Stock market anomalies in emerging markets

Allard Vollering (326181)

Erasmus University Rotterdam
Master Thesis Econometrics and Management Sciences:
Quantitative Finance

January 2016

This thesis investigates the presence and behavior of stock market anomalies in emerging markets and the objective is to answer two questions, both directly related to investing in anomalies and (emerging market) stocks. The first question is; how do anomalies behave in emerging markets? And the second, can anomalies be used effectively to forecast the stock returns? These questions are researched with hedge portfolios, (Fama-MacBeth) cross-sectional regressions, and point-forecasts (with various variable specifications and calculated with shrinkage methods).
This thesis finds that anomalies are still present in emerging markets when forming hedge portfolios and in cross sectional characteristic premiums. This is reflected by significant results on some of the hedge portfolios, even after trading costs. The characteristic premiums are significant, and some have changed over time. Using the anomalies directly in forecasting stock returns is not a fruitful exposition.

Keywords: stock market, anomalies, emerging markets, cross-section, shrinkage

Supervised by
Dr. Wing Wah Tham
Dr. Erik Kole (co-reader)

Contents

1 Introduction

This thesis investigates the presence and behavior of stock market anomalies in emerging markets and the objective is to answer two questions, both directly related to investing in anomalies and (emerging market) stocks. The first question is, how do anomalies behave in emerging markets? And the second, can anomalies be used effectively to forecast the stock returns? The first research question is broad and is divided into three sub-questions to answer it. First, do anomalies have significant influence on stock market returns in emerging markets? Second, is the effect of anomalies in emerging markets constant or time variant? Third, do anomalies in emerging stock markets differ from developed country stock markets?

In the last decades, many authors have attempted to forecast and explain asset returns. One of the oldest examples is the Capital Asset Pricing Model (Sharpe, 1964) which explains asset returns by their correlation to the market, β. The field of study has expanded much ever since, and the number of factors that show significant effect in explaining asset returns has risen to 314 in 2012 (Harvey, Liu and Zhu, 2014). Not only are these factors able to explain the cross section of stock returns in a fair extent, some are able to forecast the direction of stock returns. Some of the most important factors are the value (Fama and French, 1993) and momentum (Jegadeesh and Titman, 1993) factors. While factors should not be able to provide risk adjusted profits in an efficient market but often are, they make research in this topic especially interesting.

Chordia, Subrahmanyam and Tong (2014) argue that the increase in liquidity, trading activity and technological trading innovations in the United States have facilitated arbitrage which makes markets more efficient and should have led to a decline in anomaly effect due to an increase in market efficiency. The authors found that the anomaly effects have attenuated in the United States. This raises the question if and how anomalies are present in emerging markets, as the anomaly effects might differ from the United States.

The first (research) sub-question, how do anomalies behave in emerging markets, is answered by an analysis which is similar to that of Chordia et al. (2014) who research capital market anomalies in the United States. Hedge portfolios are formed based on the anomalies (stock characteristics) which are examined for significance of their returns. The anomaly effects are moreover examined with a two stage approach (Fama-MacBeth style regressions), which enables to draw conclusions about size and significance of the anomaly effects in the cross section of stocks.

To answer the second sub-question, if the anomaly effects are constant or time varying, an additional analysis is conducted based on Chordia et al. (2014). The authors research the hedge portfolios returns and Fama-MacBeth coefficients for trends to draw conclusions about the time (in)variance of anomaly returns. The original paper searches for accentuation or attenuation but it is a distinct possibility that the effects are constant over time in emerging markets. They found that the significance of most factors attenuated for US stocks over the last decades. To answer the third sub-question, whether anomaly effects in emerging markets differ from developed markets, a qualitative comparison to Chordia et al. (2014) (among other authors) is made.

To answer the second research question a selection of methods from Kim and Swanson (2014) is used to model the return forecasts by implementing shrinkage methods. That paper states a variety of models that can cope with a large number of explanatory variables to forecast macroeconomic variables by using soft thresholding. Here, ridge regressions, LASSO and elastic net models are used, the advantage over for example OLS is that these models add bias but reduce variance so they are better in sense of mean squared error. To make answering this question computationally feasible, a sub-sample of emerging market stocks is used: the MSCI Emerging Markets (MSCI EM) index.

These methods will be supplemented by a dynamic factor model based on PCA which lacks practical interpretation of coefficients but could add to a forecasting model as it is a method of variable reduction. A last method of constructing forecasts is by augmented autoregressive models, with the anomalies and PCA (on the anomalies).

These forecasts will be subsequently used to form mean-variance efficient portfolios, optimized to satisfy an objective such as minimum variance or maximum diversification. To make sure the covariance matrix of asset returns is invertible, the covariance matrix shrinkage method of Ledoit and Wolf (2003b).

The literature of anomalies (and factors) has grown exponentially, which makes it increasingly difficult to define a selection of anomalies that is not attributable to data mining (see for example Harvey et al., 2014). In this paper, the anomaly subset is based on Chordia et al. (2014) where the authors define a selection of anomalies with good performance history and broad academic support in the United States. That results in the following anomalies: asset growth, illiquidity, issuance, idiosyncratic volatility, size, short-term reversal, momentum, value and turnover.

Anomalous returns are studied by many authors, often these papers are focused on one anomaly. Most research is concentrated in financial markets of developed countries such as the United States. So opportunities for further exploration lie within emerging markets, which have inherently different characteristics in metrics such as variance and returns (Harvey, 1995).

Anomalous returns have been studied by many authors, often these papers are focused on one anomaly and not on a broad selection of anomalies. Most research is concentrated in financial markets of developed countries such as the United States. So opportunities for further exploration lie in emerging markets, which have inherently different characteristics in metrics such as variance and average returns (Harvey, 1995). Using these methods add to the literature by the examination of this broad set of anomalies in a broad sample of emerging market stocks (from various countries) and the analysis of the trend in these anomaly exposures.

Estimating the sensitivity of the stocks to their characteristics in order to make forecasts creates a different framework than investing directly in hedge portfolios. Using robust estimation techniques with stock specific (anomaly) characteristics this thesis adds to the literature by using the anomalies in a way that closes the gap between asset pricing and econometric forecasting. The formation of portfolios based on the forecasts for individual stocks adds to the literature because taking the MSCI EM index as a benchmark and calculating the portfolio weights based on shrinkage forecasts has not been done yet.

This paper finds that anomalies are still present in emerging markets when forming hedge portfolios and in cross sectional characteristic premiums. This is reflected by significant results on some of the hedge portfolios, even after trading costs. The characteristic premiums are significant, and some have changed over time.

Using the anomalies in forecasting stock returns is not a fruitful exposition. The anomaly based models perform worst, but by using the principal components from the (stock specific) anomalies as independent variables the accuracy is increased. The autoregressive specification with and without principal components perform best and supersede the direct anomaly specifications. All of the optimized portfolios based on forecasts struggle to beat the benchmark indices based on return and Sharpe ratio. Volatility is lowered in most cases.

This paper proceeds as follows; the emerging markets data is described in Chapter 2 and the methodology is stated in Chapter 3. The results from the analysis are reported in Chapter 4 and Chapter 5 discusses and concludes.

2 Data

To specify a sample of emerging countries the 1998 World Bank definition is used which differentiates between high (developed), middle and low (emerging) income countries (Griffin, Nardari and Stulz (2007), among other authors use this definition). This results in a sample of 20 countries, i.e. Argentina, Brazil, Chile, China, Colombia, Czech Republic, Hungary, India, Indonesia, Malaysia, Mexico, Peru, Philippines, Poland, Russia, South Africa, South Korea, Thailand, Turkey and Venezuela.

