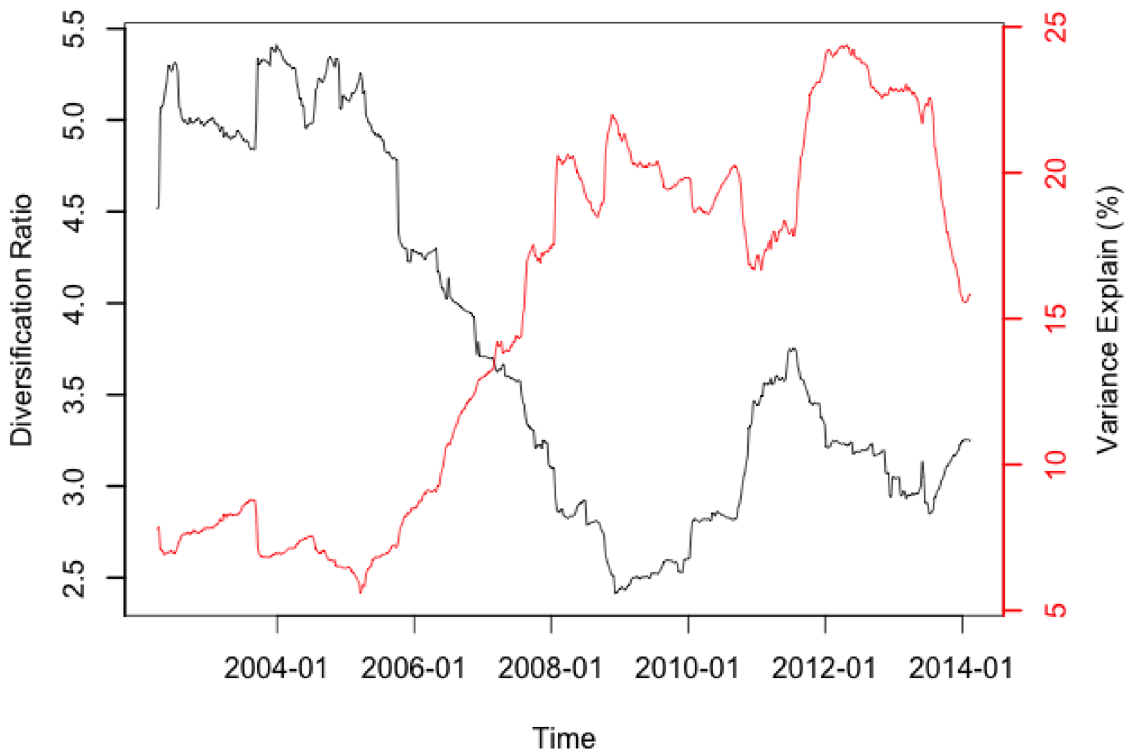
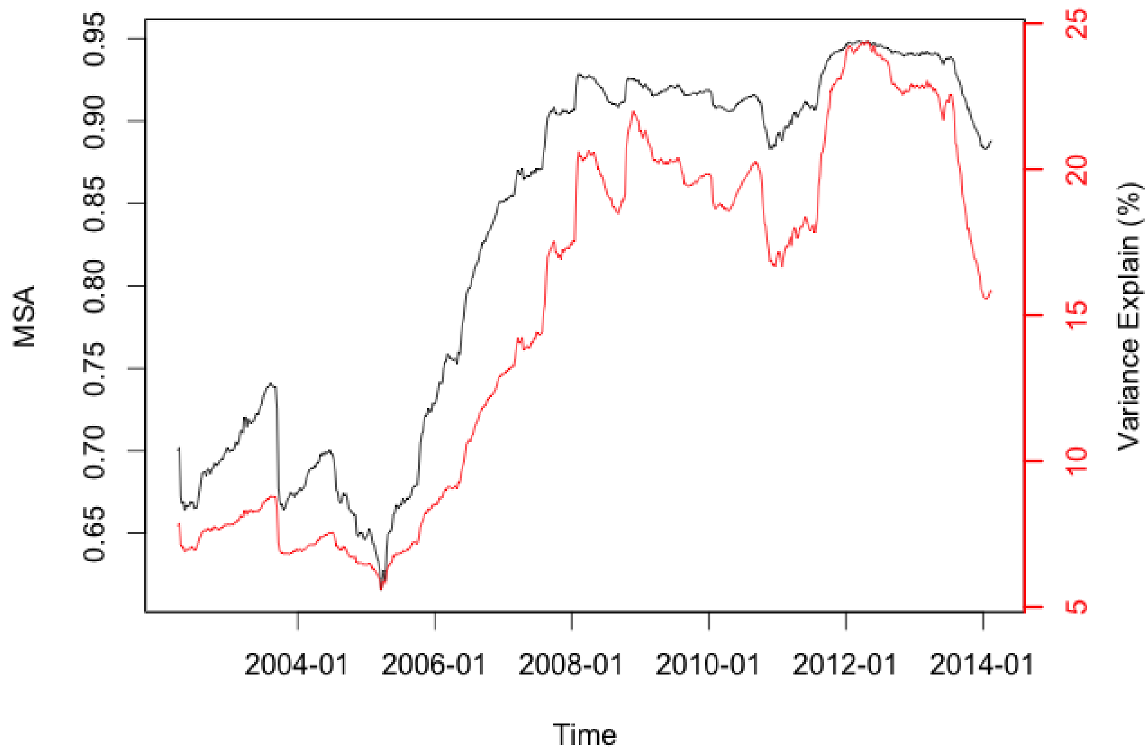


Fig. 8.3 KMO measure of sampling adequacy for 156 stocks and the variance explained by principal component one. Both measures were calculated weekly using a rolling window approach with window size two years (504 trading days).



Chapter 9

A study of principal component two

In Chapter 7, we constructed PPs which represented uncorrelated risk sources inherent in the stock market. The weight of each stock in the PPs was based on its coefficient in the eigenvector. The coefficients in the principal components are not persistent through time. With the change of the correlation of the stocks, the relative importance of each stock in the principal component is likely to change. We have calculated the performance of PPs for the last 12 years, the portfolios were rebalanced daily according to the change in the coefficients (see Chapter 7). In this chapter, we further investigate the time evolution of stock coefficients of the principal components and test whether the change of the coefficients provide a signal for a portfolio manager to trade. Moreover, we compare dynamic PPs in which the weights were rebalanced daily to the static principal portfolios that kept the weights unchanged. However, we will only include the study of principal component two because we have found it shows the most interesting results among the 10 principal components retained. For a brief discussion of other principal components, see Appendix D.

Recall the bi-plots in Figure 5.3, which were constructed using the whole study period in the investigation, we found that principal component two showed grouping of industries. While Basic Materials, Oil & Gas and Consumer Goods had positive coefficients, Financials, Health Care and Consumer Services had negative coefficients. Stocks in the Industrials group had coefficients that are both positive and negative but were close to zero. Moreover, we reported that PP2 has the highest price correlation with ASX200 index among the retained 10 principal components (see Table 7.1) and is possibly a representation of a risk related to the state of the economy (see Chapter 7).

In order to study the time evolution of coefficients in principal component two, we performed PCA daily with a window size of two years (504 trading days) to extract the coefficients and presented these using a heat map. Figure 9.1 shows coefficients of 156 stocks in principal component two together with ASX200 index value from 4 April 2002 to 17 February 2014. The horizontal axis is time in units of one trading day. The vertical axis is stocks for the heat map and prices for the ASX200 index. The stocks are sorted based on the first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 (the list of stocks is shown in Appendix E). Despite the fact that there are frequent changes of the coefficients, we observe patterns in the evolution of coefficients in principal component two. The first pattern is relative to the ASX200 index value. There are two major structures in the coefficients separated by the financial crisis in 2008. Before the price drawdown in late 2007, the reds, which indicate negative coefficients, are mostly in the bottom and the yellows, which indicate positive coefficients, are in the top. Conversely, when the index value started to collapse, the reds change to the top and the yellows in the bottom. Essentially, before the financial crisis in 2008, principal component two is short in Financials, Health Care and Consumer Services and long in Basic Materials, Oil & Gas and Consumer Goods. When the financial crisis started to burst, the coefficient structure in principal component two changed to opposite, that is, a long position in Financials, Health Care and Consumer Services and a short position in Basic Materials, Oil & Gas and Consumer Goods. The Industrial groups continue to have small positive and negative coefficients. This is also shown in the biplots of the coefficients of each stock in Figure 9.2. While the heat map offers an overall observation of the industry groups, the bi-plots provide more detailed information regarding to the coefficients of stocks with industry for a single rolling window. We present the two sets of coefficients in the middle point of the heat map (the fifth black vertical line) where two structures break to represent the two structures we discussed above. One is the last day of first structure and the other is the first day of the second structure. Figure 9.2a presents the coefficients result from the study period 22/09/05 to 19/09/07 and Figure 9.2b is for the study period of 23/09/05 to 20/09/07. We can see that when 20/09/07 entered into the rolling window, the magnitude of the coefficients remained the same while the sign of all coefficients changed to the opposite in principal component two. We note that even when we used these two sets coefficients as representative of the two structures, we do not assume that within each structure, the magnitudes of the coefficients remain the same over time. The fact is that the magnitudes of the coefficients are relatively the same for more closer rolling windows. For rolling windows that are in different periods, and result in the same sign of the coefficients, the relative importance in the components may vary. To illustrate this point, we present the coefficient and it absolute value of a single stock in principal component two across time. In Figure 9.3a, the sign of the coefficients were changing from positive to negative or the other way around from time to time, but the magnitudes of the coefficients were relatively the same for closer rolling window. Figure 9.3b shows that the magnitudes of coefficients were relatively the same for the first three years and decreased to a new level in 2006 then remained there until end of 2010. There was a sudden jump of the magnitude of the coefficients in 2011 but soon returned to a level slightly higher than it was during period 2006 and end of 2010 and remained there until the end of the study period.

This brings us to the second pattern we have found in Figure 9.1. We noticed that the colors in the heat map are brighter in the first few years, which indicates larger absolute values of the coefficients. We then produced another heat map of the square of all coefficients and sorted by industries (the list of stocks sorted by industry is shown in Appendix F). Recall that in PCA, the square of the coefficients for any principal component will sum to 1. This suggests that the square of the coefficients will well represent the relative importance of each stock in the component. In Figure 9.4, industries are separated with the horizontal dashed lines. It is clear that the time evolution of the square of coefficients can be divided into three parts. The first part shows that the Financial stocks dominate the principal component two and this lasted until the end of 2005. The second part starts from around 2006 and ends at 2011. The stocks in Financial industry stopped being dominant stocks in the component. The square of the coefficients universally shows small values, which means no industries have significant contribution to principal component two. However, we find that stocks in Basic Materials, Financials, and Oil & Gas have relative higher values than other industries. Moreover, some stocks in the Basic Materials have the highest square of coefficients. In the last part, which started from 2011, the stocks in Basic Materials were even more important in principal component two while the Oil & Gas became insignificant. The Basic Materials basically dominated the principal component two. This suggests that there are actually three different structures inherent in the evolution of principal component two coefficients rather than two structures we found in Figure 9.1 alone. The period starting from around 2011 to the end of our study period should be separated and not mixed with the structure shown during the 2008 financial crisis.

We next turn to test whether the changes of the coefficients signal a portfolio manager to trade. We choose several changes that happened at different times within our study period and each change happened in one day. Going back to Figure 9.1, there are eight vertical lines in the heat map, which indicates the eight scenarios we chose. For each scenario, we constructed portfolios based on the coefficient before and after the change respectively and compared the 12 month out-of-sample performance together with the index value. The eight time points we chose covered the periods in all three structures we discussed above. Figure 9.5 presents the test results for the eight out of sample tests. Note that we call the coefficients structure with the negative coefficients in the bottom in Figure 9.1 the "structure 1" and the positive coefficients in the bottom the "structure 2". Therefore before the price drawdown in late 2007, principal component two mainly showed structure 1 and changed to structure 2 when the price starts to decrease significantly base on Figure 9.1. All the graphs in Figure 9.5 indicate the portfolio based on the coefficients before the change (blue line) and portfolio based on the coefficients after the change (red line) move in opposite directions. This is because the magnitudes of the coefficients are the same or very similar and the only difference is the sign of the coefficients. For example, the tests in Figure 9.5e are based on the coefficients in Figure 9.2.

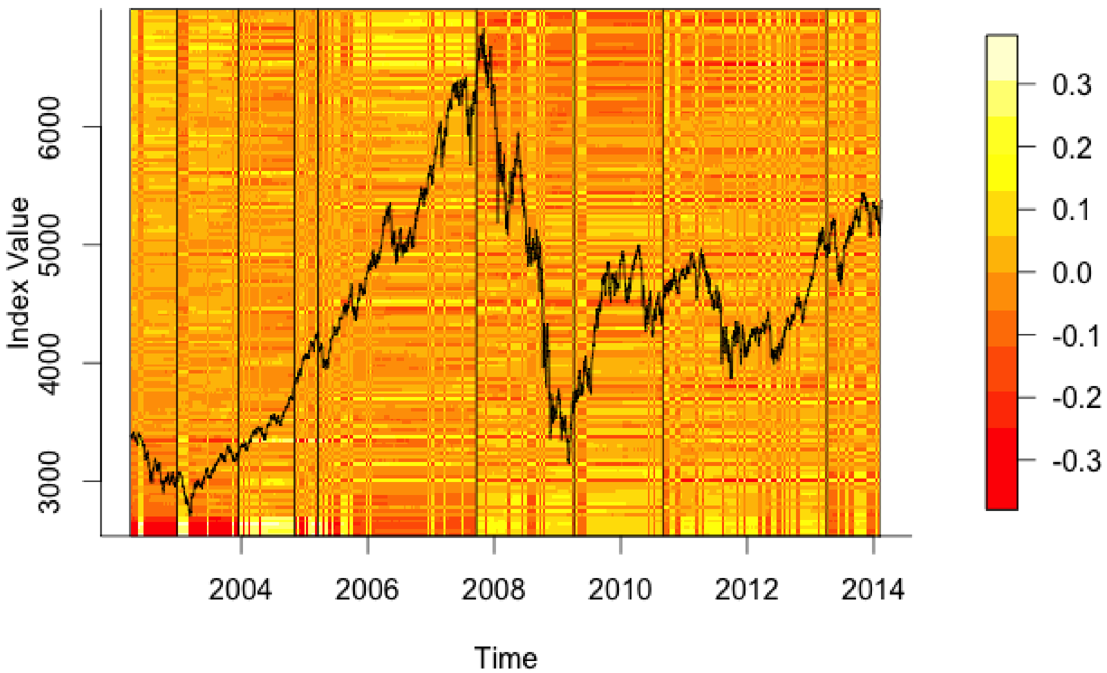
We first look at the test of the first four scenarios that occurred before the price drawdown in late 2007. Figure 9.5a shows the change from structure 1 to structure 2 when 27/12/02 entered into the rolling window. The portfolio based on structure 2 (after change) outperformed both the index and portfolio of structure 1 (before change) for about six months. Especially in the first two months, the ASX200 index portfolio dropped to about \$0.9 million and the before change portfolio decreased to around \$0.94 million while the after change portfolio value increased to \$1.06 million. Recall that PP2 has the highest correlation with the ASX200 index (see Table 7.1) and before the financial crisis in 2008 the market was mostly in structure 1. This means forming portfolio based on structure 1 coefficients should follow the ASX200 index value and portfolio based on structure 2 coefficients, on the other hand, move opposite to the index. After the change in scenario 1, there was a big drop in the index value for about two months (see Figure 9.1). This is the reason that after change portfolio increased substantially for the first two months. When the index recovered from the drop in the value, the after change portfolio no longer outperformed the index and the before change portfolio. We have similar findings in Figure 9.5d, in which the after change portfolio (structure 2) outperformed the index and the before change portfolio when there was a immediate decline in the ASX200 index. Figure 9.5b also shows a structure 1 change to structure 2. The ASX200 index increased gradually over the 12 month out of sample, the before change portfolio outperformed the after change portfolio. Conversely, Figure 9.5c presents structure 2 change to structure 1. The after change portfolio, which is the portfolio of structure 1, moved closely with the index for most of the time. Even at the end of the 12 month period, the drop in the index value caused the after change portfolio to have a substantial loss and the before change portfolio increased in an abnormal manner, we conclude that the structure 1 portfolio (after change portfolio) still performed better than the structure 2 portfolio as the other three scenarios.

We next look at the periods of structure 2 (starting from the price drawdown in late 2007). Figure 9.5e presents the change that happened during the global financial crisis (also see Figure 9.2). The structure 2 portfolio in this case is the one that followed the ASX200 index since the structure has changed. Obviously, the index has fallen due to the crisis and the structure 2 portfolio (after change portfolio) consistently under performed the structure 1 portfolio (before change portfolio) until the end of the 12-month period.

While the index value decreased and remained steady, the structure 2 portfolio (after change portfolio) recovered from the decrease in contrary to the decrease of structure 1 portfolio value. Figure 9.5f is also a test of structure 1 change to structure 2, in which the after change portfolio is the one more related to the index compared to the before change portfolio. However, we found that the after change portfolio did not follow the increase of index value in July 2009. It stays at around $1.1 million for nearly six months and back to track the index in the remaining time. Both Figure 9.5g and Figure 9.5h are tests of structure 2 change to structure 1. More importantly, neither the structure 1 portfolio nor the structure 2 portfolio had a high correlation with the index. These two scenarios are in the periods of the third structure shown in Figure 9.4. We believe that this was a period of transition in which the sign of the coefficients in the principal component two changed more often than other periods. Despite the fact that it was the Basic Materials group which consistently dominated the component during this period, its positions long and short are inconsistent over time. As a consequence, to better follow the ASX200 index, one should change the portfolios holding based on the coefficient change. This is the reason that neither the before change portfolio nor the after change portfolio has high correlation with the index.

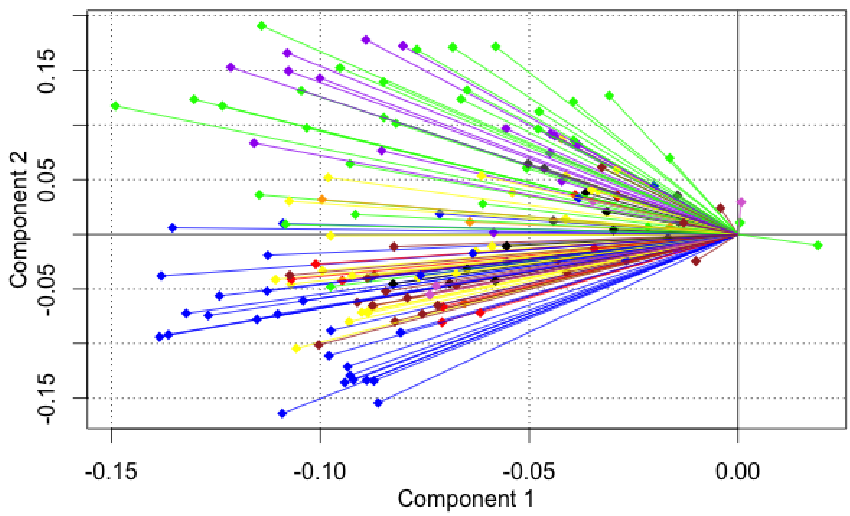
We conclude that it shows there is no benefit for a portfolio manager to trade based on the coefficient change. Even when there are frequent changes over time, principal component two is still described by the major structure, short in Financials, Health Care and Consumer Services and long in Basic Materials, Oil & Gas and Consumer Goods before 2008 financial crisis and short position in Basic Materials, Oil & Gas and Consumer Goods and long position in Financials, Health Care and Consumer Services when the crisis started to burst. This is demonstrated by the fact that it is the portfolio that was constructed based on the major structure in relevant period follows the ASX200 index. For example, before the price drawdown in late 2007, structure 1 portfolio followed the index. In the period of 2008 to 2011, while the structure 1 portfolio moved in opposite to the index, the structure 2 portfolio closely tracked the index. Trading based on the coefficient changes is considered necessary when one wants to replicate a portfolio that has high correlation with the index over time. As the cases after 2011, one can only have portfolio highly correlated to the index

Fig. 9.1 Time evolution of stocks coefficients in principal component two with the ASX200 index value. The coefficients are obtained from daily performing PCA on 156 stocks for two years of data (504 trading days). The horizontal axis are times from 4 April 2002 to 17 February 2014. The vertical axis is the 156 stocks. The stocks are ranked based on the first day's coefficients, which was obtained from the first rolling window, 3 April 2000 to 3 April 2002 (the list of stocks with their industry is shown in Appendix E).

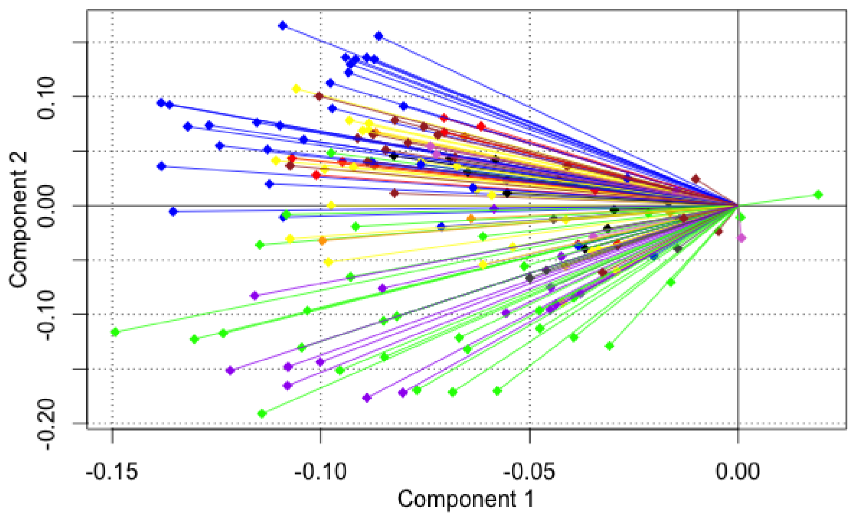


by frequently changing its position based on the coefficient changes.

Fig. 9.2 Bi-plots of relative weights of each stock in principal component one and two for the study period of 22/09/05 to 19/09/07 and 23/09/05 to 20/09/07, with respect to the Industry Classification Benchmark. The colours correspond to ICB sector classification: Financials are Blue (33 stocks), Health Care are Red (9 stocks), Industrials are Yellow (24 stocks), Consumer Services are Brown (19 stocks), Basic Materials are Green (31 stocks), Oil & Gas are Purple (16 stocks), Utilities are Orange (5 stocks), Consumer Goods are Black (9 stocks), Telecommunications are Orchid (4 stocks), Technology are Grey (6 stocks).

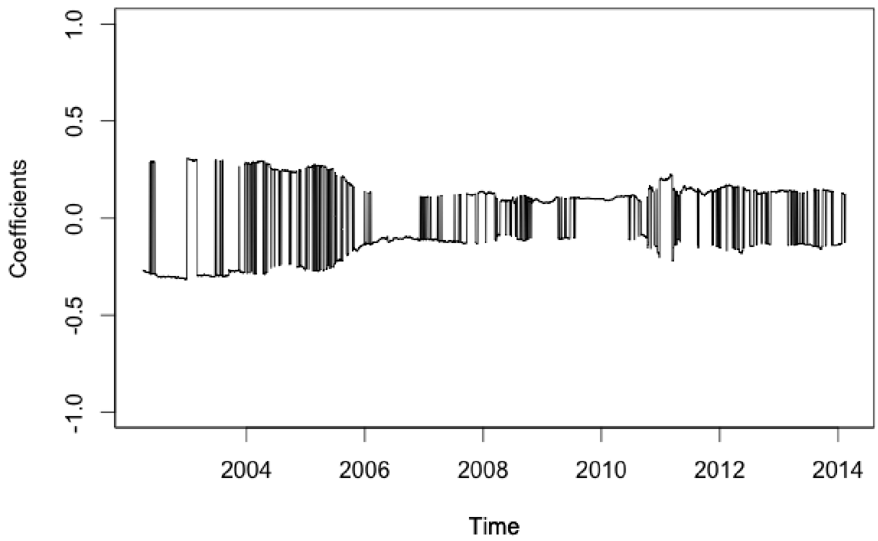


(a) Study period of 22/09/05 *to* 19/09/07.

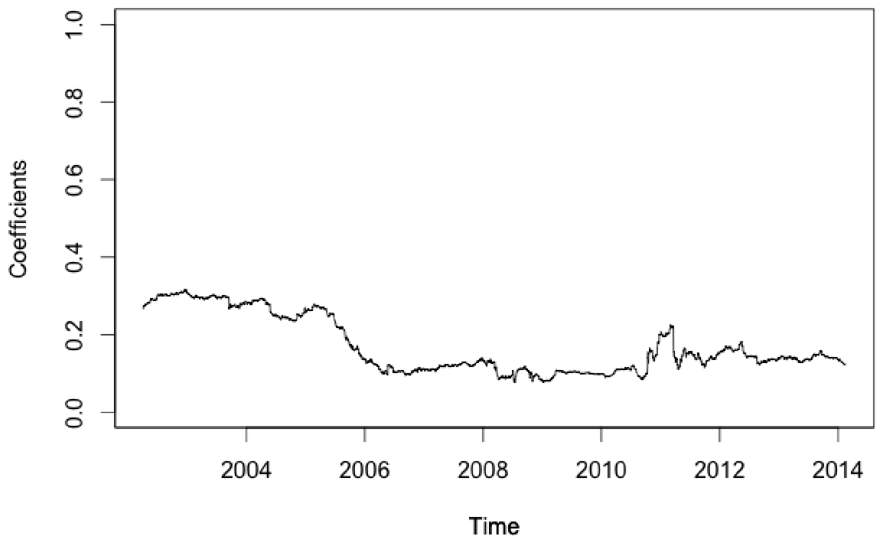


(b) Study period of 23/09/05 *to* 20/09/07.

Fig. 9.3 Time evolution of GPT coefficients in principal component two .The coefficients are obtained from daily performing PCA on 156 stocks for two years of data (504 trading days).



(a) Coefficients of GPT in principal component two across time.



(b) Absolute value of coefficients of GPT in principal component two across time.

Fig. 9.4 Time evolution of the square of the coefficients in principal component two. The coefficients were obtained from daily performing PCA on 156 stocks for two years of data (504 trading days). The horizontal axis are times from 4 April 2002 to 17 February 2014. The vertical axis are the 156 stocks ranked by industries (the list of stocks order by the industry is shown in Appendix F). The order of the industries are: Basic Materials (1-31), Consumer Goods (32-40), Consumer Services (41-59), Financials (60-92), Health Care (93-101), Industrials (102-125), Oil & Gas (126-141), Technology (142-147), Telecommunications (148-151) and Utilities (152-156). Industries are separated by the horizontal dash lines.

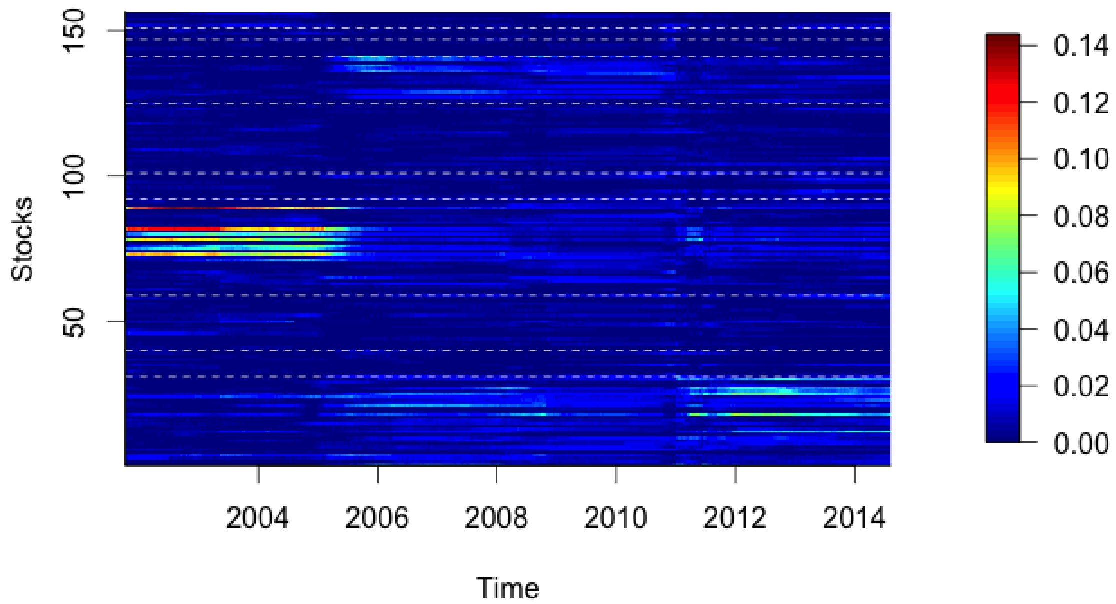
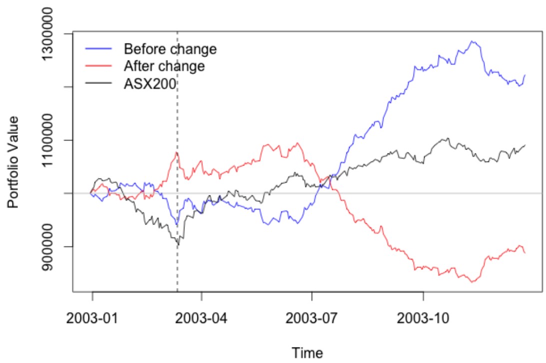
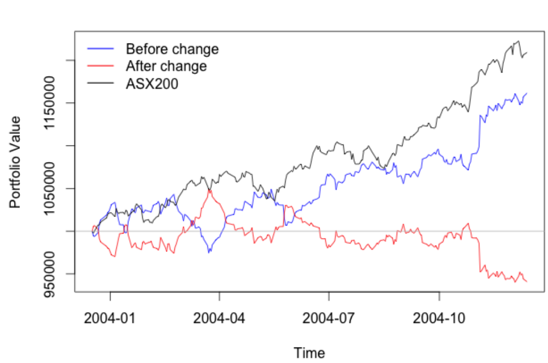


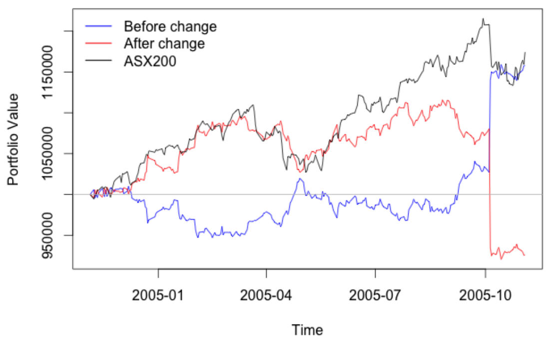
Fig. 9.5 12 month out of sample of tests of portfolios constructed based on coefficient changes in principal component two. There are eight scenarios selected and indicated by the black vertical lines in Figure 9.1. In each scenario, the change occurred in one day. The “structure 1" indicates the structure shown in Figure 9.1 when the negative coefficients are in the bottom of the heat map. The “structure 2" indicates the structure shown in Figure 9.1 when the positive coefficients are in the bottom of the heat map, opposite to structure 1.



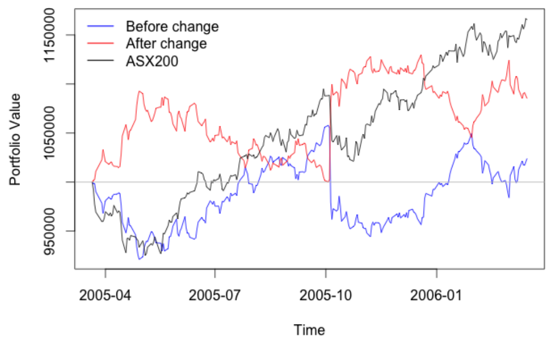
(a) Structure 1 change to structure 2 20/02/2007 to 20/02/2008 .



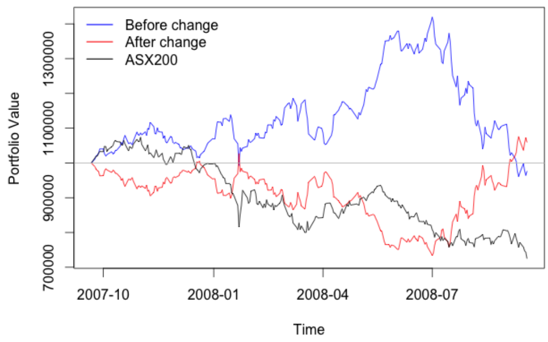
(b) Structure 1 change to structure 2.



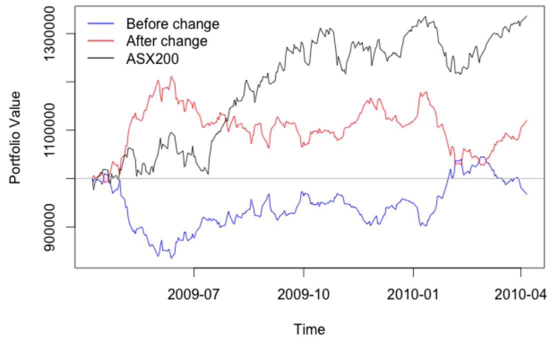
(c) Structure 2 change to structure 1.



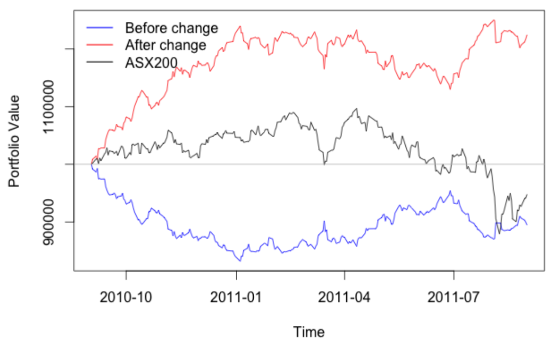
(d) Structure 1 change to structure 2.



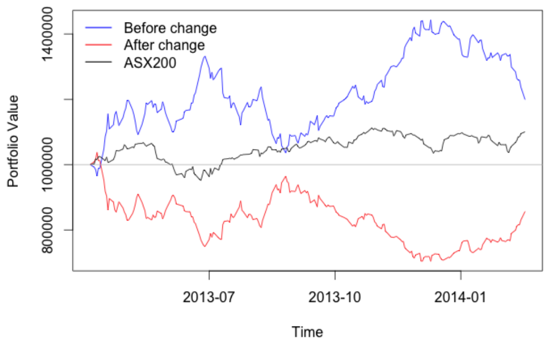
(e) Structure 1 change to structure 2.



(f) Structure 1 change to structure 2.



(g) Structure 2 change to structure 1.



(h) Structure 2 change to structure 1.

Chapter 10

How many stocks are needed to diversify a portfolio?

One way to reduce the dimensionality of a data set is to retain a much smaller number of principal components based on PCA. We have determined that the first 10 principal components are sufficient to describe a data set containing the 156 stocks in the ASX200 for which have complete data (see Chapter 5). In Chapter 7, we constructed portfolios based on each of the 10 principal components and treated them as individual investment assets. A combination of these 10 PPs created a diversified portfolio that was exposed to all major uncorrelated risk sources and significantly reduced the risks compared to the ASX200 index that is market capitalization weighted. However, even though dimensionality was reduced to 10 principal components, it is still a portfolio of all 156 stocks. We have just changed the weights of each stock in the portfolio to obtain better diversification. In this chapter, we show that to be adequately diversified a portfolio does not require 156 stocks. Jolliffe (1986) pointed out that if a data set can be successfully described by a smaller number of principal components, then it will always be true that it can be replaced by a subset of the variables. When the number of variables in a data set is large, it is often the case that many variables contain repeated information. So it will be the case that a subset of variables contains virtually all the information available in the full data set. We show below with properly choosing stocks, a much smaller portfolio will closely resemble the ASX200 index in terms of the fluctuation in portfolio value.