To keep the list of countries in line with the MSCI Emering Market index ('MEM sample'), which will be used in this paper as a benchmark and baseline model to form portfolios, the following countries are added to the sample: Egypt, Hong Kong, Israel, Jordan, Morocco, Pakistan and Taiwan. These countries are selected to have at least one stock listed in the MEM sample (on average) over the sample period. The addition of these countries increases the total to 28 countries.

The data to construct the returns, and anomaly variables for the 28 countries is retrieved from the Datastream database. The sample consists of listed and delisted stock to mitigate the risk of survivorship bias. The number of securities in a first run selection was approximately 29,000, and after an extensive screening which made sure that only ordinary shares in their home market are included and investment vehicles (REITs, investment trusts, and other non equity traded products) are excluded. This screening process is based on the comprehensive process of Griffin, Kelly and Nardari (2010), in which the authors want to use ordinary stock data for their analysis. The constituents and their weights in the MSCI Emerging Markets index is retrieved from Bloomberg from January 1995 until December 2014.

The final large sample consists of 23,644 stocks that had traded in the period from January 1994 to December 2014 (252 months). This sample gives a broad cross section to test the first two hypotheses on. The MEM sample consists of a total of 1,026 stocks, which is a sample that allows to fit a variety of models and optimize portfolios. Stock returns are available for 252 months as for the large sample. The constituents and weights for the MSCI Emerging Market index are available from January 1995 so this slightly shorter sample is used to form the portfolios on.

To make sure the results are driven by stock returns instead of inflation, all prices are converted to US Dollar before the calculation of the returns. In addition, an investor wants to have exposure to stocks and usually not to emerging market currency risk. The summary statistics for (the returns of) both samples are in Table 1.

The anomalies are based on the paper of Chordia et al. (2014), the authors use a list of 12 anomalies for their analysis in the United States and in this paper nine of the anomalies will be analyzed. Investigating the anomalies and the Datastream database showed that for three[1] balance sheet based anomalies insufficient data would be available (less than 3% data availability per stock in traded months) from which the conclusion had to be drawn that these should be set aside for this paper. This paper investigates nine stock market anomalies[2] which are defined as follows:

1. AG: Asset growth, as in Cooper, Gulen and Schill (2008) which is defined as the year-on-year growth of total assets.
2. BM: Book equity divided by market equity, as defined in Fama and French (1993).
3. ILLIQ: A measure for illiquidity that follows Amihud (2002). It measures the impact of absolute price changes per volume traded.

$$ILLIQ_d = \frac{1}{D_{it}} \sum_{d=1}^{D_{it}} \frac{|R_{itd}|}{DVOL_{itd}} \times 10^6 \tag{1}$$

4. ISSUE: New issues, as defined by Pontiff and Woodgate (2008), defined as the logarithm of number of shares divided by the number of shares eleven months ago.
5. IVOL: Idiosyncratic volatility, as proposed by Ang, Hodrick, Xing and Zhang (2006) who define the variable as the standard deviation of the regression residuals of a factor model. The variable is calculated as the standard deviation of the regression residual of a *local*

[1] Accounting accruals, profitability and standardized unexpected earnings

[2] Asset growth, book-to-market, idiosyncratic volatility, illiquidity, momentum, new issues, short term reversal, size, and turnover

Captial Asset Pricing Model (Sharpe, 1964) within a month. Several authors (for example Fama and French (1993)) have showed that R^2 is high for the CAPM and for the purpose of this analysis does not seem to be inferior to for example a Fama French model. Furthermore, the return variance is higher in emerging markets Harvey (1995) which makes the usage of the Fama French model unnecessary due to the inherent higher variance.

6. R1: The lagged one month return as in Jegadeesh (1990).

7. R212: The 11 month return lagged two months as in Jegadeesh and Titman (1993).

8. SIZE: Size as in Banz (1981), defined as the logarithm of the market value of a stock.

9. TURN: Turnover, measured as the logarithm of the trading volume divided by the number of shares outstanding (as in Datar, Naik and Radcliffe, 1998).

The anomaly variables are winsorized at the 0.005 and 0.995 percentiles and set to those levels, to limit the impact of extreme values on the results. Returns are truncated at the 0.01 and 0.99 percentile to make sure the results are not driven by extreme observations, following Barry, Goldreyer, Lockwood and Rodriguez (2002). Stocks that have zero price, no volume data or less than 5,000 shares outstanding are deleted on the related points in time. Only stocks that traded in the 12 months before an observation are included in the analysis, as in Chordia et al. (2014). In addition, at most three missing values are allowed to be interpolated from adjacent data.

The average value, median and standard deviation of the stock characteristics are in Table 1. The statistics are calculated as the time series means of the cross sectional statistics over the full sample. The statistics in the large sample differ from the results of Chordia et al. (2014) (in parenthesis), size is lower with 417 million USD (2580 million), asset growth, book-to-market, turnover, idiosyncratic volatility, and issuance are higher with 0.2 (0.14), 1.1 (0.8), 0.2 (0.09), 0.08 (0.02), and 0.06 (0.02) respectively. Average reversal returns are lower for the emerging markets sample, with R1 0.005 (0.012) and R212 0.09 (0.17). The difference in size suggests that there are on average smaller firms in emerging markets than in the US. The effects of most other statistics are more pronounced and have higher variance than in the US which shows that the sample is diverse and has different characteristics than a sample of US stocks. The statistics for the MEM sample are more in line with the results from Chordia et al. (2014).

Panel A	Returns			Panel B	Returns		
Mean	0.005			Mean	0.012		
SD	0.161			SD	0.121		
Avg #	11191			Avg #	726		
Min #	5530			Min #	373		
Max #	15415			Max #	942		
Panel C	Mean	Median	SD	Panel D	Mean	Median	SD
AG	0.200	0.080	0.611	EM.AG	0.209	0.116	0.497
BM	1.076	0.752	1.941	EM.BM	0.802	0.602	0.911
ILLIQ	211.883	19.185	501.173	EM.ILLIQ	78.893	10.315	213.946
ISSUE	0.061	0.000	0.279	EM.ISSUE	0.088	0.000	0.302
IVOL	0.084	0.037	0.098	EM.IVOL	0.072	0.028	0.085
R1	0.005	-0.003	0.161	EM.R1	0.012	0.004	0.121
R212	0.088	-0.034	0.563	EM.R212	0.173	0.060	0.503
SIZE.USD	417	39.156	1328	SIZE.USD	2900	1327	3200
TURN	0.212	0.026	0.688	EM.TURN	0.204	0.053	0.607

Table 1: Panel A: The mean and standard deviation of the large sample, with the average, minimum and maximum number of stocks per time period. Panel B: The mean and standard deviation of the MEM sample, with the average, minimum and maximum number of stocks per time period. Panel C: Gives the mean, median and standard deviation for the nine anomalies in the large sample. Size is denoted in million US Dollar. Panel D: Gives the mean, median and standard deviation for the nine anomalies in the MEM sample. Size is denoted in million US Dollar.

3 Methodology

This chapter introduces the methodology that is used in this thesis. The anomalies will be analyzed in the cross section which is explained in Section 3.1. Forecast models based on shrinkage methods will be explained in Section 3.2 and the portfolio specifications will be explained in Section 3.3.

3.1 Cross section

The cross section of emerging market stock will be researched by their anomaly return and risk premium, with respectively the formation of hedge portfolios and Fama-MacBeth coefficients in Section 3.1.1. To research the time variation of the anomaly returns and risk premia, hedge portfolio returns and Fama-MacBeth coefficients are analyzed for a trend with an exponential model in Section 3.1.2.

3.1.1 Hedge portfolios and Fama-MacBeth regressions

Hedge portfolios are formed based on the anomaly values which are long the high anomaly value and short the low anomaly value. The methodology is based on Chordia et al. (2014) but it is fairly common in practice, for example Fama and French (1993) and Jegadeesh and Titman (1993) use a similar approach. By forming such a portfolio, market movements should be averaged out and only the anomaly return should remain. The hedge portfolio returns are calculated by taking long the 10% stocks with the highest anomaly value and short the 10% with the lowest anomaly value. Following Chordia et al. (2014), the characteristics are lagged two periods (except R212 which is already lagged two periods, and R1 that is already lagged one period), to make sure they use only data that is available on the moment of investment. The holding period of the portfolio is one month.

The methodology for the Fama-MacBeth regressions is based on Brennan, Chordia and Subrahmanyam (1998). The authors adjust the model from Fama and MacBeth (1973) which is based on risk factors to incorporate stock specific characteristics which should make it less sensitive to portfolio grouping approaches (Lo and MacKinlay, 1990). The model calculates risk adjusted returns based on five statistical factors in the first step, and regresses these returns cross-sectionally on stock specific anomaly characteristics in the second step. The statistical factors are the five factors that explain most of the variation in the returns, based on asymptotical principal component analysis as in Connor and Korajczyk (1988). The motivation of this procedure is that as the number of assets N is larger than the number of time periods T, regular principal component analysis is not feasible. Therefor this procedure estimates PCA on the $(T \times T)$ matrix $\hat{\Omega}_T$, which is essentially a covariance matrix for time (instead of assets). This method is showed to converge to the regular PCA estimates for large N. The asset returns are then regressed on the factors, and the returns are scaled with the inverse of the residual standard deviation to calculate a refined covariance matrix $\hat{\Omega}_T^*$. The choice for five factors follows Brennan et al. (1998).

Fama and MacBeth (1973) introduced a method to calculate the factor premium in a cross section of stocks. The idea is that the stocks have constant sensitivity to the factors, but that market wide influence of the characteristic can change over time. The first step in that procedure calculates the time series coefficients on the relevant (risk) factors, and the second step estimates the cross sectional risk premia from the first step coefficients on the returns. In Fama and MacBeth (1973) the premium is expected to be constant and measurement errors in the first step regressions do not influence the final estimates. These assumptions do not seem feasible while the second step regressions are based on the regression estimates from the first step, the approach of Brennan et al. (1998) solves this problem by making sure the second step estimates are not driven by error in variable, due to the estimation uncertainty in the first step.

The method in this paper is based on a two step regression procedure: The first step calculates the risk adjusted returns with regard to risk factors ($F_{M,t}$) and the risk free rate, for each stock n (in 1 to N). The sensitivity to the factors is calculated with a rolling window of at least 30 and at most 60 observations. The sensitivity $\beta_{m,n,t}$ of the returns to the risk factors is estimated via asymptotical PCA:

$$R_{n,t} = \beta_{1,n,t} F_{1,t} + ... + \beta_{M,n,t} F_{M,t} + \varepsilon_t \tag{2}$$

Where $R_{n,t}$ is the excess return over the risk free rate (3 month T-bill rate). The risk adjusted returns are then calculated as:

$$R^*_{n,t} = R_{n,t} - \sum_{m=1}^{M} \hat{\beta}_{i,n,t} F_{m,t} \tag{3}$$

The second step estimates the cross sectional premium on the stock specific anomalies in each time period t (in 1 to T) by regressing the returns on the stock specific characteristics $Z_{k,n}$ for each characteristic k in 1 to K and each stock n in 1 to N:

$$R^*_n = c_{0,n} + c_1 Z_{1,n} + ... + c_K Z_{K,n} + \varepsilon_n \tag{4}$$

The statistics are then calculated as the time series averages of the estimates c_k, and the standard errors as the sample standard deviation $\sigma(\hat{c}_k)/\sqrt{T}$. The coefficients in the second stage regressions are standardized to have zero mean and unit standard deviation, in this way the coefficients are interpretable as increases in returns per standard deviation increase. The main advantage is that the coefficients from this analysis are directly interpretable as premia and show what the compensation has been for each of the anomalies.

3.1.2 Time variation

To research the if the effect of anomaly returns and premia from section 3.1.1 changed over time, these will be tested for an exponential trend. This approach is similar to Chordia et al. (2014). This approach tests the following model:

$$y_t = a \exp(bt + \varepsilon_t), \tag{5}$$

with y_t the time series of anomaly returns or premia, and t the equally spaced series on the interval $[-1, 1]$. The trend is calculated and assessed on its significance. For the Fama-MacBeth coefficients the trend is analyzed in the premia, after the second step regression.

3.2 Forecasts

Five models (based on the anomalies, Principal Components and Autoregressive lags, in various specifications) are fitted using four estimation techniques (Ordinary Least Squares, Least Angle Shrinkage and Selection Operator, Ridge Regressions and Elastic Net). The models will be used to forecast stock returns on a monthly basis, and the models are assessed based on their Mean Squared Prediction Error, Mean Absolute Error and Root Mean Squared Error. Variable selection would be possible using hard thresholding methods but the choice is made to implement soft thresholding methods in this thesis, as the results (coefficient estimates) are expected to be smoother. To test the forecasts on directional accuracy the Pesaran and Timmermann (1992) (PT) statistic will be calculated, and to assess the forecast accuracy mutually the Diebold and Mariano (1995) (DM) statistic will be calculated. The following will introduce the estimation techniques and the statistics.

3.2.1 Estimation

The first method is Ordinary Least Squares (OLS), this is one of the most widely used estimation techniques. It is often used as a baseline for more advanced techniques as it gives unbiased estimates if the data generating process is linear and allows for straight forward estimation. Moreover, it has a closed form solution which makes it efficient to fit the coefficients. A general linear model is used;

$$y_t = \alpha + \beta_1 x_{1,t-K} + \ldots + \beta_P x_{P,t-K} + \varepsilon_t \tag{6}$$

In which, y_t are the monthly stock returns, and $x_{i,t-K}$ (for each i in 1 to P) is the (K-lagged) independent variable series. The coefficient estimates are calculated with the closed form solution;

$$\hat{\beta}^{OLS} = (X'X)^{-1}X'y \tag{7}$$

The other three estimation techniques are all penalized regressions in order to implement variable selection and shrinkage, while still being reasonably computationally efficient. All shrinkage methods are based on coordinate descend, in which the objective function is univariately optimized one direction at a time. The models have the advantage over OLS that they reduce the variance of the point estimate, but have the disadvantage that they introduce a bias. In forecasting, high variance introduces much uncertainty so a trade-off between bias and variance is important and relevant.

The three techniques are the Least Angle Shrinkage and Selection Operator (Lasso), Ridge regression and the Elastic Net. These add penalty terms to the OLS estimation objective, Ridge and Lasso differ in penalty term, and Elastic Net combines the penalty terms of both.

A Ridge regression is in terms of objective function very similar to OLS but it introduces a penalty term on the coefficients. It minimizes penalized sum of squares as with a penalty term for β_i estimates:

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{P} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{P} \beta_i^2 \right\} \tag{8}$$

With P the number of explanatory variables and N the number of observations. The objective function is similar to OLS if $\lambda = 0$, and $\hat{\beta}^{ridge} = \hat{\beta}^{OLS}$ in which case the coefficients are unrestricted. The penalty does not apply to the intercept of the regression model because doing so would make the results dependent on the choice of origin. The input variables are standardized and centered to make the model invariant to the scale of the variables. The model shrinks the coefficient estimates of the variables with lowest variance to zero, it assumes variables with the highest variance have the highest relevance in the model. This model has a closed form solution (the variable matrix X is assumed to be standardized, this solution follows easily if the objective formula is written in matrix form);

$$\beta^{ridge} = (X'X + \lambda I_p)^{-1} X'y \tag{9}$$

Ridge regression adds a constant to the matrix $X'X$ before inversion, which makes it nonsingular even if $X'X$ is not full rank. This property was the most important reason for introducing ridge regressions in the statistical literature.