Many researchers have tried to answer the question - how many stocks make a diversified portfolio? They mainly based their studies on random selection and/or industry selection (Domian et al., 2003, 2007; Statman, 1987). For randomly selected stocks, all stocks are assumed to be equally valuable. If randomly selected within industry groups, it assumed that all stocks in the same industry are equivalent from an investment stand point. Even when one has found the number of stocks that exploit all the diversification benefits, it is nearly impossible to replicate the best combination of stocks that has the promised diversification because stocks do not have same mean return, variance and covariance. Blume and Friend (1978) reported the actual diversification in 70 percent of the investors in their study was much lower than the number of securities in the portfolio suggested. It is very unlikely that investors are randomly selecting stocks. Rather they have preferences for certain types of stocks which make their portfolios under diversified.

Randomly choosing stocks to add to a portfolio, even when the number of stocks required is reached, may not result in the promised diversification if the chosen stocks are more correlated than expected. This means that finding the number of stocks needed to diversify a portfolio is only useful from a theoretical point of view and is impractical because one can not know which stocks should be held. Jacob (1974) pointed out that investors can reduce idiosyncratic risk significantly if they can choose their securities judiciously. We propose a new method that provides the investors the means to select securities judiciously. This method is based on PCA and selects stocks according to their correlation structure. Given a set of stocks, we will not only able to determine the number of stocks that is sufficient to describe the full set of stocks, we also have a way to identify which stocks are needed.

10.1 Stock selection using 156-stocks data set

As with choosing principal components (see Chapter 5), there is more than one method of variable selection. We followed Jolliffe (1986) and used the variable selection method that he claimed to retain the "best" subsets more often than other methods considered. This method is related to Kaiser's rule (see Chapter 5). The selection procedure is described below:

1. Apply PCA to the correlation matrix of a data set.

2. Associate one variable with the highest coefficient in absolute value with each of the last m_1 principal components that have eigenvalue less than a certain level l which we call the deletion criteria, then delete those m_1 variables. For example, one can use Kaiser's rule. Recall that in the case of a correlation matrix, a principal component with eigenvalues smaller than 1 contains less information than one of the original variables.

3. A second PCA is performed on remaining variables. The same procedure was applied that associates one variable with each m_2 principal components that have an eigenvalue less than l, and delete those m_2 variables.

4. The procedure is repeated until no further deletions are considered necessary based on a stopping criteria. One can decide to stop the selection procedure based on the eigenvalue of the last principal component. For example, the stopping criteria can be delete variables until the retaining variables all have eigenvalue not less than 0.7.

For example, a selection procedure with a deletion criteria 1 and a stopping criteria 0.7 was applied to the 156 stocks for whole study period. The further investigation of the results will be discussed later this chapter. First of all, we performed PCA on the correlation matrix of 156 stocks and there were 107 principal components with eigenvalues lower than 1. Then we found the stocks with the highest coefficient in each 107 principal components and there were 84 unique stocks. Note that some stocks have the highest coefficient in more than one principal component. It is not necessary that the number of stocks deleted be equal to the number of principal components with eigenvalue lower than the deletion criteria. We removed the 84 stocks and performed a PCA on the remaining 72 stocks. The eigenvalue of the last principal component was 0.49, which was still lower than 0.7. So the deletion procedure was continued. There were 47 principal components had eigenvalues lower than 1 and there were 40 unique stocks associated with the components. We deleted these 40 stocks and performed a PCA on the remaining 32 stocks. The last principal component had an eigenvalue of 0.64. This was closer but still lower than 0.7. We again deleted the stocks associated with the principal components which had eigenvalues lower than 1 and 15 stocks remained. A PCA was performed on the 15 stocks and the last eigenvalue was 0.77. This was higher than 0.7. We stopped the deletion and there were 15 stocks selected after three cycles of deletion.

The idea behind this method is that low eigenvalue principal components are often associated with near-constant relationships among a subset of variables (see Chapter 6). If such variables are detected and deleted, little information will be lost. With each step of the deletion procedure, the eigenvalues of the new set of variables will converge. In the example discussed above, most of principal components from the selected 15 stocks have eigenvalues close to each other. The second largest eigenvalue was 1.12 and the smallest is 0.77. This means each principal component contains a similar amount of information as one individual stock. The principal components obtained from the selected 15 stocks were approximately the same as the original 15 stocks. This is the case when there is low correlation among the original stocks a PCA is less relevant. Moreover, one can control the deletion speed by adjusting the deletion criteria. Jolliffe (1986) suggested that deleting principal components that have eigenvalue less than 1 is too aggressive and likely to result in a loss of useful information, a more conservative level is 0.7. Thus, we can set the deletion criteria to 0.7 which will slow the deletion process and be less likely to delete stocks that contain useful information.

While we have described the stock selection procedure with a deletion criteria of 1 and a stopping criteria of 0.7 on 156 stocks for the whole study period, we further investigated the performance of selected stocks. We will discuss the three sets of stocks which were selected from the three levels of the deletion cycles. Recall that a deletion criteria of 1 and stopping criteria of 0.7 required three deletion cycles and retained 15 stocks eventually. With two deletion cycles, 32 stocks remained and the eigenvalues of the principal components were not lower than 0.64. With only one deletion cycle, 72 stocks were retained and all the eigenvalues of the principal components of the remaining 72 stocks were higher than 0.49.

In Chapter 7, we found equal investment in uncorrelated PPs dominated other strategies considered. Selecting stocks based on the above selection procedure retains stocks that are almost uncorrelated. Most of the correlations between stocks in all three different sized portfolios lie in the range of 0 to 0.1. This suggests principal components arise from the selected uncorrelated stocks will be almost the same as the selected stocks. Constructing portfolios based on the principal components and individual stocks in this case makes little difference. As a consequence, it is reasonable to assume that 1/N is also the best strategy to apply to the selected stocks. We note that with more stocks retained, the maximum correlation in the portfolio increased. The maximum correlation in the 15-stock portfolio is 0.17. When an extra 17 stocks were added to the portfolio, the maximum correlation rose to 0.27. The maximum correlation in the 72-stock portfolio is even higher, 0.45. However, this is still much lower than the maximum correlation of 0.71 in the full data set of 156 stocks. It is clear that our method of selecting stocks is based on the correlations of stocks in the selection pool. With each step of deletion procedure, stocks with the highest correlations with the others were deleted gradually.

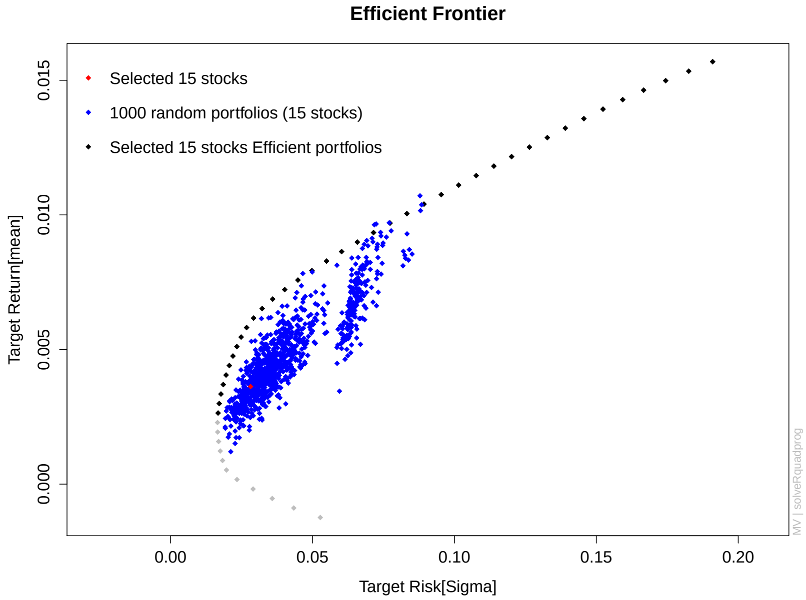
Figure 10.1a presents the efficient frontier constructed based on the selected 15 stocks together with the mean and standard deviation of an equally weighted portfolio of the selected 15 stocks and the means and standard deviations of 1000 equally weighted randomly selected portfolios of 15 stocks. For the random portfolios, the stocks were selected from the 156 stocks in our data set without replacement. It is clear that except for four portfolios, all random portfolios of 15 stocks lie in the achievable region, which is inside the efficient frontier in Figure 10.1a. This means there will be at least one portfolio constructible from the selected 15 stocks that has the mean and volatility corresponding to each of the 996 random portfolios.

We also constructed an efficient frontier based on one of the random 15 stock portfolios for comparison purposes and this is presented in Figure 10.1b. On the other hand, lots of the random portfolios lie outside the efficient frontier in Figure 10.1b. Within the 1000 random portfolios of 15 stocks, one can replicate 996 portfolios with the 15 stocks selected by PCA selection procedure and the 15 randomly selected stocks do not do as well as the ones selected by the PCA. Thus, the 15 stocks selected from our method explain the original 156 stocks well.

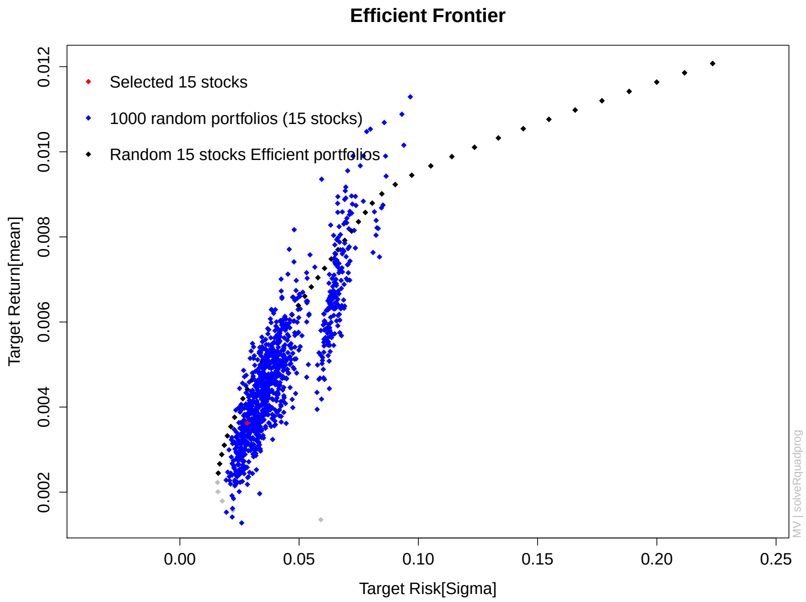
We find that the selected 32 stocks describe the original data set even better than the selected 15 stocks. In Figure 10.2a, all the random portfolios lie in the achievable region. Compared to the 32 stocks selected by our method, the 32 stocks randomly picked from the full data set can not achieve all the mean and volatility corresponding to the 1000 random portfolios (see Figure 10.2b). The selected 72 stocks from our method is not superior to the 72 stocks randomly selected in terms of describing the full data set (see Figure 10.3). Moreover, by comparing the mean and standard deviation of portfolios of selected stocks to the random portfolios of the same size, we find that the 15-stock portfolio and 72-stock portfolio lie in the middle of the random portfolio cluster. In contrast, the selected 32-stock portfolio lies in the left edge of the random portfolio cluster. Intuitively, the portfolio of selected 32 stocks tends to have lower risk for the given level of return or higher return for the given level of risk compared to the random portfolios.

We have found that three different numbers of selected stocks all explain well the original 156 stocks. When comparing the risk and return of portfolios of selected stocks to the random portfolios, 32-stock portfolio stands out. We further compared the risk and return of three selected portfolios to try to find the point where the benefits of diversification are virtually exhausted. Figure 10.4 presents the mean and standard deviation of the weekly returns for three portfolios together with the efficient frontier of the 156 stocks. The portfolio of 32 stocks has slightly reduced the risk and had higher returns compared to the portfolio of 15 stocks. When the portfolio size was increased to 72, the return increased but the risk was higher compared to the portfolio of 32 stocks. All three portfolios lie close to the global minimum variance point, which is the lowest possible variance a portfolio of 156 stocks. We conclude that 15 stocks are not enough to diversify a portfolio and 32 stocks are where all the diversification benefits are exploited when using the whole study period for the investigation. Further spreading the portfolio's investment to include 72 stocks is superfluous

Fig. 10.1 The first plot presents the efficient frontier constructed based on the 15 stocks selected by PCA, the mean and standard deviation of equally weighted portfolio of the selected 15 stocks and mean and standard deviation of 1000 equally weighted random portfolio of 15 stocks selected from the 156 stocks in our data set. The second plot is similar to the first plot except the efficient frontier is constructed from one of the random 15 stock portfolios. All the returns are on a weekly basis.

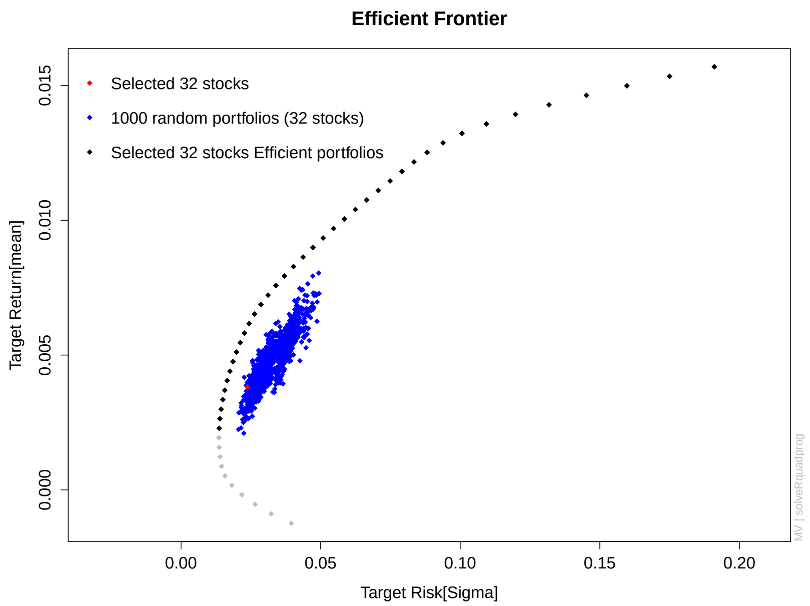


(a) Efficient frontier of 15 stocks selected by PCA.

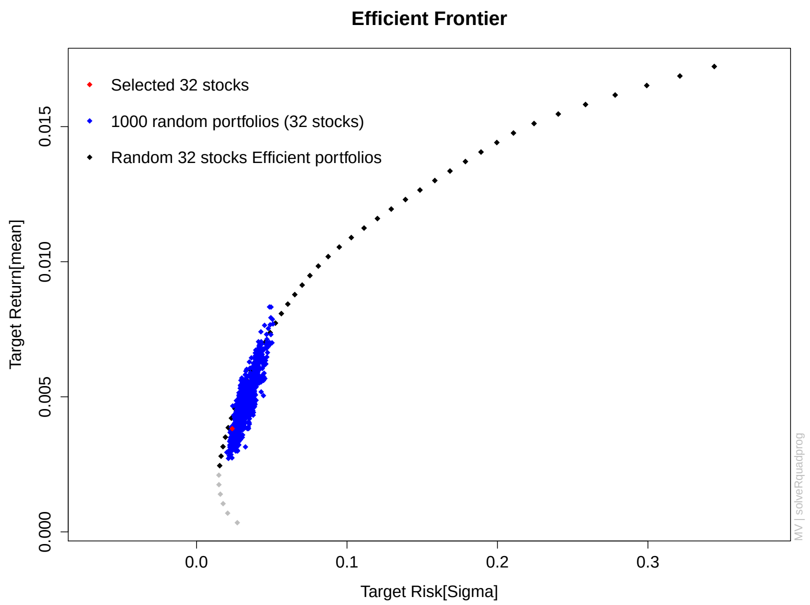


(b) Efficient frontier of 15 randomly selected stocks.

Fig. 10.2 The first plot presents the efficient frontier constructed based on the 32 stocks selected by PCA, the mean and standard deviation of equally weighted portfolio of selected 32 stocks and mean and standard deviation of 1000 equally weighted random portfolio of 32 stocks selected from the 156 stocks in our data set. The second plot is similar to the first plot except the efficient frontier is constructed from one of the random 32 stock portfolios. All the returns are on a weekly basis.

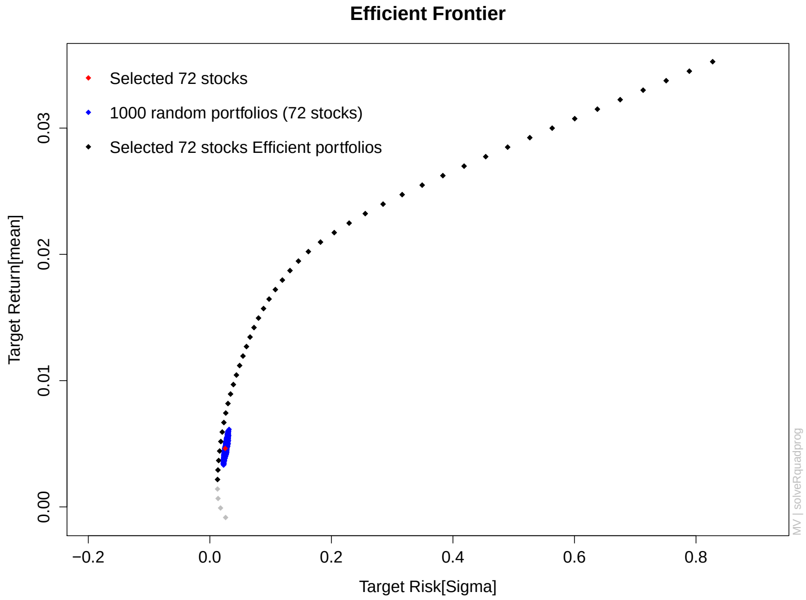


(a) Efficient frontier of 32 stocks selected by PCA.