Least Angle Shrinkage and Selection Operator (Lasso), fits coefficients on the independent variables that are the most correlated with the dependent variable. The model starts with all coefficient estimates b_i equal to zero and in each iteration in the algorithm adds the independent variable that is most correlated to the dependent variable to the model. In comparison to for example forward stagewise regressions the Lasso model moves the coefficient estimate for a certain variable towards its least squares estimate until the correlation with the residuals is equal to the next variable outside the active set its correlation with the residual. Then, the algorithm adds this variable to the active set and the algorithm continues. In the algorithm, the coefficients that are zero are dropped from the active set and a new joint least squares estimate is made on the new active set. It minimizes;

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_i| \right\} \tag{10}$$

The objective function shows much resemblance with the ridge model, but it has a differently shaped penalty area. The difference is best visualized with an illustration of a two-variable case. An illustration of the two models is in Figure 1, the red ellipticals around the OLS coefficient estimate $\hat{\beta}$ gives the residual sum of squares and the blue areas define the constraint of both penalty functions. The left figure shows a visualisation of the Lasso model, that is indifferentiable in the corners there the coefficient is set zero.

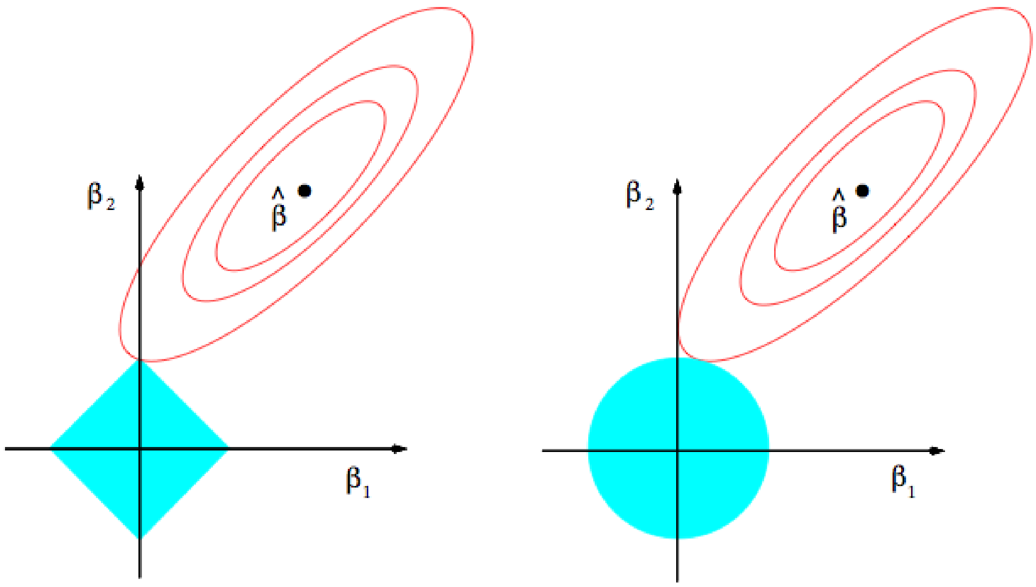

Figure 1: *The models with their constraints visualized for a two-variable case, the red ellipticals give the residual sum of squares function around the OLS estimate $\hat{\beta}$. The blue area is the model specific constraint, which is defined by $|\beta_1| + |\beta_2| \leq t$ for Lasso (on the left) and $\beta_1^2 + \beta_2^2 \leq t^2$ for ridge (on the right). Source: Google.*

The elastic net procedure combines the errors of the lasso and ridge regressions into the following specification;

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} (\alpha|\beta_i| + (1-\alpha)\beta_i^2) \right\} \tag{11}$$

The advantage is that it selects variables as in Lasso, and shrinks the coefficients as in ridge. While varieties of L_q ($\sum_{j=1}^{p} |\beta_i|^q$) norms (penalties) could result in the same penalty area, this technique will not be able to set coefficients at exactly zero (as is possible when using a Lasso penalty).

The shrinkage intensity for all three shrinkage methods is determined by using 10 fold cross validation. This means that a leave one out estimator is calculated 10 times, each run leaving 10% of the data out. The shrinkage intensity is then set as the value that minimizes the estimation error.

3.2.2 Models

The four estimation techniques will be used with a selection of independent variables. The base case is naturally the selection of 9 anomalies as described in Chapter 2. In addition to the anomalies, an autoregressive model of order 12 (AR12) is added. This model is fairly large but shrinkage is implemented by the penalized regression models. Nardi and Rinaldo (2011) implement a lasso shrinkage on a large autogregressive model with good results. The OLS base case will probably perform worse than the other three models, while no selection is based on other statistics (e.g. significance of coefficients), for the AR model it should give reasonable forecasts as it is often regarded as a base-line model and if the return series has autoregressive properties this model will maintain the time series structure of the data.

The anomalies are substituted by the principal components that explain 95% of the variance (from Principal Component Analysis (PCA) on the anomalies) as this is a straight forward alternative to variable selection. PCA converts a set of correlated observations or variables to an orthogonal set where the first principal component explains most variance, the second second most etc. This method finds the factors in the variables that explain most of the common variance and it is expected to improve forecasts by using the most important and disregarding the least important principal components. The idea is similar to ridge as factors with high variance are expected to have higher explanatory power so should get (more) weight. For more information about PCA, the reader is referred Jolliffe (2002). The PCA specification is set up to incorporate a varying number of principal components, dependent on the proportion of variance explained.

The first model is based on the lagged anomaly values, as described earlier. The variables R1 and R212 are not lagged since these are lagged in construction:

$$\begin{aligned} R_t = \alpha &+ \beta_1 AG_{t-1} + \beta_2 BM_{t-1} + \beta_3 ILLIQ_{t-1} + \beta_4 ISSUE_{t-1} + \beta_5 IVOL_{t-1} \\ &+ \beta_6 R1_t + \beta_7 R212_t + \beta_8 SIZE.USD.log_{t-1} + \beta_9 TURN_{t-1} + \varepsilon_t \end{aligned} \tag{12}$$

The second model is an autoregressive model:

$$R_t = \phi_0 + \phi_1 R_{t-1} + \ldots + \phi_{12} R_{t-12} + \varepsilon_t \tag{13}$$

The third model is a combination of the first and the second, it includes the lagged anomalies and autoregressive lags. Note how the first AR lag is not included since this is equal to the variable R1.

$$\begin{aligned} R_t = \alpha &+ \beta_1 AG_{t-1} + \beta_2 BM_{t-1} + \beta_3 ILLIQ_{t-1} + \beta_4 ISSUE_{t-1} + \beta_5 IVOL_{t-1} \\ &+ \beta_6 R1_t + \beta_7 R212_t + \beta_8 SIZE.USD.log_{t-1} + \beta_9 TURN_{t-1} \\ &+ \phi_1 R_{t-2} + \ldots + \phi_{12} R_{t-12} + \varepsilon_t \end{aligned} \tag{14}$$

The fourth model is based on PCA and includes K PCs:

$$R_t = \alpha + \beta_1 PC^1_{t-1} + \ldots + \beta_K PC^K_{t-1} + \varepsilon_t \tag{15}$$

The fifth and last model is based on PCA and autoregressive lags:

$$R_t = \alpha + \beta_1 PC^1_{t-1} + \ldots + \beta_K PC^K_{t-1} + \phi_1 R_{t-2} + \ldots + \phi_{12} R_{t-12} + \varepsilon_t \tag{16}$$

3.2.3 Statistics

The forecast accuracy is assessed with three statistics; the Mean Squared Prediction Error (MSPE), the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE).

$$MSPE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y_i})^2 \tag{17}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (= \sqrt{MSPE}) \tag{19}$$

All three statistics quantify the accuracy of the model, a lower value means a better point forecast. All statistics give a slightly different interpretation to the accuracy so they are stated in this combination. The results are shown as average cross sectional statistics which makes it in a sense additive to the analysis form Section 3.1.1. The purpose from this analysis is not to assess the profitability of the forecasts, but to assess the accuracy of the various specifications in forecasting returns. The profitability of the forecasts is researched in portfolio context.

To assess the forecasts on their directional accuracy, the statistic of Pesaran and Timmermann (1992) (PT) statistic is calculated. This statistic measures the proportion that the forecast x_t has the same direction (sign) as the predicted variable y_t:

$$S_n = \frac{\hat{P} - \hat{P}^*}{\sqrt{\hat{V}(\hat{P}) - \hat{V}(\hat{P}^*)}}, \tag{20}$$

with:

$$\hat{P} = \frac{1}{n}\Sigma_{t=1}^{n} I(y_t x_t)$$

$$\hat{P}^* = \hat{P}_x\hat{P}_y + (1 - \hat{P}_x)(1 - \hat{P}_y)$$

$$\hat{V}(\hat{P}) = \frac{1}{n}\hat{P}^*(1 - \hat{P}^*)$$

$$\hat{V}(\hat{P}^*) = \frac{1}{n}(2\hat{P}_x - 1)^2\hat{P}_y(1 - \hat{P}_y) + \frac{1}{n}(2\hat{P}_y - 1)^2\hat{P}_x(1 - \hat{P}_x) + \frac{4}{n^2}\hat{P}_x\hat{P}_y(1 - \hat{P}_x)(1 - \hat{P}_y),$$

and $I(z)$ is an indicator for $z > 0$. The statistics $\hat{P}_x$ and $\hat{P}_y$ denote the proportions that respectively x and y are larger than 0 (over n observations). All these formulae relate closely to the distribution of binominal variables. S_n is standard normally distributed under the null-hypothesis of no predictability.

The Diebold and Mariano (1995) statistic compares two forecasts in their forecast accuracy. The statistic quantifies the difference in forecast accuracy. It is based on the two series of (here, squared) forecast errors $g(e_{it}) = e_{ij}^2$ of the respective models which form the loss differential $d_t = g(e_{1t}) - g(e_{2t})$. The null-hypothesis of equal forecast accuracy ($E(d_t) = 0$) is tested.

$$DM = \frac{\overline{d}}{\sqrt{[\hat{\gamma}_d(0) + 2\Sigma_{k=1}^{h-1}\hat{\gamma}_d(k)]/T}} \tag{21}$$

Where $\hat{\gamma}_d(i)$ is the covariance $cov(d_t, d_{t-i})$ for each i in 0 to $h-1$ lags. The statistic is asymptotically standard normal under the null hypothesis of equal predictability. The DM statistic often does not reject the null hypothesis in small samples.

3.3 Portfolio formation

This section will explain the portfolio formation procedure. Portfolios are formed based on optimization, value weighing and 1/N.

The trade-off between economic and statistical considerations results in a moving window of 60 observations. And no missing values in this window are allowed. This slightly lowers the sample size in the period between 1994 and 2000. Including missing values or interpolating returns would bias the sample so this is not desirable. Furthermore, the optimization procedure will not work with missing values in the data.

3.3.1 Non-optimized portfolios

In this chapter six models are compared, two varieties of the value weighed market index (MSCI EM), an equal weight portfolio and three mean-variance optimal portfolios based on diversification, minimum variance and Sharpe ratio.

This section will introduce the three non-optimized portfolios; the value weighed indices and the equal weight portfolio. The first value weighed index is the MSCI Emerging Markets index, calculated as the weighed average of the returns its constituents. The second value weighed index takes into account the 60 month return requirement and no missing values in the sample. This results in a slightly smaller sample than in the usual value weighed index. The third method is the equal weighed method, based on the filtered stock sample. The last two methods are included because the optimized portfolios will be based on the sample sample of stocks so it makes them comparable.

In addition to these three portfolios, three optimized portfolios are calculated which will be introduced in the following section.

3.3.2 Optimized portfolios

A mean-variance efficient portfolio maximizes the expected return for a risky asset given level of volatility. The most widely known portfolio is the minimum variance portfolio, in which the investor constructs a portfolio that has the lowest possible variance. In that case, the investor optimizes the weight in order to minimize the portfolio volatility:

$$\sigma^{2*} = \min_{w} w'\hat{\Sigma}w \tag{22}$$

Where $w = [w_1..w_N]$ is the vector of portfolio weights for assets $i = 1,..,N$ and $\hat{\Sigma}$ is an estimate of the covariance matrix. The sample covariance matrix is a noisy estimate of the real covariance matrix if the number of assets is greater than the number of observations ($N > T$) (Ledoit and Wolf, 2003a), to mitigate this problem covariance matrix shrinkage is applied. The details of this solution are covered in Section 3.3.3.

A second optimization objective is the diversification ratio, as defined by Choueifaty and Coignard (2008). The authors define the ratio by:

$$DR^* = \max_{w} \frac{w'\hat{v}}{\sqrt{w'\hat{\Sigma}w}} \tag{23}$$

In this formula, $\hat{v} = [\sigma_1 \quad ... \quad \sigma_N]'$ denotes the vector of volatilities for asset $i = 1,..,N$ (which is equivalent to the square root of the diagonal entries of the covariance matrix). This ratio divides the volatility of the portfolio assuming zero correlations between stocks by the diversified portfolio volatility (with consideration of correlations). Therefor it measures the amount of diversification, given that nonzero correlations between assets will lower the portfolio variance.

The third optimization is based on the Sharpe ratio. Maximum return is often a wish of an investor but high returns often go hand in hand with high volatility. In this case, the investor might not be compensated in the sense of returns based on the risk of the portfolio. The same logic applies reversely, many investors wish to have very low variances but in that case the return might be too low based on the risk-return profile. The Sharpe ratio solves this problem measuring the compensation for risk with regard to a benchmark.

$$SR^* = \max_w \frac{w'E(r) - r_B}{\sqrt{w'\hat{\Sigma}w}} = \max_w \frac{E(r_P) - r_B}{\sigma_P} \tag{24}$$

The last part of this objective is the classic Sharpe ratio, and in the remainder of this paper the risk free rate r_f will be used as the benchmark r_B. The expected portfolio return ($E(r_P) = w'E(r)$, with $E(r)$ the $(1 \times N)$ vector of expected asset returns) and the portfolio volatility (σ_P) are based on the efficient frontier.

3.3.3 Covariance matrix

It is obvious from the previous formulas that the covariance matrix has to be estimated to optimize the portfolios. Many authors have found that the sample covariance matrix is a noisy estimate of the real covariance matrix (among other authors Ledoit and Wolf, 2003a and Pafka and Kondor, 2002). In a setting where the number of assets is large relative to the number of observations, the sample covariance matrix gets increasingly noisy due to the small number of observations that are available to calculate the covariates. When the number of assets gets larger than the number of observations ($N > T$) the sample covariance matrix has rank smaller than N so it is not invertible. In calculating the mean variance efficient portfolio this is a huge problem. In addition, Ledoit and Wolf (2003a) have found that implementing alternative methods to estimate the covariance matrix yields better results. The methods include but are not limited to factor model covariance estimation, Asymptotic Principal Component Analysis and various covariance matrix shrinkage procedures such as constant correlation shrinkage.