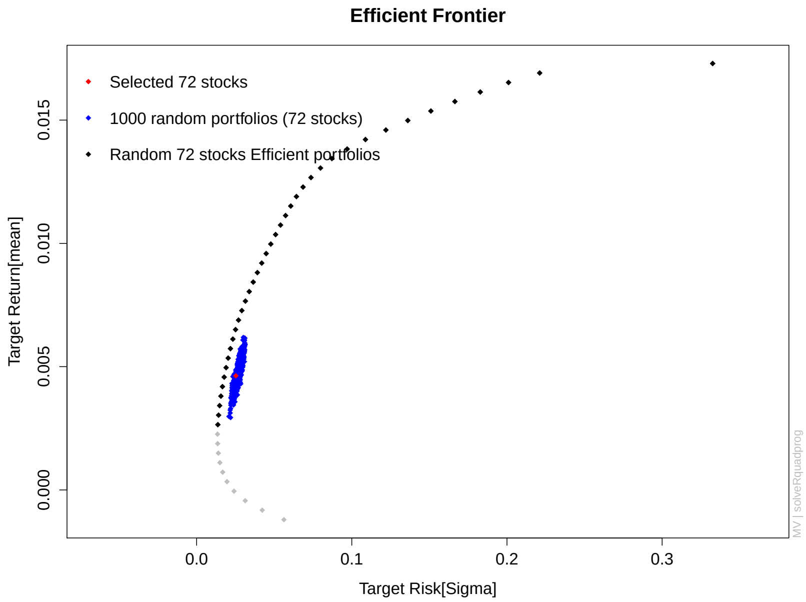


(b) Efficient frontier of 32 randomly selected stocks.

Fig. 10.3 The first plot presents the efficient frontier constructed based on the 72 stocks selected by PCA, the mean and standard deviation of equally weighted portfolio of selected 72 stocks and mean and standard deviation of 1000 equally weighted random portfolio of 72 stocks selected from the 156 stocks in our data set. The second plot is similar to the first plot except the efficient frontier is constructed from one of the random 72 stock portfolios. All the returns are on a weekly basis.

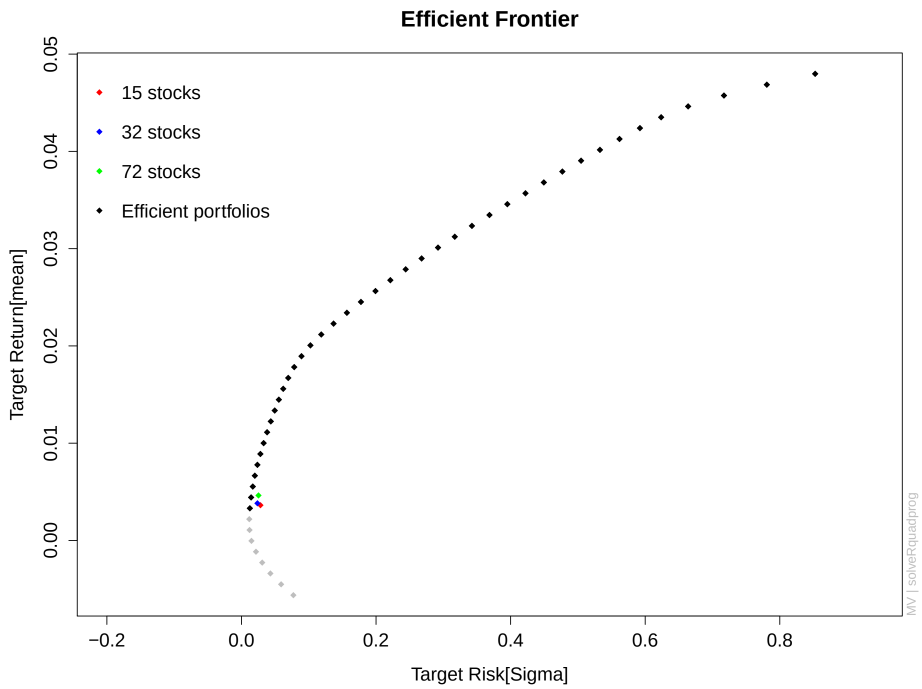


(a) Efficient frontier of 72 stocks selected by PCA.



(b) Efficient frontier of 72 randomly selected stocks.

Fig. 10.4 The efficient frontier of 156 stocks and mean and standard deviation of three different sized equally weighted portfolios. PCA was performed on 156 stocks, whole study period. 72 stocks was a portfolio which contains stocks that were retained based on only one deletion (Equivalent to deletion criteria 1 and stop criteria 0.49). 32 stocks was portfolio which contains stocks that have had two deletions (Equivalent to deletion criteria 1 and stop criteria 0.64). 15 stocks was portfolio which contains stocks that have had three deletions. (Equivalent to deletion criteria 1 and stop criteria 0.77). All three portfolios are equally weighted and all the returns and standard deviations are weekly.



diversification and should be avoided.

Table 10.1 and Table 10.2 lists the selected 15 stocks and selected 32 stocks together with their industry information respectively. The stocks selected were spread across almost all industries. Recall that there are a total of 10 industries represented in the ASX200 index based on ICB industry classification. The 15-stock portfolio included nine out of the 10 industries while the 32 stock portfolio contained all industries. We found that when the number of stocks doubled from 15 to 32, the stocks added were also spread over all industries. Moreover, we noticed that in both the 15 and the 32 stocks selected stocks, major companies such as BHO and RIO in Basic Materials, the four big banks in the Financial industry, WPL and STO in Oil & Gas industry, were not selected. The explanation may be

Table 10.1 The 15 stocks that were selected from the 156 stocks used for the whole study period, based on a deletion criteria of an eigenvalue 1 and stop criteria of 0.77.

Stocks Code	**Industry**
MAH	Basic Materials
TRY	Basic Materials
AVG	Consumer Goods
ELD	Consumer Goods
MTS	Consumer Services
VRL	Consumer Services
DJW	Financials
IBC	Financials
IOF	Financials
RHC	Health Care
AJL	Industrials
HIL	Industrials
AUT	Oil & Gas
SMX	Technology
HTA	Telecommunications

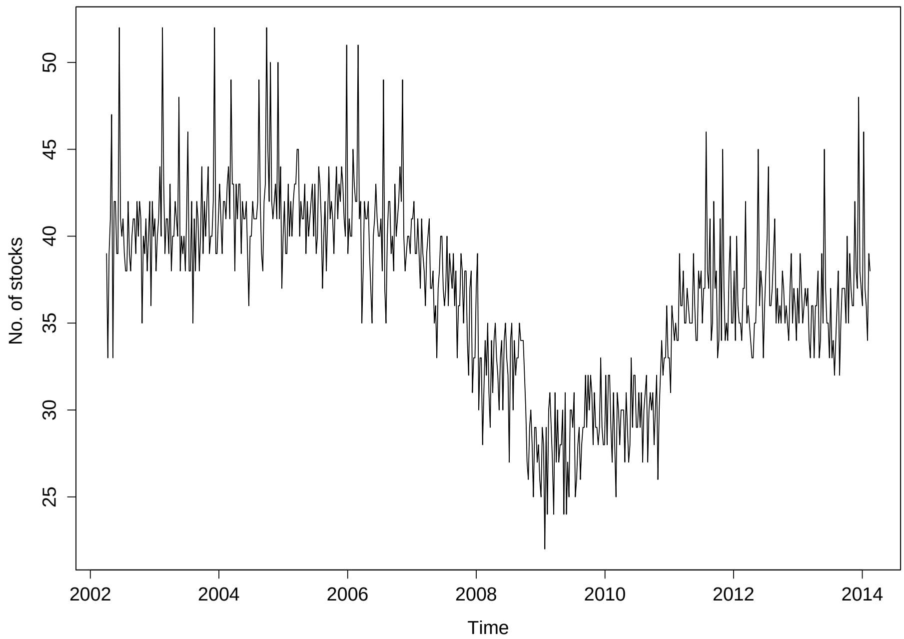
twofold. These stocks were identified in the last few principal components that were near linearly correlated (see Chapter 6). They all have high coefficients and were deleted based on our stock selection procedure. The other explanation is the major companies are exposed to multiple risk sources and move with the market. The stock selection procedure tends to select stocks that represent the uncorrelated risk sources in the market.

The correlations between stocks changed over time and this affects the number of stocks selected. We have investigated the stocks market connectedness in Chapter 8 and found that when the market is more connected, the variance is concentrated in a smaller number of principal components. Based on this, we suspected that, during the periods of a more connected market, there should be less risk sources. This means one should expect a smaller number of stocks being selected to describe the market. In order to show this we applied a stock selection procedure of 0.7 for an eigenvalue of the deletion criteria and 0.5 for the stopping criteria to our 156 stocks sample on a rolling window basis. The window size is two years (504 trading days). In Figure 10.5, the number of stocks selected decreased starting 2007 and reached the lowest in late 2009. The market had already become concen-

Table 10.2 The 32 stocks that were selected from the 156 stocks used for the whole study period, based on a deletion criteria of an eigenvalue 1 and stop criteria of 0.64. Stocks that are retained in the 15 stock portfolio are highlighted.

Stocks Code	**Industry**
AGG	Basic Materials
MAH	Basic Materials
MDL	Basic Materials
RSG	Basic Materials
TRY	Basic Materials
AVG	Consumer Goods
ELD	Consumer Goods
GUD	Consumer Goods
MTS	Consumer Services
PRT	Consumer Services
SWM	Consumer Services
VRL	Consumer Services
AOG	Financials
BOQ	Financials
CPA	Financials
DJW	Financials
IBC	Financials
IOF	Financials
REA	Financials
RHC	Health Care
RMD	Health Care
AJL	Industrials
HIL	Industrials
MRM	Industrials
PMP	Industrials
SKE	Industrials
SLX	Industrials
AUT	Oil & Gas
MLB	Technology
SMX	Technology
HTA	Telecommunications
AGK	Utilities

Fig. 10.5 The number of stocks selected by PCA over time. A stocks selecting procedure of 0.7 for deletion criteria and 0.5 for stop criteria was used. The selecting procedure was applied on a rolling window basis with window size of two years (504 trading days).



trated and offered fewer diversification opportunities before the 2008 financial crisis started. When the market became less tightly coupled, the number of stocks increased. This trajectory of number of selected stocks moved in an opposite way to the level of systemic risk (see Figure 8.2). This illustrates that the number of stocks needed to diversify a portfolio is not constant through time. With the number of major stock market risk sources changing, a portfolio can be considered diversified consistently only if it is adaptive to the change. In other words, the number of stocks included to diversify major risk sources should change based on the number of risk sources in the market. Thus, a portfolio that holds the same number of stocks or same constituents can only be the best combination to create a diversified portfolio at a single point of time. Holding more stocks than necessary when the number of major risk sources decreases is redundant. On the other hand, holding fewer stocks than required when the number of major risk sources increases means that the portfolio is under-diversified.

10.2 Stock selection using full data set

Our final test of stock selection was to examine the performance of the selected stocks compared to the ASX200 index value. We found that the fluctuation of the ASX200 index value can be replicated with a much smaller portfolio. We use the ASX200 index that includes 200 stocks as our benchmark portfolio and compared portfolios of smaller size to it. The index is used as an example of an attainable and fairly diversified portfolio of 200 stocks and we do not assume that we cannot obtain better diversified portfolio of 200 stocks. The index funds provide opportunities for investors to acquire a diversified portfolio at low cost. There are incentives for investors to hold individual stocks. One is reduction in brokerage commissions and management fees. Another is buying individual stocks also gives investors better control in timing of realisation of capital gains and losses.

All the tests we have done so far were based on the 156 stock data set. We were concerned that stock selection was sensitive to the selection pool. With different stocks available to be chosen, the selection procedure may result in a very different set of stocks. In order to better compare with ASX200 index, using more complete constituents was considered more appropriate. So we divided the whole study period into seven subsets, each with a sub study period of two years (around 504 trading days)[1], except for the last sub period which is less than two years and only had 472 trading days. We extracted the stocks that had complete returns information in the relevant periods. Table 10.3 summarizes the number of stocks in selection pool in each two year sub period.

We performed in sample and out of sample tests of stocks selected for each two years sub-period. We compared the portfolio value of stocks selected in the first year to the portfolio value of stocks selected in the second year. This is the comparison of the portfolio which would have been held and the portfolio which should have been held. We note that all the investigation have been done were based on two years length data of 156 stocks as the KMO statistic suggested (see Chapter 4). We also performed the KMO test on each two years sub-period data, and the shortest length of data to efficiently apply PCA was one year. Table 10.4 presents the KMO statistic of each year. From 2006, the KMO statistics were all

[1] All the sub study periods are two years exactly but the actual number of trading days may vary.

Table 10.3 The number of stocks in the selection pool in each two year sub-period.

Study Period	No. of stocks
2000-2002	171
2002-2004	172
2004-2006	175
2006-2008	187
2008-2010	195
2010-2012	190
2012-2014	194

above 0.7. There was only one year, 2004 to 2005, the KMO statistic went below 0.5, the lowest acceptable value. The portfolio construction was carried out in following steps:

1. Within each two year sub period, a stock selection procedure was applied to the first year and second year separately. This created two sets of selected stocks, one is based on the first year data and another is based on second year data.

2. For each set of selected stocks, construct a portfolio that has equal investment in those stocks. We call the portfolio of stocks selected based on first year data the “first period model portfolio" and the portfolio of stocks selected based on second year data the “second period model portfolio".

3. For first period model portfolio, we assumed a $1 million investment at the beginning of second year. This means the portfolio was constructed based on the price at the first day of second year. The portfolio value will converge to $1 million at the first day of second year and diverge since then.

4. We also assume $1 million portfolio value at the first day of second year for the ASX200 index and computed its value for whole two years.

5. For second period model portfolio, we only computed the second year portfolio value and also assumed $1 million investment at the beginning of the second year.

The test results are presented in Figure 10.6. The gray vertical line indicates the first day of the second year, where the portfolio value equal to $1 million. The right hand side of the gray line shows the portfolio value of stocks held, stocks that should have been held and the ASX200 index value. Obviously, portfolios of selected stocks, regardless of the period from which they were selected, moved closely relative to the index. Especially in periods of 2004 to 2006 and 2012 to 2014, the trajectory of first period model portfolio and second period model portfolio approximately matched the index. Moreover, the selected stock portfolios not only tracked the index well, they also consistently outperformed the index in most of the second periods. The outperformance of the selected stock portfolios might be the result of the dividends. Recall that the ASX200 index did not adjust for dividends paid but we assumed all dividends were reinvested for all constituents. However, as the number of stocks in the selected stock portfolios was small and the study period in this test was short, we believed the dividends would not influence the result much. We found that the first period model in general more closely resembled the ASX200 index in out-of-sample testing except for the period of 2008 to 2010. While the second period model portfolio closely evolved with the index, the first period model portfolio was far more volatile in 2009. One explanation for this is that during the 2008 financial crisis, the market condition changed significantly and the affects caused by the crisis lasted a long time. The first period in 2008 to 2010 is completely different from its second periods. Consequently, stocks selected from the first period market conditions was not adapted to the second period market conditions.

While the trajectory of selected stocks portfolios shows that the ASX200 index can be described by smaller portfolios, we next investigated the number of stocks that were selected in these portfolios. Table 10.5 presents the number of stocks selected in the first period and second period of each two years sub period. For the first three two year sub-periods, the number of stocks selected was all above 20 and the maximum is 25. The difference between the first period and second period was not larger than two stocks. This minor difference also reflected in the selected portfolio values in Figure 10.6. The red line (second period model portfolio) and the blue line (first period model portfolio) are almost matched. The difference in the number of stocks selected between the first period and second period increased for the subsequent study periods. The trajectories of two period portfolios are less similar. (See Figure 10.6). The number of stocks selected declined to below 20. This is the reflection of

Table 10.4 Measure of Sampling Adequacy for each two year sub-period.