While there are many methods to estimate the covariance matrix that overcome the problems that were described earlier, the choice is made to only use the procedure of Ledoit and Wolf (2003b). The implementation of an estimator for the covariance matrix other than the sample covariance matrix is important but not one of the main goals of this paper. Other research might explore this topic further and compare various models, specifications and performance.

In the following portfolio optimizations, the constant covariance shrinkage methodology of Ledoit and Wolf (2003b) is used. In the shrinkage target, this procedure imposes constant correlations across all assets and keeps the variances equal to the sample variance per asset. The shrinkage target F looks like:

$$F = \begin{pmatrix} s_{11} & f_{12} & \dots & f_{1N} \\ f_{21} & s_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ f_{N1} & \dots & \dots & s_{NN} \end{pmatrix}$$

With s_{ii} the i^{th} entry of the diagonal of the sample covariance matrix S, and $f_{ij} = \overline{\rho}\sqrt{s_{ii}s_{jj}}$ with $\overline{\rho}$ the average sample correlation.

The method finds the optimal $\hat{\delta}^*$, such that the Frobenius norm of the shrinkage estimator ($\hat{\Sigma}_{LW}$) and the true covariance matrix is minimal.

$$L(\delta) = ||\delta F + (1-\delta)S - \Sigma||^2 \tag{25}$$

While the true covariance matrix is unobserved and unknown, the authors derive a consistent estimator for the optimal shrinkage constant $\hat{\delta}^*$ such that;

$$\hat{\Sigma}_{LW} = \hat{\delta}^* F + (1-\hat{\delta}^*)S \tag{26}$$

Ledoit and Wolf (2003b) prove that the optimal value $\hat{\delta}^*$ behaves like a constant over T for large T. Moreover the authors show that the optimality still holds in finite samples. The constant is formulated as:

$$\kappa = \frac{\pi - \rho}{\gamma} \tag{27}$$

In formula 27, π is the sum of the asymptotic variances of the elements of the sample covariance matrix scaled by $\sqrt{T}$. Then, ρ denotes the sum of asymptotic covariances of the elements of the shrinkage target with the elements of the covariance matrix scaled by $\sqrt{T}$. Last, γ measures the misspecification of the shrinkage target. The authors derive consistent estimators for each of the parts from κ: $\hat{\pi}$, $\hat{\rho}$, and $\hat{\gamma}$.

The definition for π is (where AsyVar stands for asymptotic variance):

$$\pi = \Sigma_{i=1}^{N}\Sigma_{j=1}^{N} AsyVar(\sqrt{T}s_{ij}) \tag{28}$$

And a consistent estimator is:

$$\hat{\pi} = \frac{1}{T}\Sigma_{t=1}^{T}[(y_{it} - \overline{y}_i)(y_{jt} - \overline{y}_j) - s_{ij}]^2 \tag{29}$$

3.3.4 Expected returns

The mean-variance efficient portfolio is based on the covariance matrix and the expected returns. To compare models, two specifications will be examined. In the first specification, the average (historical) returns are used as a proxy for expected returns and in the second specification a selection of forecasts from the previous chapter is used. In that second specification the models from the last chapter can be assessed on their ability to proxy expected returns, although the shrinkage models are expected to be biased. The models that show the most directional accuracy in the sense of Pesaran and Timmermann (1992) are expected to provide the best portfolio returns. These models can determine best which stocks will have negative or positive results, a trait that improves the portfolio optimization.

4 Results

This chapter will present the results for the analysis. As introduced in the previous chapter the following will start with the results for the hedge portfolios and Fama and MacBeth (1973) (FMB) style analysis in Section 4.1, then the results for the stock forecasts in Section 4.2 and finally the results for the portfolio procedure in Section 4.3.

4.1 Hedge portfolios and Fama MacBeth regressions

Only stocks are included that are traded for at least 12 months preceding the 'current' month. This filter is included to assure that the results are not driven by changes in listing. The formation of portfolios is based on two-period lagged variables and the return for the 'next' period is calculated in the hedge portfolios.

4.1.1 Hedge portfolios and Fama-MacBeth coefficients

The returns for the hedge portfolios and the FMB style analysis are stated in Table 2, and the results for the portfolios with regard to trading costs can be found in Section A.1 (Table 10). Eight out of nine anomaly hedge portfolios show (highly) significant returns. These returns are analyzed gross, and with regard to trading costs which are assumed to be 1%[3]. After adjusting returns for trading costs, some hedge portfolio returns are still highly significant. This is probably due to the large number of observations (249) which makes the denominator in the calculation of the t-statistic small and results in higher t-statistics. The coefficients on the FMB analysis show highly significant results for all anomalies. The following paragraphs will go into detail on the results for the hedge portfolios and FMB regressions.

Unreported analysis showed that the hedge portfolio returns are extremely high if the returns are only winsorized at the 0.005 and 0.995 percentile. Therefor, in each month returns are truncated on the 1% and 99% percentile, following Barry et al. (2002). This results in less extreme returns and t values for asset growth and book-to-market. In addition, more hedge

[3] Assuming trading costs are 1%, and that these can be subtracted from the gross returns net returns can be approximated by: $r_{net} = r_{gross} - (1\% \times turnover)$.

portfolios show significant results due the deletion of extreme values.

The average hedge portfolio return for value (book-to-market) is highly significant with 2.1% (t=9.2) and still highly significant after regarding trading costs 1.9% (t=8.5). The FMB analysis shows that the value premium is also highly significant and positive in the cross section. These results confirm the findings of for example Fama and French (1993), Chordia et al. (2014) and Brennan et al. (1998) among many others, these authors have all found significant value premia in developed markets. The size of the returns is roughly in line with Barry et al. (2002) in which the authors find a highly significant value effect in emerging markets of 2.9% with extreme returns, and 1.5% without extreme returns. Unreported analysis shows that calculating the returns of this anomaly portfolio in smaller subsamples leads to less extreme t-statistics. The size of the returns is driven by the significant results for both the long and the short leg which are significantly positive and negative respectively (see Section A.1).

The observation that both analyses show significant results in emerging market stocks assures that the value anomaly is still in the market. Stocks with higher book-to-market value tend to perform better than low book-to-market stocks, so the stocks that have high market prices and little assets (low book-to-market value) tend to drift away from the inherent value tend to perform worse. Emerging markets have more uncertainty and volatility so investors prefer value stocks which are showed to be attractively priced.

The size effect is insignificant in the hedge portfolio analysis. The FMB analysis shows that size has positive significant explanatory power for returns in the cross-section. This confirms the results from Barry et al. (2002) where the authors research value and size effect in emerging markets, and find no significant results if extreme observations are removed.

The size effect is commonly explained in the literature as an investment in small stocks that should get higher returns due to the risk return trade-off (small stocks are argued to be more risky). The results from the FMB analysis shows a risk premium that is positive and significant, which confirms the cross-sectional findings of Chordia et al. (2014), Fama and French (1993) and Barry et al. (2002) (among many other authors). In combination with the last paragraph, this paragraph finishes the analysis of the Fama and French (1993) anomalies which are confirmed in various levels of conclusiveness.