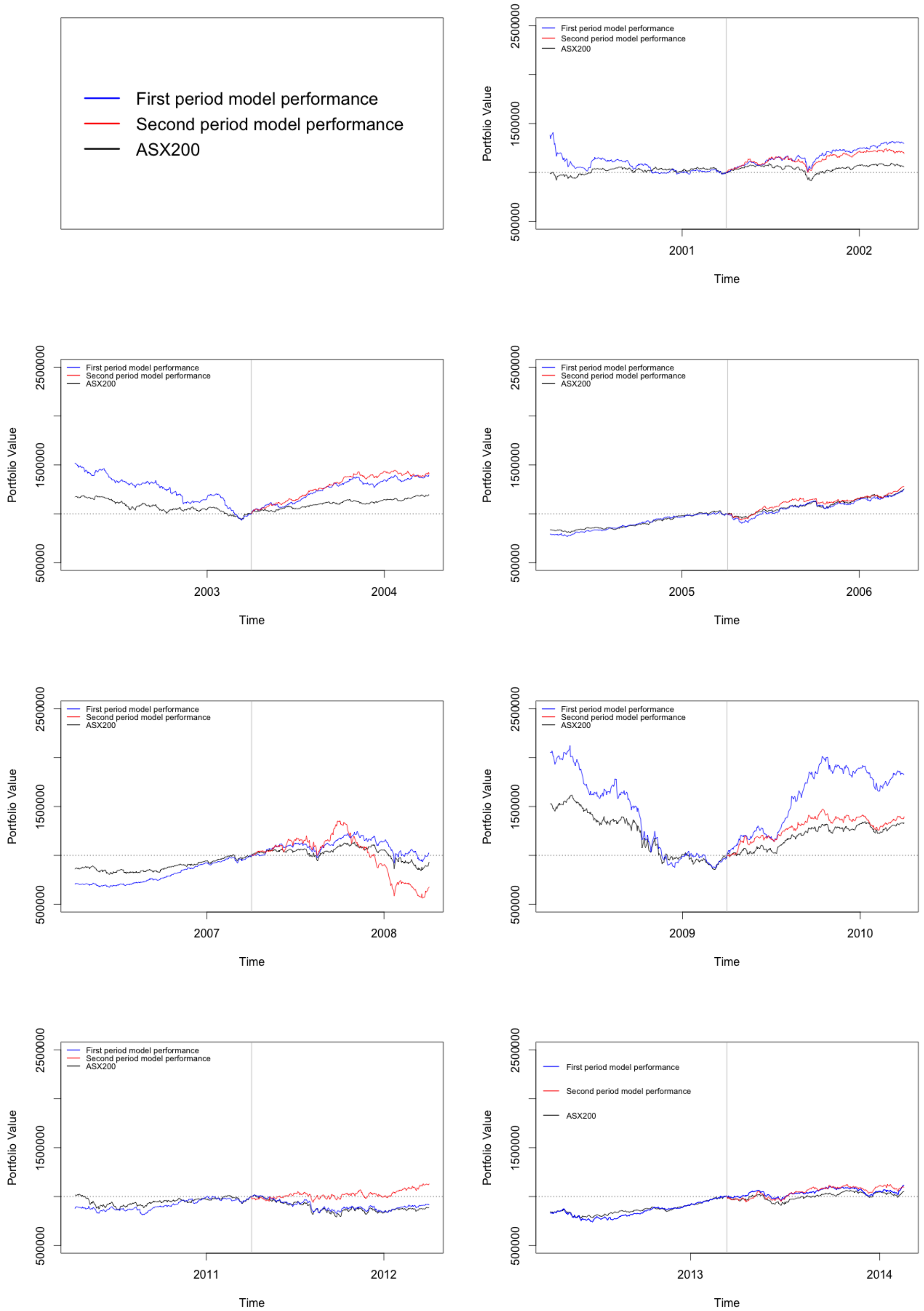
	1st period	**2nd period**
2000-2002	0.51	0.54
2002-2004	0.58	0.50
2004-2006	0.45	0.65
2006-2008	0.75	0.86
2008-2010	0.77	0.73
2010-2012	0.81	0.90
2012-2014	0.61	0.71

Table 10.5 The number of stocks selected for each year. A deletion criteria of an eigenvalue 0.7 and stop criteria of 0.5 was used.

Study Period	**1st period**	**2nd period**
2000-2002	21	20
2002-2004	23	25
2004-2006	22	21
2006-2008	18	14
2008-2010	13	17
2010-2012	19	12
2012-2014	21	17

more connected market. For the first year of the last study period, the number of stocks rises to 21 but in the second year, this number decreased. Our results indicate that to adequately diversify a portfolio, one does not have to include all 200 stocks. A portfolio with about 20 stocks well describes the 200 stock portfolio. For investors who want to buy individual stocks and replicate the fluctuation of the index, our method of stock selection provides a way to make this possible.

Fig. 10.6 In sample and out of sample test of portfolios of selected stocks against the ASX200 index value. Stocks selection was based on deletion criteria of an eigenvalue of 0.7 and stop criteria of 0.5.



Chapter 11

Conclusions and Further Research

11.1 Conclusions

In the last 14 years in the Australian stock market, there were approximately 10 major risk sources inherent within the stocks (see Chapter 5). In Chapter 7, we constructed principal portfolios based on the 10 retained components. A principal portfolio constructed based on the first component, which is a market component contains the most systematic risk compared to all other components, was essentially a 1/N portfolio on all stocks considered. The principal portfolio constructed based on the second component had the highest correlation with the ASX200 index among the first 10 principal components. When portfolio allocation was determined based on the principal portfolios that represent the 10 major risk sources, the risk decreased substantially compared to the same allocation strategies based on the underlying stocks. Even more so, one could have avoided the significant drop during the 2008 financial crisis. Among the 1/N allocation strategy, ERC, and capitalization weighted allocation strategy, the 1/N portfolio dominated the others regardless of whether it was based on principal portfolios or underlying stocks.

In Chapter 9 we also investigated the time evolution of the relative importance of each stock in the first 10 principal components respectively. The first four principal components illustrated certain structures related to the industries. According to the test of principal component two, changing the portfolio position based on the coefficient change would not have resulted in better performance but rather provided a more close relationship between principal portfolio two and the ASX200 index.

In Chapter 10 we showed that a portfolio of at most 25 stocks closely resembled the ASX200 index over time. Our results also revealed that it is not just any combination of stocks which can used to represent the whole data set. It must be a group of carefully selected stocks using the PCA selection procedure. This is an important finding for two reasons. The traditional researchers have only determined the number of stocks that make a diversified portfolio and implicitly assumed all combination of stocks are the same, which clearly they are not. The other reason is we provide means for investors to select stocks that will result in the promised diversification that was not accomplished by the other papers.

The stock market was more concentrated during crisis period (see Chapter 8). A portfolio that holds the same position over time will be less diversified when the market is tightly coupled. The variance explained by principal component one was an effective measure of the level of systemic risk and served as a leading indicator of financial crisis. We also found the KMO measure of sampling adequacy was highly correlated with the variance explained by principal component one and hence could be used in the same role.

It is not only the first ten principal components which have important applications in portfolio management. The last few principal components successfully identify stocks which have near linear correlations and which have useful implications in portfolio construction (see Chapter 6). This was also the idea behind the stock selection method discussed above.

11.2 Further research

1. In our research, we did not look into the financial interpretation of the principal components. The bi-plots and the heat maps of the coefficients of stocks in the principal components are both ways to find possible interpretations of the principal components. However, the bi-plots and heat maps did not always show any information which could be used to interpret the components. Normally more than one method

should be used to achieve the goal. Lohre et al. (2012) regressed each principal components against a set of well-known factors, such as a market return, size factor, liquidity factor. They reported this method was quite effective shaping the understanding of the principal components. This method should be applied to Australian market PCA.

2. When we investigated the market connectedness over time, we measured the level of systemic risk without further studying the change of systemic risk. Kritzman et al. (2011) and Zheng et al. (2012) researched the change of systemic risk and reported that it is a possible way to predict the market turbulence before it happened.

3. We pointed out that all allocation strategies that can be applied to the individual stocks could also be applied to the principal portfolios. We have only tested 1/N and ERC in this research. We believed it is valuable to compare all allocation strategies available to portfolio selection based on stocks and principal portfolios.

References

Adrian, T. (2007). Measuring risk in the hedge fund sector. *Current Issues in Economics and Finance*, 13(3).

Aguirre, S. M. and Saidi, R. (2004). Japanese banks liability management before and during the banking crises and macroeconomic fundamentals. *Journal of Asian Economics*, 15:373–397.

Allen, G. C. (2010). The Risk Parity Approach to Asset Allocation. [online]http://www.top1000funds.com/attachments/TheRiskParityApproachtoAssetAllocation2010.pdf.

Artzner, P. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997). Thinking coherently. *Risk*, 10(11):68–71.

Bahansali, V., Gingrich, R. M., and Longstaff, F. A. (2008). Systemic credit risk: What is the market telling us? *Financial Analysts Journal*, 64(4):16–24.

Behr, P., Gütter, A., and Miebs, F. (2008). Is minimum-variance investing really worth the while? An analysis with the robust performance inference. *Working paper*. [online] http://www.cfr-cologne.de/download/kolloquium/2009/behr_et_al.pdf.

Bernstein, W. (2001). *The Intelligent Asset Allocator*. McGraw-Hill.

Billioand, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104:535–559.

Blume, M. E. and Friend, I. (1978). *The Changing Role of the Individual Investor: A Twentieth Century Fund Report* . John Wiley & Sons, New York.

Boudt, K., Peterson, B., and Croux, C. (2008). Estimation and decomposition of downside risk for portfolios with non-normal returns. *Journal of Risk*, 11(2):79–103.

Brana, S. and Lahet, D. (2009). Capital requirement and financial crisis: The case of Japan and the 1997 Asian crisis. *Japan and the World Economy*, 21:97–104.

Brandt, O. D. and Hartmann, P. (2000). Systemic risk: a survey. *Working paper*. [online] https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp035.pdf.

Butler, K. C. and Joaquin, D. S. (2001). Are the gains from international portfolio diversification exaggerated? The influence of downside risk in bear market. *Working paper*. [online] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=221992.

Campbell, J., Martin, L., Burton, M., and Xu, Y. X. (2001). Have Individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance*, 56(1):1–43.

Campbell, R., Koedijk, K., and Kofman, P. (2002). Increased correlation in bear markets. *Review of Financial Studies*, 58(1):87–94.

Cappiello, L., Engle, R., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4(4):385–412.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research*, 1:245–276.

Chopra, V. K. and William, T. Z. (1993). The effect of errors in mean, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11.

Choueifaty, Y. and Coignard, Y. (2008). Towards maximum diversification. *The Journal of Portfolio Management*, 35(1):40–51.

Clarke, R., de Dilva, H., and Thorley, S. (2006). Minimum variance portfolio in the U.S. market. *The Journal of Portfolio Management*, 33(1):10–24.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Oxford Journals*, 22(5):1915–1953.

Domian, D. L., Louton, D. L., and Racine, M. D. (2003). Portfolio diversification for long holding periods: How many stocks do investors need? *Studies in Economics and Finance*, 21:40–64.

Domian, D. L., Louton, D. L., and Racine, M. D. (2007). Diversification in portfolios of individual stocks: 100 stocks are not enough. *The Financial Review*, 42:557–570.

Driesson, J., Melenberg, B., and Nijman, T. (2003). Common factors in international bond returns. *Journal of International Money and Finance*, 22:629–656.

Evans, J. L. and Archer, S. H. (1968). Diversification and the reduction of dispersion: An empirical analysis. *Journal of Finance*, 23:761–767.

Farmer, S. A. (1971). An investigation into the results of principal component analysis of data derived from random numbers . *Statistician*, 20:63–72.

Feeney, G. J. and Hester, D. D. (1967). *Risk Aversion and Portfolio Choice*. Wiley, New York.

Fenn, D. J., Porter, M. A., Williams, S., MacDonald, M., Johnson, N. F., and Jones, N. S. (2011). Temporal evolution of financial-market correlations. *Physical Review E*, 84:026109.

Ferreira, M. and Gama, P. (2004). Correlation dynamics of global industry portfolios. *ISCTE Business School*. [online] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=517742.

Frahm, G. and Wiechers, C. (2011). On the diversification of portfolios of risky assets. *Discussion paper in statistic and econometrics*. [online] https://ideas.repec.org/p/zbw/ucdpse/211.html.

Francis, J. C. (1986). *Investment: Analysis and Management*. McGraw-Hill, New York, Fourth edition.

Gup, B. E. (1983). *The Basics of Investing*. John Wiley & Sons, New York, Second edition.

Haugen, R. A. and Baker, N. (1991). The efficient market inefficiency of the capitalisation-weighted stocks portfolios. *Journal of Portfolio Management*, 17:35–40.

Hong, Y., Tu, J., and Zhou, G. (2007). Asymmetries in stock returns: statistical tests and economic evolution. *Review of Financial Studies*, 20(5):1547–1581.

Jacob, N. L. (1974). A Limited-Diversification Portfolio Selection Model for The Small Investor. *Journal of Finance*, 29:837–857.

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer, New York.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, 58(3):259–278.

Jorion, P. (2007). *Value at Risk: the new benchmark for managing financial risk*. McGraw Hill, New York, Third edition.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4):401–415.

Kaiser, H. F. and Rice, J. (1974). Little jiffy. *Educational and Psychological Measurement*, 34(1):111–117.

Kenett, D. Y., Raddant, M., Lux, T., and Ben-Jocob, E. (2012). Evolvement of uniformity and volatility in the stressed global financial village. *PLoS ONE*, 7:e31144.

Kim, D.-H. and Jeong, H. (2005). Systematic analysis of group identification in stocks markets. *Physical Review E*, 72:046133.

Kind, C. (2013). Risk-based allocation of principal portfolios. *Working paper*. [online] http://ssrn.com/abstract=2240842.

Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2011). Principal Components as a measure of systemic risk. *Journal of Portfolio Management*, 37:112–126.

Lee, W. (2011). Risk-based asset allocation:a new answer to an old question? *The Journal of Portfolio Management*, 37(4):11–28.

Lehar, A. (2005). Measuring systemic risk: A risk management approach. *Journal of Banking and Finace*, 29:2577–2603.

Lohre, H., Neugebauer, U., and Zimmer, C. (2012). Diversified risk parity strategies for equity portfolio selection. *Journal of Investing*, 21(3):111–128.

Lohre, H., Opfer, H., and Orszag, G. (2014). Diversified risk parity. *Journal of Risk*, 16(5):53–79.

Lowenfeld, H. (1909). *Investment, an Exact Science*. Financial Review of Reviews.

Maclean, B. and Nocera, J. (2010). *All the Devils Are Here: The Hidden History of the Financial Crisis*. Penguin Group, New York.

Maillard, S., Roncalli, T., and Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, 36:60–70.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, page 392–417.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7:77–91.

Meucci, A. (2009). Managing Diversification. *Risk*, 22(5):74–79.

Michaud, R. O. (1989). The Markowitz optimisation Enigma: is 'optimised' optimal? *Financ. Anal. J.*, 45(1):31–42.

Newbould, G. D. and Poon, P. S. (1993). The minimum number of stocks needed for diversification. *Financial Practice and Education*, 3:85–87.

Newbould, G. D. and Poon, P. S. (1996). Portfolio risk, portfolio performance, and the individual investor. *Journal of Investing*, 5:72–78.

Partovi, M. H. and Caputo, M. (2004). Principal portfolios: recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10.

Pérignon, C., Smith, D. R., and Villa, C. (2007). Why common factors in international bond returns are not so common. *Journal of International Money and Finance*, 26:284–304.

Pfaff, B. (2013). *Financial Risk Modelling and Portfolio Optimization with R*. Wiley, New York.

Qian, E. (2006). On the financial interpretation of risk contribution: risk budgets do add up. *Journal of Investment Management*, 4:1–11.

Qian, E. (2011). Risk parity and diversification. *Journal of Investing*, 20:119–127.

Reilly, F. K. (1985). *Investment Analysis and Portfolio Management*. Dryden Press, San Francisco, Second edition.

RiskMetrics Group (1994). Riskmetrics technical document. Technical report, JP Morgan.

Rudin, A. and Morgan, J. S. (2006). A portfolio diversification. *The Journal of Portfolio Management*, 32:81–89.

Sharpe, W. F. (1963). Mutual Fund Performance. *Journal of Business*, 39:119–138.

Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, 22:353–363.

Statman, M. (2004). The diversification puzzle. *Financial Analysts Journal*, 60:44–53.

Stevenson, R. A. and Jennings, E. H. (1984). *Fundamentals of Investments*. West Publ. Co., San Francisco, Third edition.

Wang, D., Podobnik, B., and Stanley, D. H. H. (2011). Quantifying and modelling long-range cross correlations in multiple time series with applications to world stock indices. *Phy. Rev. E*, 83:046121.

Zangari, P. (1996). A VaR methodology for portfolios that include options. *RiskMetrics Monitor*, 39:4–12.

Zheng, Z., Podobnik, B., Feng, L., and Li, B. (2012). Changes in cross-correlations as an indicator for systemic risk. *Scientific Reports*, 2:888.

Appendix A

Data: detailed description

The ASX200 index is reconstructed from time to time. We had to identify all the stocks which had ever been in the ASX200 for the whole study period, from 03/04/2000 to 17/02/2014. SIRCA provided historical constituents at any time point, and historical deletions and additions to the index. However, there was no list of constituents for whole study period. In order to get this list, we compared the information on deletions and additions to the current constituent list (constituent list on 17/02/2014) and this gave a list of all the stocks, which had ever been in the index for whole study period. This list contained 596 stocks initially. After we obtained the list of all constituent stocks, we summited the list to SIRCA and requested the price and dividend information. When we got the prices and dividends for all 596 stocks, we soon noticed there were some problems associated with the data:

1. There were 188 non-AUD dividends. To deal with this problem, we obtained exchange rate data from Australian Reserve Bank Website[1] in date order. Then we manually converted the non-AUD dividends to AUD. Three dividends were paid for which no exchange rate was available. We used the nearest day's exchange rate. The three dividends are:

    - BHP on 28/09/2008
    - FOX on 16/04/2014

[1]http://www.rba.gov.au/statistics/tables/index.html#exchange-rates

- SGT on 01/08/2005

Two dividends were on percentage yield instead of dollars and one of them was 100%. That was unrealistic so we decided to delete it. The other dividend in percentage did not have a stock price (not available in wither the ASX website or the SIRCA database) on the dividend paid day, we assumed it was recorded mistakenly and deleted it as well. These two dividends were:

- GPT on 21/08/2009
- MDL on 03/12/2010

2. After we adjusted the dividend data, we merged the price and dividend into one large file, then dealt with the name changes, mergers and acquisitions of all stocks. The changes of name, a merger or acquisition were recorded under symbology change database in SIRCA. We compared all the symbology changes that happened within our study period with the original 596 unique stocks to find any stocks that were related to the 596 stocks. We found an extra 63 stocks that are either the same stock with one of the 596 stocks with different symbology, or stocks that had been merged or acquired by any of the 596 stocks. We then summited them to SIRCA and requested price and dividend data as well. There were 10 of the 63 that did not have price and dividend data in SIRCA. We checked those 10 stocks and it turned out that they had been updated for symbology change in SIRCA database. We merged stocks that either were the same stocks with different symbol or stocks that had been merged or acquired and used the latest stock's name.

3. There were 531 unique stocks (not including the index itself) after adjusting for any name changes or mergers and acquisitions. But we noticed that there were some days for some stocks which did not have price information. These days were neither weekend nor public holidays. After checking online, those stocks with no trades were because of trading halts[2]. We used zero returns for days with trading halts.

[2] Trading Halt is a temporary suspension in the trading of a particular security on one or more exchanges, usually in anticipation of a news announcement or to correct an order imbalance.

4. The prices obtained from SIRCA had not been adjusted for stock splits or consolidations. We adjusted them manually. All stocks split and consolidation information was obtained from ASX or Yahoo! Finance[3].

5. There were seven stocks which did not have price and dividend information at all and we deleted them. These seven stocks were:

   - DVT
   - GAS
   - PDP
   - WFA
   - WFT
   - WSF
   - UTB

This left us with a final data set which included 524 unique stocks for our research.

[3] https://nz.finance.yahoo.com/

Appendix B

Standard deviations of the 156 stocks with complete data

Table B.1 A list of standard deviations of the 156 stocks with complete data, based on daily returns and sorted by standard deviations.

	Stocks	**Std. Dev.**
1	MGL	39.79%
2	ALZ	7.83%
3	RSG	7.68%
4	NRT	7.11%
5	SAR	6.30%
6	SDL	6.15%
7	FMG	5.93%
8	EWC	5.93%
9	CDU	5.92%
10	PDN	5.63%
11	CYO	5.43%
12	AUT	5.42%
13	LYC	5.37%

Continued on next page

Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
14	MPO	5.19%
15	IAU	5.18%
16	OEC	5.14%
17	CVN	5.13%
18	AZZ	5.10%
19	PNA	5.07%
20	ALU	4.83%
21	DLS	4.72%
22	SBM	4.58%
23	GBG	4.53%
24	MDL	4.52%
25	MAQ	4.45%
26	HZN	4.36%
27	PMP	4.28%
28	SLX	4.28%
29	MCR	4.27%
30	SXY	4.21%
31	REA	4.20%
32	HTA	4.09%
33	ASL	4.08%
34	UML	4.06%
35	IMD	3.99%
36	MAH	3.90%
37	AIX	3.74%
38	CAA	3.68%
39	AGG	3.65%

Continued on next page

Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
40	AQP	3.63%
41	TWR	3.60%
42	ELD	3.55%
43	OZL	3.51%
44	KCN	3.46%
45	MRM	3.42%
46	SDG	3.40%
47	MLB	3.37%
48	IIN	3.36%
49	IDT	3.33%
50	TRY	3.30%
51	AJL	3.29%
52	AVG	3.26%
53	SMX	3.20%
54	AOG	3.13%
55	ERA	3.12%
56	ROC	3.09%
57	ALL	3.02%
58	SFH	3.00%
59	SKE	2.92%
60	BPT	2.89%
61	TNE	2.88%
62	BRG	2.82%
63	ASB	2.79%
64	NWS	2.75%
65	DOW	2.73%

Continued on next page

Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
66	CGF	2.71%
67	ENE	2.71%
68	AWC	2.69%
69	AWE	2.69%
70	TAP	2.66%
71	API	2.59%
72	NCM	2.59%
73	OSH	2.54%
74	VRL	2.53%
75	CQR	2.52%
76	TOL	2.49%
77	CSR	2.47%
78	RMD	2.43%
79	ILU	2.42%
80	APN	2.41%
81	ABC	2.41%
82	PRT	2.40%
83	MQG	2.37%
84	CTX	2.36%
85	LEI	2.36%
86	ALQ	2.33%
87	GPT	2.33%
88	TEN	2.32%
89	BXB	2.32%
90	HIL	2.31%
91	QBE	2.31%

Continued on next page

Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
92	CPU	2.30%
93	NUF	2.29%
94	CSL	2.29%
95	RIO	2.27%
96	JHX	2.26%
97	AAD	2.24%
98	FLT	2.23%
99	FWD	2.20%
100	SGM	2.20%
101	MGR	2.20%
102	IOF	2.20%
103	UGL	2.18%
104	QAN	2.16%
105	PPT	2.16%
106	GWA	2.15%
107	SWM	2.15%
108	RIC	2.14%
109	BHP	2.12%
110	DJS	2.09%
111	FXJ	2.08%
112	HVN	2.06%
113	COH	2.01%
114	GNC	1.99%
115	AMP	1.98%
116	STO	1.97%
117	BLD	1.96%

Continued on next page

Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
118	ENV	1.95%
119	FBU	1.94%
120	MTS	1.94%
121	GUD	1.94%
122	LLC	1.89%
123	SGP	1.88%
124	PRY	1.87%
125	ORI	1.87%
126	CAB	1.86%
127	BEN	1.86%
128	SUN	1.85%
129	AHD	1.85%
130	RHC	1.84%
131	WPL	1.82%
132	ORG	1.81%
133	ANN	1.80%
134	TAH	1.78%
135	BOQ	1.77%
136	CPA	1.74%
137	TEL	1.72%
138	TCL	1.69%
139	BWP	1.69%
140	SHL	1.68%
141	WES	1.68%
142	ASX	1.67%
143	NAB	1.64%

Continued on next page

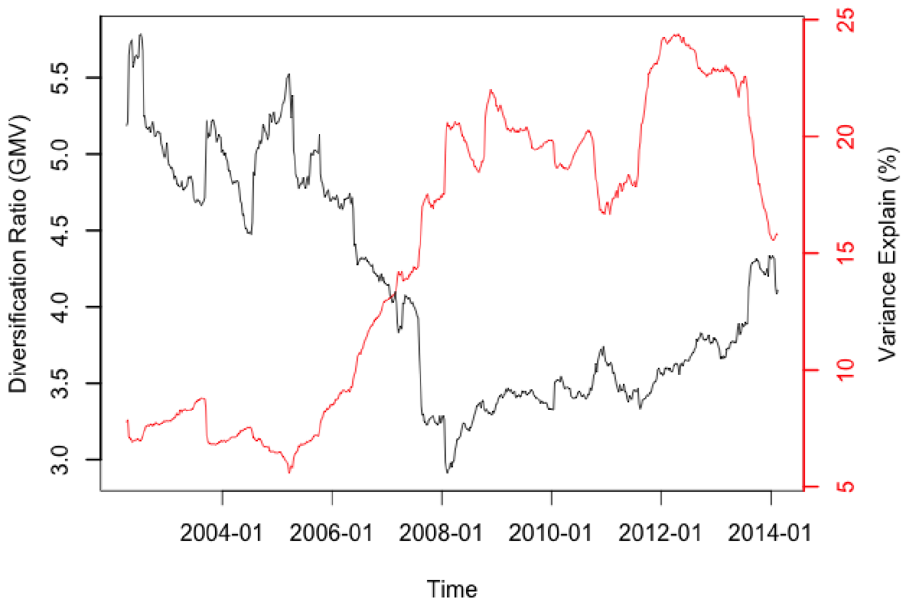
Table B.1 – *Continued from previous page*

	Stocks	Std. Dev.
144	IBC	1.63%
145	CCL	1.61%
146	ANZ	1.61%
147	AMC	1.58%
148	WBC	1.55%
149	CFX	1.49%
150	CBA	1.46%
151	AGK	1.45%
152	DJW	1.43%
153	TLS	1.34%
154	WOW	1.28%
155	AFI	1.21%
156	ARG	1.15%

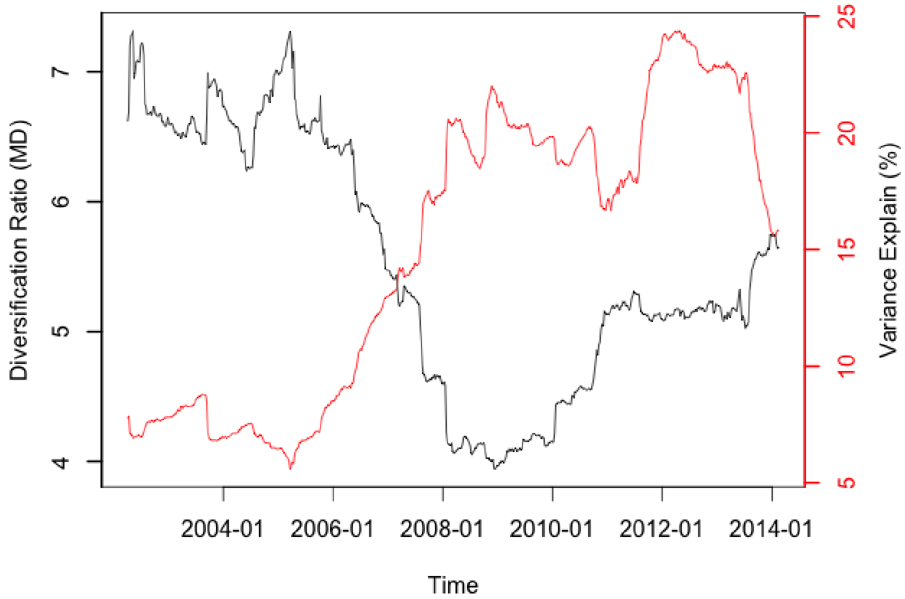
Appendix C

Time evolution of diversification ratio

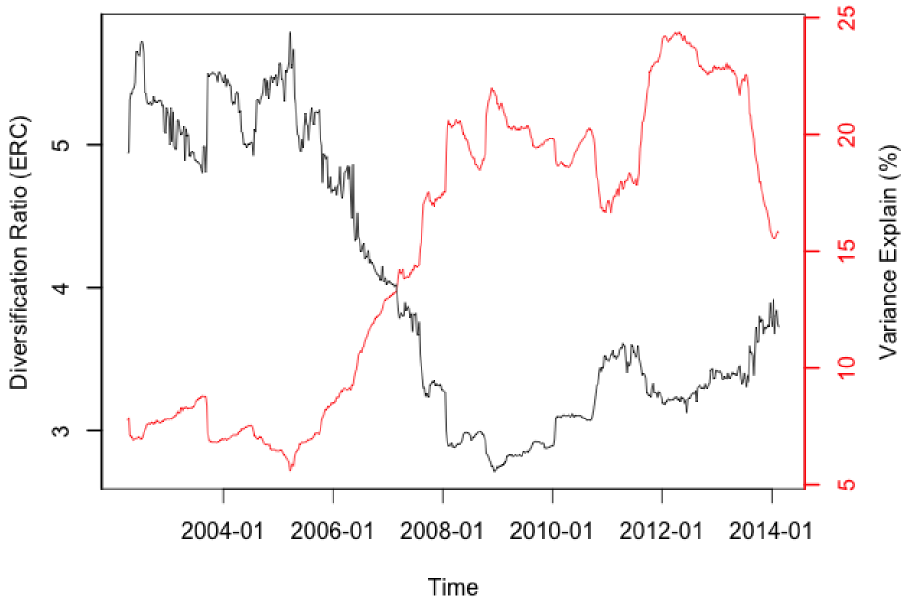
Fig. C.1 Variance explained by principal component 1 with diversification ratio of global minimum variance portfolio, maximum diversification portfolio, and equal risk contribution portfolio respectively. All measures were calculated weekly using a rolling window size of two years (equivalent to 504 trading days) on 156 stocks. The diversification ratios were calculated using $DR_{\omega\in\Omega} = \frac{\omega'\sigma}{\sqrt{\omega'\Sigma\omega}}$.



(a) Diversification ratio of global minimum variance portfolio.



(b) Diversification ratio of maximum diversification portfolio.



(c) Diversification ratio of equal risk contribution portfolio.

Appendix D

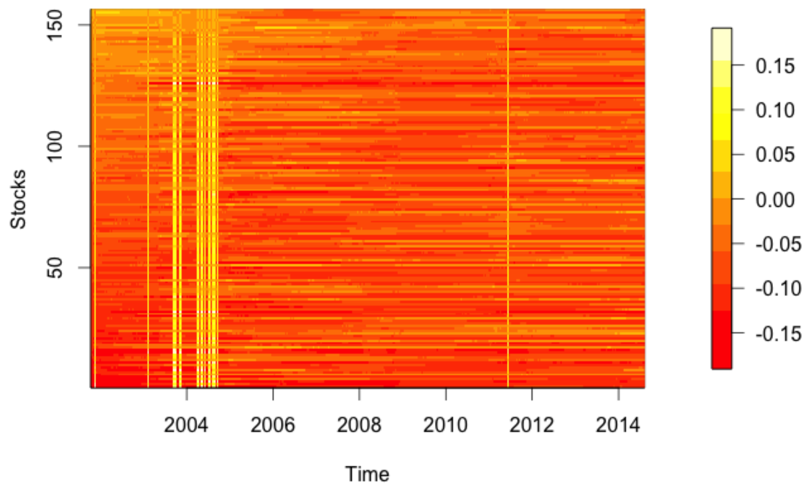
Time evolution of coefficients in principal components: a brief discussion

The coefficients in principal component one were relatively stable over time compared to the other 9 principal components retained (see Figure D.1a). Most stocks had negative coefficients over time except for the periods around 2004 and a day in 2011. In Chapter 7, we pointed out that normally the sign of the coefficients in principal component one were positive because it is a market component with approximately equal contribution of all stocks. But all negative coefficients are still explainable. Principal component one with all negative coefficients is a measure of size. Figure D.1b illustrated that almost all industries have equal importance in principal component one and this was quite robust for different times. This reaffirmed the fact that principal component one represents a market-wide influence and its stability in time.

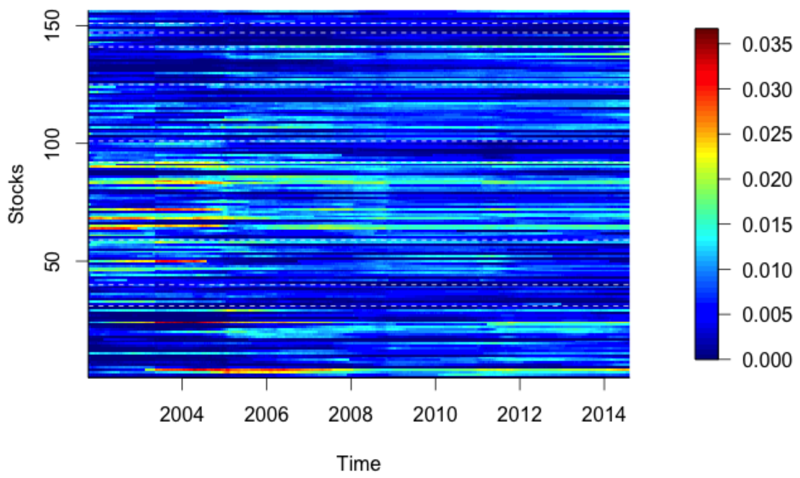
From principal component three, the change in the coefficients across time was more frequent. The coefficients in each principal component changed more often than lower number principal components. There are no patterns evident in the yellow-red colour heat maps from principal component three (see Figure D.1c, Figure D.1e etc.). However, the blue-red colour heat maps, which is the square of coefficients and sorted by industries, exhibit patterns for principal component three and four. In Figure D.1d, we found that principal component three was dominated by Basic Materials and Financials over time. For some periods, it is just one of these two industries dominated the component. From the end of 2003 to the end of 2005, principal component three is dominated by stocks in Basic materials. Around the end of 2005, Basic Materials stopped dominating principal component three and only stocks in Financial industry had significant contributions. Principal component four was mainly dominated by Financial stocks across time except in 2007 when it was dominated by the Oil & Gas industry and in 2014 when it was dominated by Health Care. Interestingly, stocks in Financial industry appeared to be dominants moving between principal component three and four. When the Financial stocks stopped dominated principal component three, they appeared to have high coefficients in principal component four. When the Financial stocks re-dominated principal component three, we can see that the Financial stocks had low coefficients in principal component four. Moreover, the Financial stocks were mostly dominated principal component four when the 2008 financial crisis started. The blue-red colour heat maps for other principal components shows that they are dominated by different stocks in different times and exhibit no patterns related to industries.

The results of our study of the time evolution of the coefficients in principal components revealed that the information contained in particular principal components change over time. This is different from the results of fixed sectors in relevant principal components observed in a static correlation matrix.

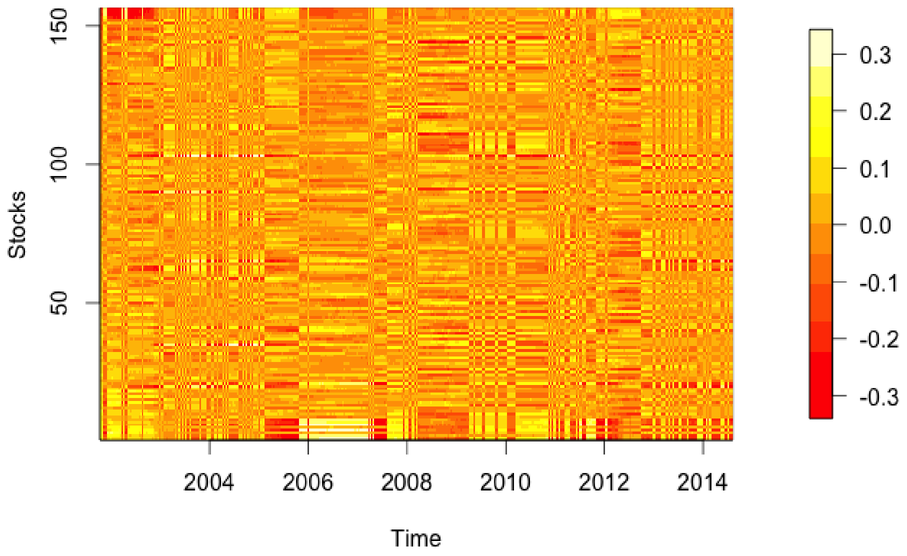
Fig. D.1 Time evolution of the coefficients and square of coefficients respectively in principal component 1 and 3 to 10. The coefficients are obtained from daily performing PCA on 156 stocks for two years data (504 trading days). The horizontal axis are times from 4 April 2002 to 17 February 2014. The vertical axis in the yellow-red colour heat map are the 156 stocks ranked by first day coefficients, which obtained from the first rolling window, 3 April 2000 to 3 April 2002 (the list of stocks with the industry is shown in Appendix C). The vertical axis in the blue-red heat map are 156 stocks order by industry. The order of the industries are: Basic Materials (1-31), Consumer Goods (32-40), Consumer Services (41-59), Financials (60-92), Health Care (93-101), Industrials (102-125), Oil & Gas (126-141), Technology (142-147), Telecommunications (148-151) and Utilities (152-156). Industries are separated by the horizontal dash lines.