	Panel A					Panel B			
	r (%)	SD (%)	TO (%)	t-stat	Sign.	Coef	SE	t-stat	Sign.
Intercept						1.33	0.06	20.80	***
AG	2.09	2.73	14.9	11.79	***	0.55	0.04	15.50	***
BM	2.06	3.55	15.1	9.19	***	-0.56	0.05	-11.30	***
ILLIQ	-0.91	6.38	50.9	-2.23	**	0.11	0.05	2.22	**
ISSUE	-0.53	2.32	15.0	-3.49	***	-0.10	0.02	-4.18	***
IVOL	-0.88	5.10	58.6	-2.72	***	0.85	0.13	6.47	***
R1	0.94	4.93	84.3	3.01	***	-0.71	0.09	-8.26	***
R212	1.18	5.93	30.7	3.06	***	-0.33	0.08	-4.04	***
SIZE	0.39	4.86	15.4	1.27		0.57	0.10	5.45	***
TURN	0.71	5.76	37.7	1.96	*	1.53	0.07	22.11	***

Table 2: *Panel A: This table gives the monthly average raw returns (r), standard deviation (SD) and turnover (TO) of the long-short decile hedge portfolios. Significance for a two sided test with the null hypothesis that individual returns are zero at the 10%, 5% and 1% level is indicated by *, **, and *** respectively. To calculate the t-statistic,* $T = 249$ *is used. Panel B: This table gives the FMB coefficients and their standard errors (SE), multiplied by 100 for readability. The regressions follow the empirical methodology of Brennan, Chordia and Subrahmanyam (1998). Significance at the 10%, 5% and 1% level is indicated by *, **, and *** respectively.*

The hedge portfolio return for idiosyncratic volatility is significant with a mean of (-)0.88%. This result is driven by the observation that the high idiosyncratic volatility part (the long part) of the hedge portfolio gets a (insignificant) negative return and the short part a positive significant return, these results are in Section A.1. The portfolio turnover is 59% which diminishes the significance, both statistically and economically (the mean would be -0.30% with $t = -0.92$). The FMB coefficient that implies that a one standard deviation increase of idiosyncratic volatility increases the return by 0.85%.

The hedge portfolio results confirm Ang et al. (2006) and Ang, Hodrick, Xing and Zhang (2009) who found that stocks with higher idiosyncratic volatility earn lower returns. These results are furthermore in line with the results from for example Blitz, Pang and Van Vliet

(2013) that researched stocks with low volatility and found that the relation between risk and return is declining or flat, and Baker and Haugen (2012) and Blitz and Van Vliet (2007) that research low volatility stocks in developed markets which yield similar results. Recent studies have showed that these results might not be as strong as perceived earlier for example Li, Sullivan and Garcia-Feijóo (2014) show that an equally weighed long short portfolio on volatility does not get anomalous returns.

The positive FMB coefficient implies that higher idiosyncratic volatility is related to higher returns, which would fit in the efficient market hypothesis but is in contradiction to the literature. These results might be driven by investors that expect emerging markets (EM) to be increasingly risky, so in addition to demanding a premium for holding any emerging market EM stock the investor could demand a premium for risky EM stocks. In addition, emerging markets are showed to be more volatile than developed markets and these results could imply that stocks with high idiosyncratic volatility in the long run do show positive returns and the risk is compensated. But it also raises the question if idiosyncratic volatility is measured correctly and whether the relation between volatility and returns is linear.

Momentum hedge portfolio shows significant returns of 1.2% (t=3.0), and these returns as robust to trading costs: 0.9% (t=2.3). These results are in line with the literature, for example Jegadeesh (1990) and many other authors. Momentum is unexpectedly negative and significant in the Fama and MacBeth (1973) analysis, which means that a one standard deviation increase in the momentum returns leads to a monthly extra return of -0.33%. These results show that more momentum cross-sectionally leads to smaller returns, a form of reversal. And it is possible to capitalize on the winners by forming hedge portfolios.

The hedge portfolio return on asset growth is significant with 2.1% and the turnover of the portfolio is only 14.9% which translates to net returns of 1.9% (t=11.1), but its sign is different than expected. Cooper et al. (2008) show and argue that stock with high asset growth will have low returns, and the same holds vice versa. The positive sign on the estimated hedge portfolio is therefor unexpected, and this result is driven by small but significant returns on the low asset growth portfolio (see Section A.1) and the negative significant returns on the high asset growth portfolio. The FMB coefficient is highly significant and shows that an increase asset growth of one standard deviation leads to an increase of 0.55% of returns.

These results could arise from inherently different characteristics of emerging and developed markets. In developed markets, high asset growth leads to lower returns and Cooper et al. (2008) argue that this result is driven by overreaction of investors to past performance. In that case investments in asset base decrease the profit for the share holders which lowers the returns. The results in in emerging markets in this paper show that stocks with high asset growth earn higher returns. This indicates that investors see the additional investments made by asset growth will lead to higher growth and increasing prices.

The issuance anomaly, as reported by Pontiff and Woodgate (2008) is significant in this data set for the hedge portfolio returns and earns (-)0.5% per month (t=3.5) and the returns are robust to trading costs: -0.4% (t=-2.5). The FMB coefficient is also significant. The sign and impact of the hedge portfolio return is comparable to Chordia et al. (2014), which makes the cross sectional significance is not surprising while Pontiff and Woodgate (2008) finds fairly strong significant results in FMB cross sectional regressions in most periods. The effect confirms the usual interpretation of the anomaly, firms are reckoned to use the ability to repurchase and issue shares in order to react opportunistically to changes in expected returns. Section 4.1.2 will show if and how this effect has changed during the sample period.

Liquidity has significant results in the hedge portfolio with an average return of (-)0.9% (t=2.2). This portfolio shows one of the highest turnover in comparison to the other portfolios and the portfolio does not show significant returns after trading costs: -0.4% (t=-1.0). The FMB analysis shows a positive significant coefficient which means that more illiquidity leads to higher returns. Therefor a liquidity premium is present in this sample. The results from the hedge portfolios in turn show that the most liquid stocks (the short leg of the hedge portfolio, the lowest illiquidity stocks) earn positive significant returns. These results fit well in the results from Chuhan (1994) that shows that low liquidity is a major concern for institutional investors in the choice of investing in emerging markets. In that light, more liquid stocks are more in favor of institutional investors which drives demand and has a positive influence on returns.

The coefficients on short term reversal (R1) are significant for the hedge portfolio, but not robust to trading costs (then the average return is 0.10% (t=0.3)). The explanation of the anomaly is given by the overreaction of investors (high R1) which is followed by lower returns. The FMB coefficient is negative and significant which confirms the results from Chordia et al. (2014) and the existing literature on this topic. Jegadeesh (1990) has stated that stock returns do not follow a random walk and this analysis confirms that finding.

Turnover is significant in its hedge portfolio with an average return of 0.71% (t=2.0) which is not robust to trading costs 0.34% (t=0.93). This (non-)result is similar to the findings of Bekaert, Harvey and Lundblad (2007) who test this measure against other stock specific characteristics. The characteristic premium from the FMB analysis is positive and significant, which is unexpected from Chordia et al. (2014). The turnover variable is introduced as a measure for liquidity in response to the Amihud and Mendelson (1986) paper. The estimates were expected to have the opposite sign, now the results are interpreted as a reverse liquidity premium. Higher turnover leads to higher returns. These results might be driven by herding investors in popular stock. In that case, the turnover of a stock will be very high and prices will rise, leaving the less popular stock (with lower turnover) out of their portfolios. While the variables in this analysis might have measurement error, it remains extremely difficult to distinguish between measurement error and inherently different characteristics of emerging market stock.