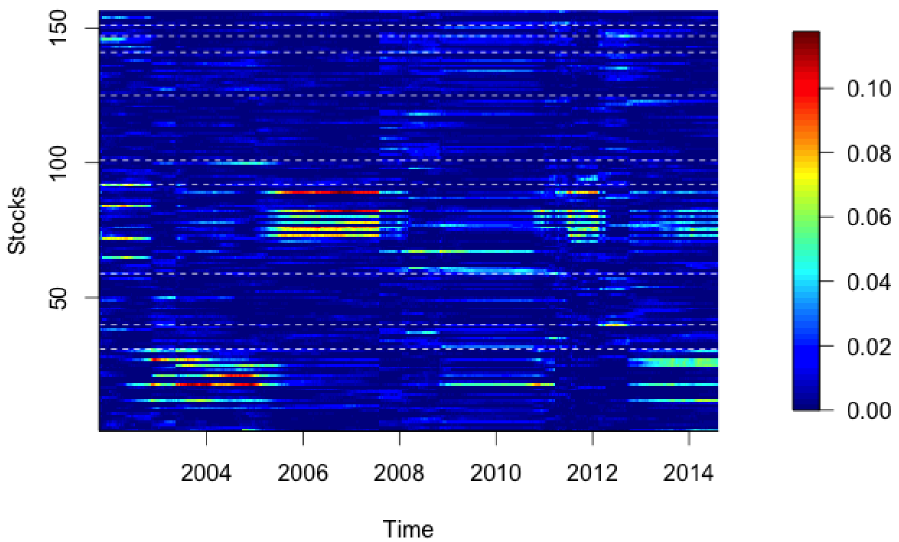
(a) Principal component one, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



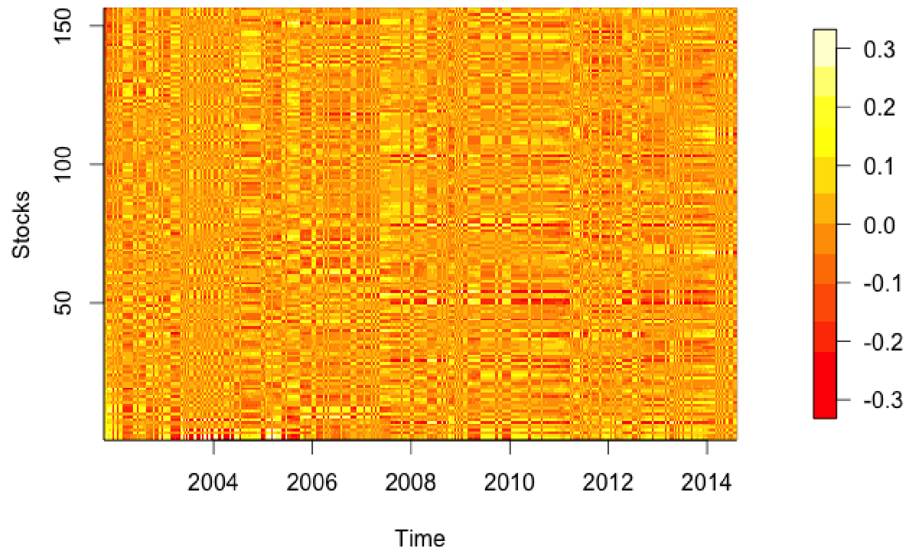
(b) Principal component one, stocks sorted by industry.



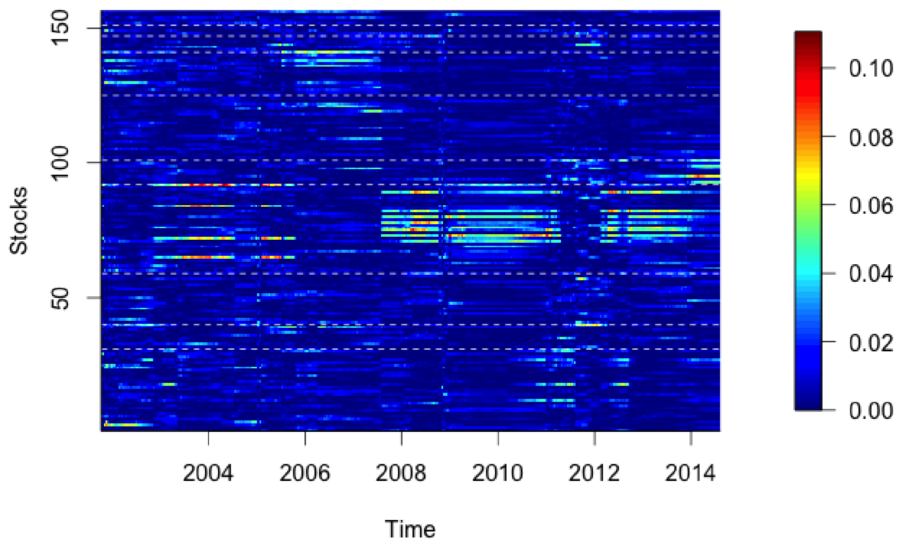
(c) Principal component three, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



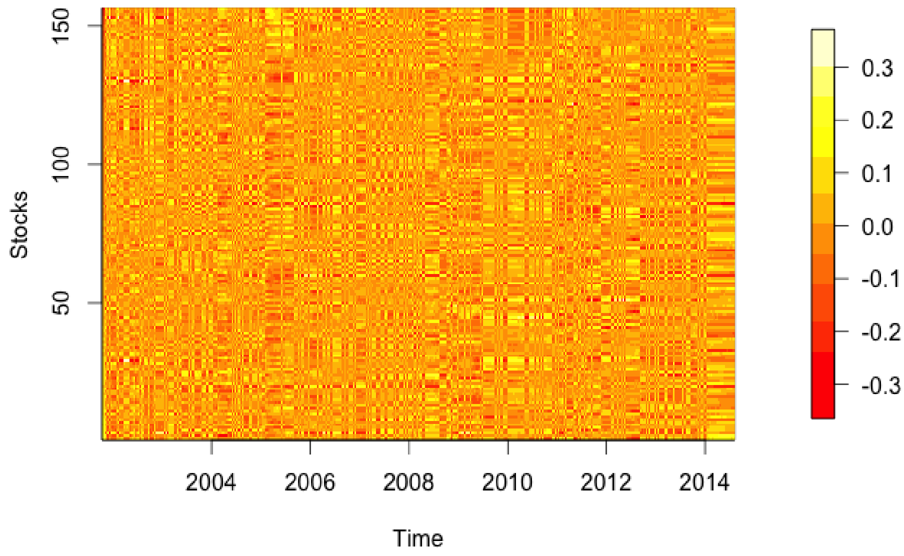
(d) Principal component three, stocks sorted by industry.



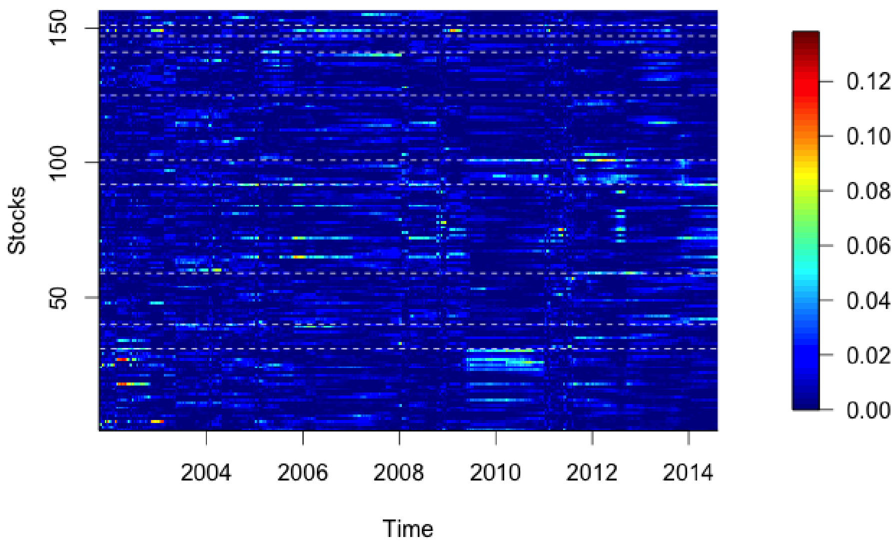
(e) Principal component four, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



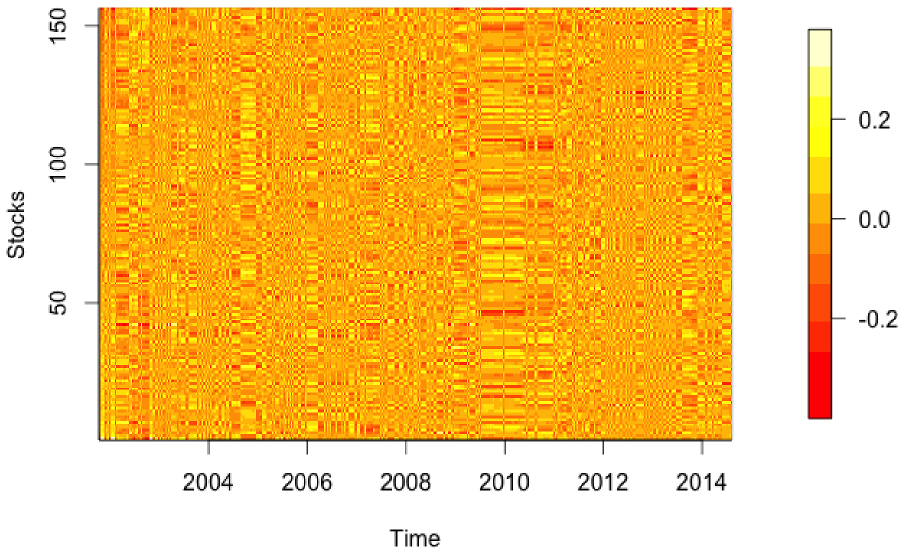
(f) Principal component four, stocks sorted by industry.



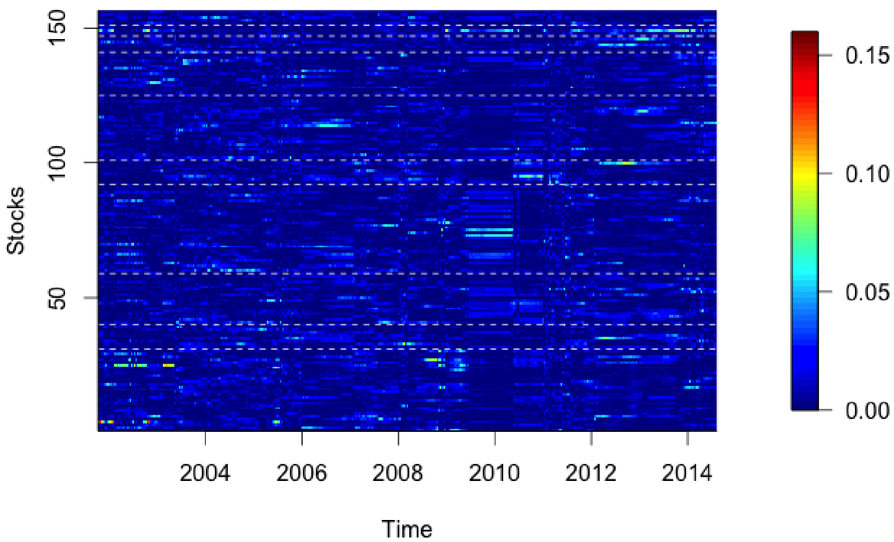
(g) Principal component five, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



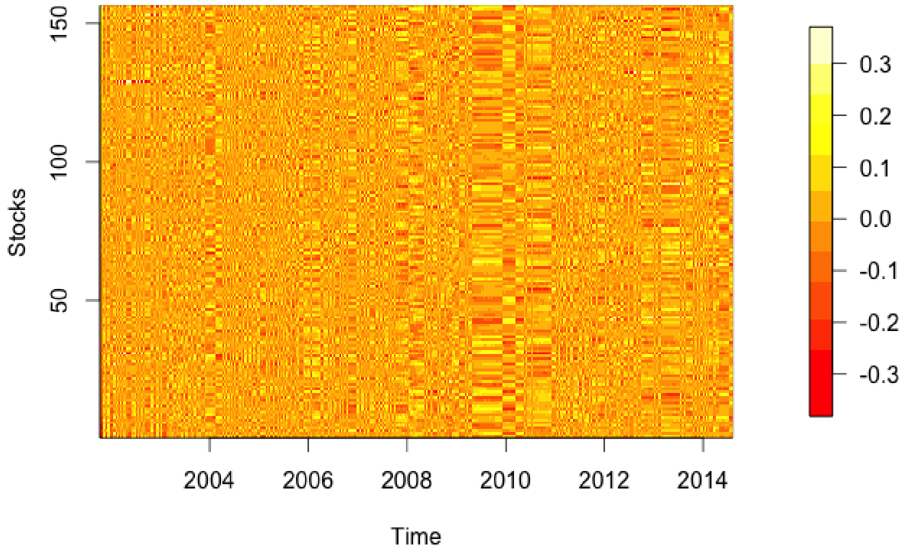
(h) Principal component five, stocks sorted by industry.



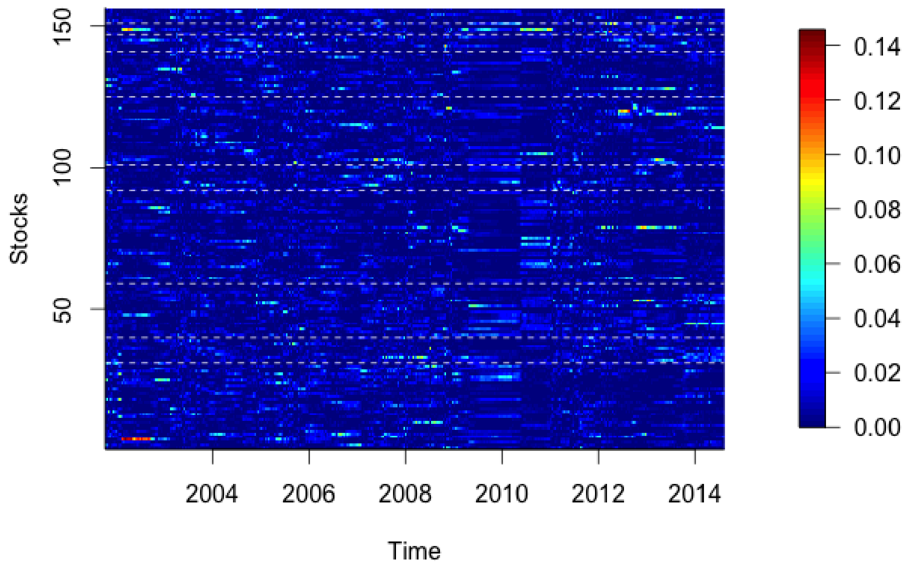
(i) Principal component six, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



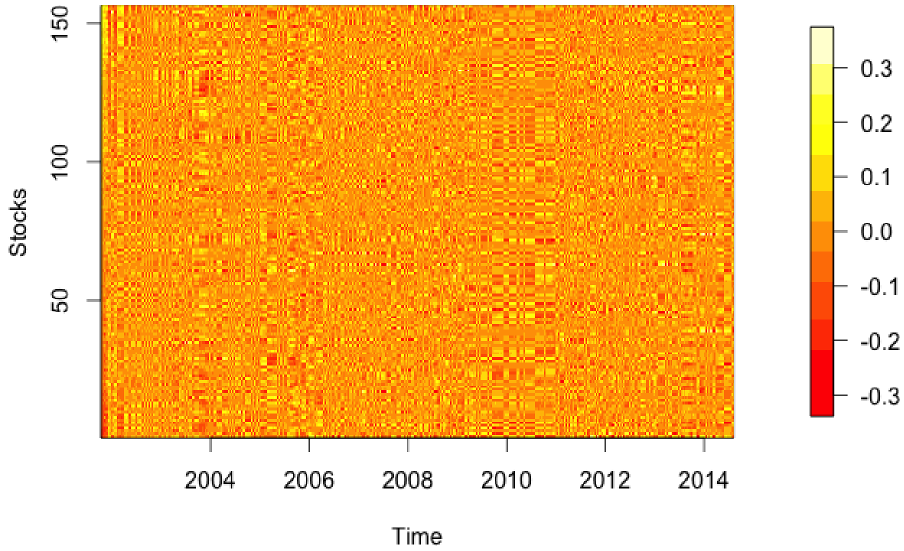
(j) Principal component six, stocks sorted by industry.



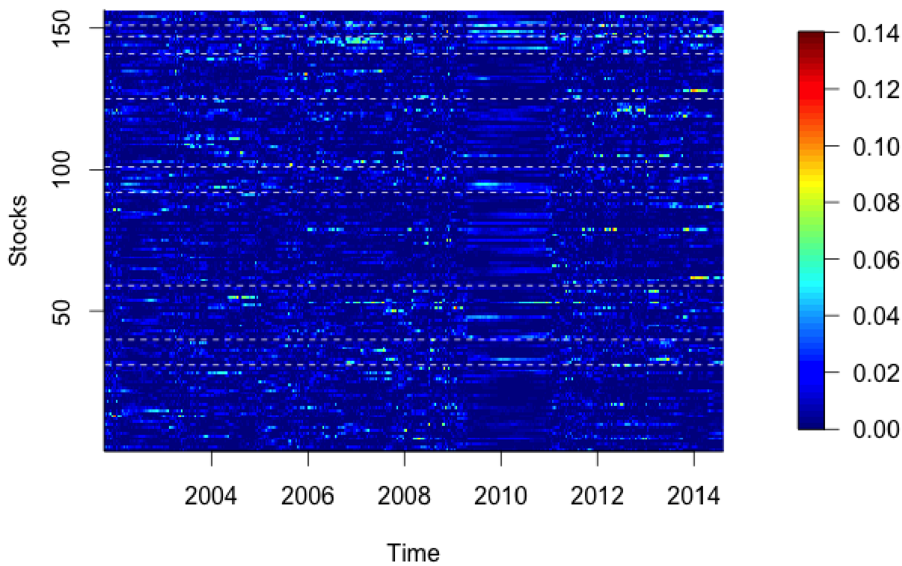
(k) Principal component seven, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



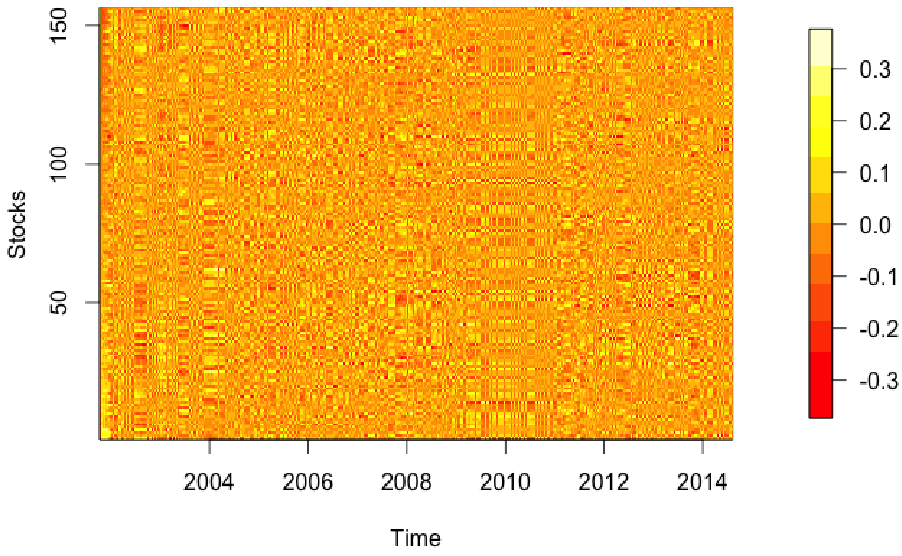
(l) Principal component seven, stocks sorted by industry.



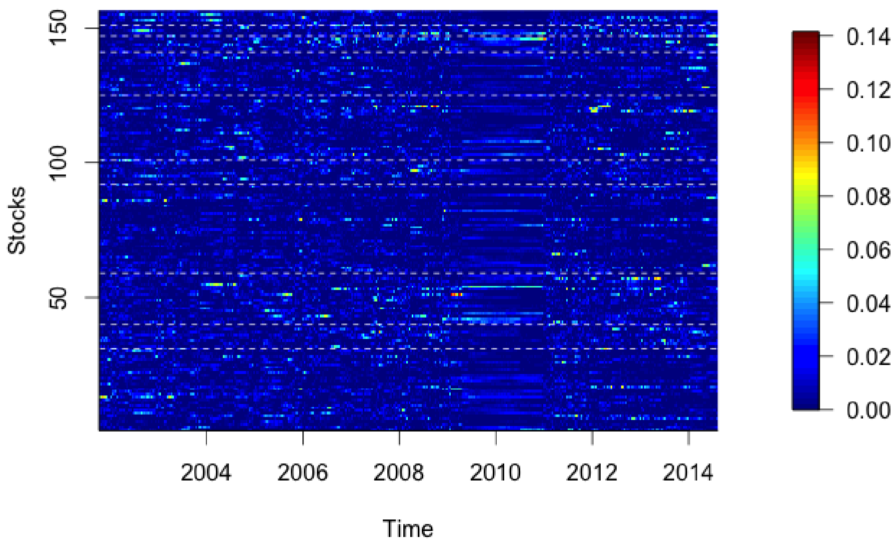
(m) Principal component eight, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



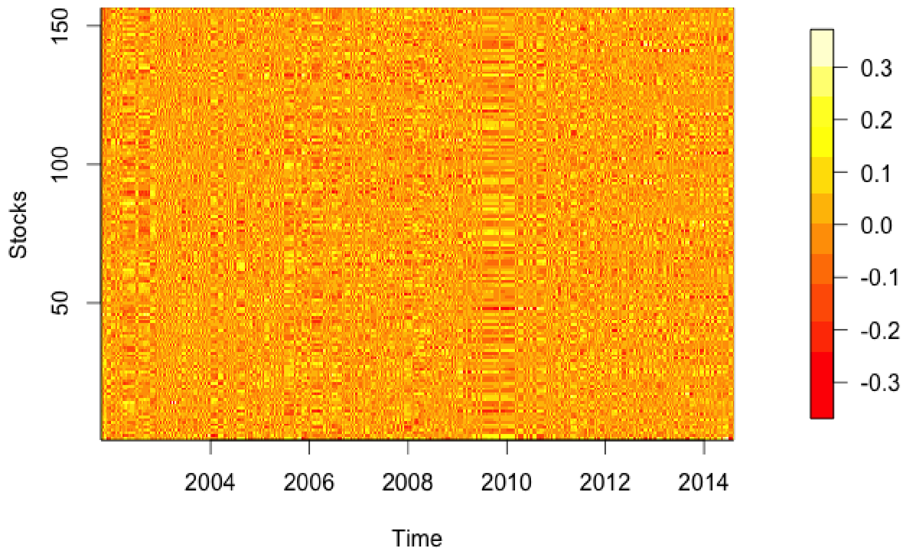
(n) Principal component eight, stocks sorted by industry.



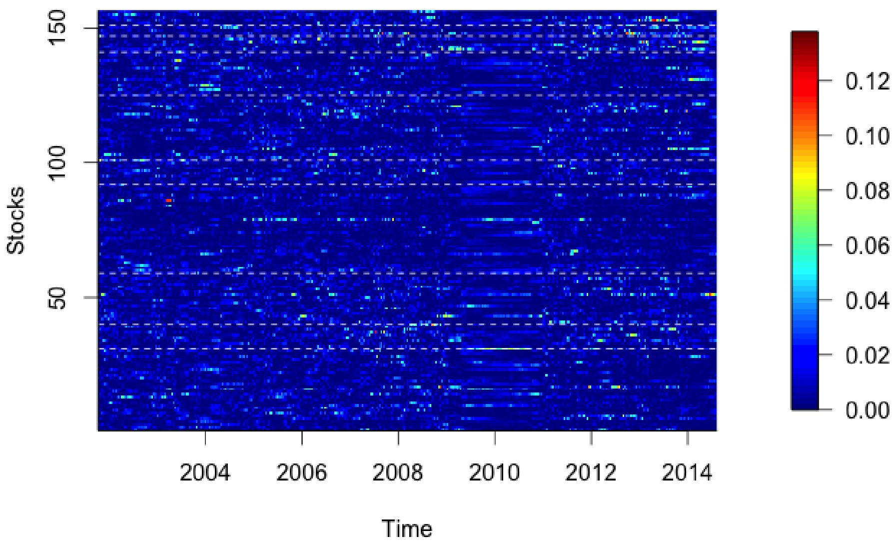
(o) Principal component nine, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



(p) Principal component nine, stocks sorted by industry.



(q) Principal component 10, stocks sorted by first day coefficients, which were obtained from the first rolling window, 3 April 2000 to 3 April 2002 .



(r) Principal component 10, stocks sorted by industry.

Appendix E

A list of the 156 stocks with complete data

Table E.1 A list of the 156 stocks with complete data and thier respective industries, sorted by the coefficients obtained from the first rolling window, 3 April 2000 to 3 April 2002, in principal component 2. This table refers to the stock ordering in Figure 9.1, Figure D.1a, Figure D.1c, Figure D.1e, Figure D.1g, Figure D.1i, Figure D.1k, Figure D.1m, Figure D.1o, and Figure D.1q.

	Stocks	**Industry**
1	MGR	Financials
2	GPT	Financials
3	CFX	Financials
4	SGP	Financials
5	CPA	Financials
6	IOF	Financials
7	WES	Consumer Services
8	CBA	Financials
9	AFI	Financials
10	WBC	Financials
11	NAB	Financials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
12	ANZ	Financials
13	TCL	Industrials
14	COH	Health Care
15	SWM	Consumer Services
16	AGK	Utilities
17	NCM	Basic Materials
18	WOW	Consumer Services
19	CCL	Consumer Goods
20	AMP	Financials
21	JHX	Industrials
22	STO	Oil & Gas
23	BEN	Financials
24	CSR	Industrials
25	ENV	Utilities
26	ILU	Basic Materials
27	TEN	Consumer Services
28	APN	Consumer Services
29	CQR	Financials
30	WPL	Oil & Gas
31	SUN	Financials
32	TAH	Consumer Services
33	SGM	Basic Materials
34	ELD	Consumer Goods
35	BWP	Financials
36	GNC	Consumer Goods
37	BOQ	Financials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
38	PPT	Financials
39	VRL	Consumer Services
40	CSL	Health Care
41	AGG	Basic Materials
42	AAD	Financials
43	CTX	Oil & Gas
44	DJW	Financials
45	SAR	Basic Materials
46	PMP	Industrials
47	AUT	Oil & Gas
48	LEI	Industrials
49	TEL	Telecommunications
50	HVN	Consumer Services
51	GWA	Industrials
52	AHD	Consumer Services
53	AIX	Financials
54	DOW	Industrials
55	OEC	Industrials
56	MQG	Financials
57	LLC	Financials
58	ARG	Financials
59	DJS	Consumer Services
60	ORI	Basic Materials
61	TWR	Financials
62	QBE	Financials
63	ASL	Industrials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
64	SFH	Consumer Services
65	AJL	Industrials
66	NRT	Health Care
67	QAN	Consumer Services
68	ALZ	Financials
69	ERA	Basic Materials
70	OSH	Oil & Gas
71	ENE	Utilities
72	TLS	Telecommunications
73	ALQ	Consumer Goods
74	AMC	Industrials
75	ALL	Consumer Services
76	BXB	Industrials
77	CAA	Basic Materials
78	LYC	Basic Materials
79	SHL	Health Care
80	EWC	Utilities
81	DLS	Oil & Gas
82	CPU	Industrials
83	TOL	Industrials
84	KCN	Basic Materials
85	SDG	Financials
86	NUF	Basic Materials
87	ANN	Health Care
88	AQP	Basic Materials
89	ABC	Industrials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
90	RHC	Health Care
91	GUD	Consumer Goods
92	MGL	Basic Materials
93	ASX	Financials
94	BLD	Industrials
95	HIL	Industrials
96	API	Consumer Services
97	FBU	Industrials
98	ROC	Oil & Gas
99	RMD	Health Care
100	TRY	Basic Materials
101	MTS	Consumer Services
102	CVN	Oil & Gas
103	HTA	Telecommunications
104	RIC	Consumer Goods
105	PRT	Consumer Services
106	IDT	Health Care
107	PNA	Basic Materials
108	ORG	Utilities
109	CAB	Industrials
110	ALU	Technology
111	IBC	Financials
112	TNE	Technology
113	ASB	Industrials
114	IMD	Oil & Gas
115	FMG	Basic Materials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
116	BRG	Consumer Goods
117	UGL	Industrials
118	PRY	Health Care
119	RIO	Basic Materials
120	MAH	Basic Materials
121	SKE	Industrials
122	AVG	Consumer Goods
123	REA	Financials
124	SMX	Technology
125	HZN	Oil & Gas
126	UML	Basic Materials
127	MLB	Technology
128	SLX	Industrials
129	MCR	Basic Materials
130	MAQ	Telecommunications
131	FWD	Consumer Goods
132	SXY	Oil & Gas
133	IAU	Basic Materials
134	CYO	Technology
135	IIN	Technology
136	GBG	Basic Materials
137	AOG	Financials
138	SDL	Basic Materials
139	AZZ	Oil & Gas
140	RSG	Basic Materials
141	OZL	Basic Materials

Continued on next page

Table E.1 – *Continued from previous page*

	Stocks	Industry
142	FLT	Consumer Services
143	TAP	Oil & Gas
144	MPO	Oil & Gas
145	BHP	Basic Materials
146	MRM	Industrials
147	AWE	Oil & Gas
148	AWC	Basic Materials
149	MDL	Basic Materials
150	FXJ	Consumer Services
151	NWS	Consumer Services
152	BPT	Oil & Gas
153	SBM	Basic Materials
154	CDU	Basic Materials
155	PDN	Basic Materials
156	CGF	Financials

Appendix F

A list of the 156 stocks, sorted by industry

Table F.1 A list of the 156 stocks with complete data and their respective industries, ordered by industry, refers to the stock ordering in Figure 9.4, Figure D.1b, Figure D.1d, Figure D.1f, Figure D.1h, Figure D.1j, Figure D.1l, Figure D.1n, Figure D.1p, and Figure D.1r

	Stocks	Industry
1	AGG	Basic Materials
2	AQP	Basic Materials
3	AWC	Basic Materials
4	BHP	Basic Materials
5	CAA	Basic Materials
6	CDU	Basic Materials
7	ERA	Basic Materials
8	FMG	Basic Materials
9	GBG	Basic Materials
10	IAU	Basic Materials
11	ILU	Basic Materials
12	KCN	Basic Materials

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
13	LYC	Basic Materials
14	MAH	Basic Materials
15	MCR	Basic Materials
16	MDL	Basic Materials
17	MGL	Basic Materials
18	NCM	Basic Materials
19	NUF	Basic Materials
20	ORI	Basic Materials
21	OZL	Basic Materials
22	PDN	Basic Materials
23	PNA	Basic Materials
24	RIO	Basic Materials
25	RSG	Basic Materials
26	SAR	Basic Materials
27	SBM	Basic Materials
28	SDL	Basic Materials
29	SGM	Basic Materials
30	TRY	Basic Materials
31	UML	Basic Materials
32	ALQ	Consumer Goods
33	AVG	Consumer Goods
34	BRG	Consumer Goods
35	CCL	Consumer Goods
36	ELD	Consumer Goods
37	FWD	Consumer Goods
38	GNC	Consumer Goods

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
39	GUD	Consumer Goods
40	RIC	Consumer Goods
41	AHD	Consumer Services
42	ALL	Consumer Services
43	API	Consumer Services
44	APN	Consumer Services
45	DJS	Consumer Services
46	FLT	Consumer Services
47	FXJ	Consumer Services
48	HVN	Consumer Services
49	MTS	Consumer Services
50	NWS	Consumer Services
51	PRT	Consumer Services
52	QAN	Consumer Services
53	SFH	Consumer Services
54	SWM	Consumer Services
55	TAH	Consumer Services
56	TEN	Consumer Services
57	VRL	Consumer Services
58	WES	Consumer Services
59	WOW	Consumer Services
60	AAD	Financials
61	AFI	Financials
62	AIX	Financials
63	ALZ	Financials
64	AMP	Financials

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
65	ANZ	Financials
66	AOG	Financials
67	ARG	Financials
68	ASX	Financials
69	BEN	Financials
70	BOQ	Financials
71	BWP	Financials
72	CBA	Financials
73	CFX	Financials
74	CGF	Financials
75	CPA	Financials
76	CQR	Financials
77	DJW	Financials
78	GPT	Financials
79	IBC	Financials
80	IOF	Financials
81	LLC	Financials
82	MGR	Financials
83	MQG	Financials
84	NAB	Financials
85	PPT	Financials
86	QBE	Financials
87	REA	Financials
88	SDG	Financials
89	SGP	Financials
90	SUN	Financials

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
91	TWR	Financials
92	WBC	Financials
93	ANN	Health Care
94	COH	Health Care
95	CSL	Health Care
96	IDT	Health Care
97	NRT	Health Care
98	PRY	Health Care
99	RHC	Health Care
100	RMD	Health Care
101	SHL	Health Care
102	ABC	Industrials
103	AJL	Industrials
104	AMC	Industrials
105	ASB	Industrials
106	ASL	Industrials
107	BLD	Industrials
108	BXB	Industrials
109	CAB	Industrials
110	CPU	Industrials
111	CSR	Industrials
112	DOW	Industrials
113	FBU	Industrials
114	GWA	Industrials
115	HIL	Industrials
116	JHX	Industrials

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
117	LEI	Industrials
118	MRM	Industrials
119	OEC	Industrials
120	PMP	Industrials
121	SKE	Industrials
122	SLX	Industrials
123	TCL	Industrials
124	TOL	Industrials
125	UGL	Industrials
126	AUT	Oil & Gas
127	AWE	Oil & Gas
128	AZZ	Oil & Gas
129	BPT	Oil & Gas
130	CTX	Oil & Gas
131	CVN	Oil & Gas
132	DLS	Oil & Gas
133	HZN	Oil & Gas
134	IMD	Oil & Gas
135	MPO	Oil & Gas
136	OSH	Oil & Gas
137	ROC	Oil & Gas
138	STO	Oil & Gas
139	SXY	Oil & Gas
140	TAP	Oil & Gas
141	WPL	Oil & Gas
142	ALU	Technology

Continued on next page

Table F.1 – *Continued from previous page*

	Stocks	Industry
143	CYO	Technology
144	IIN	Technology
145	MLB	Technology
146	SMX	Technology
147	TNE	Technology
148	HTA	Telecommunications
149	MAQ	Telecommunications
150	TEL	Telecommunications
151	TLS	Telecommunications
152	AGK	Utilities
153	ENE	Utilities
154	ENV	Utilities
155	EWC	Utilities
156	ORG	Utilities