4.1.2 Trend in coefficients

The hedge portfolio returns and FMB coefficients are researched for an exponential trend as explained in Chapter 3 and follows the methodology of Chordia et al. (2014). The model that is tested is:

$$y_t = a\exp(bt + \varepsilon_t) \tag{30}$$

The results from this analysis are stated in Table 3. The analysis for the trend in the hedge portfolio returns shows significant change in anomaly returns for the book-to-market, illiquidity and idiosyncratic volatility portfolios. These results are in contrast with the results from Chordia et al. (2014) in which all anomaly returns are found to have attenuated in the United States.

This indicates the possibility to invest based on the significant results stated in the last section. The analysis of the trend in the characteristic premia shows that other premia have changed over time than for the hedge portfolio returns.

The hedge portfolio returns have attenuated significantly for the book-to-market portfolio, which means that the positive effect from the last paragraph has become less pronounced over the last decades. The hedge portfolios for illiquidity and volatility have accentuated significantly, which means that the return on stock with those characteristics has rose. That indicates that investors have been rewarded for taking on risk more in the last years.

The asset growth premium shows a significant attenuation, while it is still positive and significant (as seen in the last section). This means that the anomaly premium has become weaker over the last decades, with a rate of -0.15% per month. Stocks with high asset growth value have earned higher premium a decade ago than now. With regard to increased market liquidity and transparency in (emerging) markets this result is logical while anomalous returns should be impossible.

Analysis of the trend in the characteristic premium for ILLIQ shows that it accentuated in the sample period. This indicates that the premium for owning illiquid stocks has become higher, which contradicts the assumption of increased market efficiency. This observation is best explained in the light of increased market wide liberalization in emerging markets (Bekaert et al. (2007)), while most of the stocks get increasingly liquid an investor demands a higher premium for an illiquid stock. The negative characteristic premium from the analysis in the last section is thus accompanied by an increasing premium which could indicate that more of an illiquidity premium is present in emerging markets in the end of the sample period than at the beginning.

Momentum premia show a positive significant trend which does not fit in the results from Chordia et al. (2014) and general asset pricing theory. Since the discovery of the anomaly in Jegadeesh and Titman (1993) market participants have had time to capitalize on this anomaly. As it grows older the premium is expected to become less pronounced from an efficient market point of view. From an behavioural finance point of view the anomaly can exist as one could argue that it is self-containing as investors are irrational.

The characteristic premium on size shows a negative and significant trend. This implies that the characteristic premium has declined in the last decades. While the sign of the coefficient in Table 2 is not in line with the literature, this result shows that it is moving towards the expected sign. The significant trend could indicate that the anomaly is self enforcing as market participants have expected it to be present in (small) emerging markets and consequently buying these stocks so driving up the price (and returns).

	Panel A			Panel B		
	b	t-stat	Sign.	b	t-stat	Sign.
AG	-3.84	-1.29		-1.48	-2.45	**
BM	-6.43	-1.68	*	0.23	0.30	
ILLIQ	15.81	2.19	**	0.26	0.41	
ISSUE	2.01	0.77		-0.22	-0.58	
IVOL	22.20	3.98	***	0.92	1.17	
R1	-0.63	-0.12		0.24	0.17	
R212	-3.92	-0.58		6.12	4.54	***
SIZE	-7.50	-1.38		-4.24	-2.81	***
TURN	-2.35	-0.36		1.64	1.28	

Table 3: This table states the coefficients for the exponential trend model ($\times 1000$). Panel A: Trend in hedge portfolio returns. Panel B: Trend in FMB coefficients ($\times 1000$.

Analysis of the hedge portfolios showed that various anomalies are present in emerging market stocks and significant returns can be obtained if such strategies are implemented. The analysis of the trend in the hedge portfolio returns shows that there has been little significant attenuation or accentuation of these anomalies so even now these returns are capitalizable. All FMB coefficients show significant results. Various characteristic premiums have showed to have undergone a change in the last decades. With these conclusions, the large sample research questions are answered and the following section will state the results on the forecast models in the MSCI EM sample.

4.2 Forecasts

4.2.1 Accuracy

The results for the five models each with four estimation techniques are stated in Table 4. These models all use in sample $(1 \dots T)$ observations to estimate the coefficients to forecast the out of sample $(T+1)$ return:

$$\hat{y}_{T+1} = \alpha + X_T \hat{\beta}_T, \tag{31}$$

with X_T the vector of variables for each model, and $\hat{\beta}$ the vector of coefficient estimates for each method (estimated over the in sample period). Panel A and B show that the models that include the anomalies in the forecasts have in general high statistics (so low accuracy) in comparison to the other three models. The model that includes only the anomalies shows very high statistics (compared to the other specifications) for all four models which shows that there is much error in the forecasts. The addition of 12 autoregressive lags improves the forecasts in the sense of the regarded statistics. The conclusion can be drawn that including the anomalies in forecasting stock returns does not yield accurate forecasts, while there is much error and the lagged values do not directly forecast returns efficiently. The shrinkage models in the anomaly plus autoregressive specification do not improve much overall over the anomaly specification, compared to the models in Panels C to E.

The autoregressive specification (Panel C) improves much over the models in Panels A and B. The forecast errors are lowest for all models, even when the models in Panel D and E are analyzed. OLS has highest errors compared to the shrinkage methods but as a baseline model, the autoregressive specification is hard to beat.

PCA is used as a method to summarize the information in the anomaly variables, and select the most important PC's accordingly. The PCA model in Panel D estimated with OLS results in higher statistics than for the AR model (in Panel C), but still much lower than for the models in panel A and B. The shrinkage methods show that for the PCA model, the errors lie very close to the AR model and are much improved over the anomaly models in Panel A and B. The PCA model is based on the anomalies in the model in Panel A, but the forecasts are much more accurate. Even the ridge model in Panel A, which follows an approach close to PCA due to the singular value decomposition of the explanatory variables performs much worse than the OLS model in the PCA model. The PCA model does show overall slightly less accurate forecasts than the AR model.

While the inclusion of autoregressive terms as in Panel E is expected to add to the accuracy of the model because both the PCA model and autoregressive model perform well, the accuracy is not improved much for the shrinkage methods. The OLS model does not perform well and does not give a good baseline model, but this is intuitively correct because adding twelve more variables in OLS increases the chance of overfitting and the in-sample fit might be high but the out of sample accuracy is disregarded. The shrinkage methods rely heavily on the correlation of the independent variables, and PCA will probably have higher correlation with returns than the AR lags. Therefor it might have happened that the PC's get higher coefficients than the AR lags.

The shrinkage methods add to the accuracy of the forecasts for all models, with large differences in the improvement for these techniques. So using continuous variable selection in forecasting stock returns does improve over OLS. Ridge shows that for each specification it gives the most accurate forecasts, and OLS in generally the least accurate.

Although even the relatively most accurate forecasts (AR12 and PCA) show much improvement over the other three specifications, the errors are still too large to use effectively in forecasting. The lowest MAE is 8.3% which makes it improbable that a forecast will be close to the realized value or even have the same sign. To formalize this statement, the forecasts are examined on their directional accuracy. Moreover, the models will be compared mutually to assess which models are best in the sense of accuracy. To explore the question whether the forecasts have value in aggregate, an analysis in portfolio context will be conducted in Section 4.3.

		Model	MSPE	MAE	RMSE
A	9 Anomalies	OLS	17886.522	1.314	9.655
		LASSO	1041.300	0.221	1.571
		Ridge	555.067	0.264	1.622
		ElNet	662.733	0.197	1.295
B	9 Anomalies + AR12	OLS	12369.311	1.139	7.569
		LASSO	583.981	0.176	1.078
		Ridge	128.240	0.172	0.834
		ElNet	402.604	0.162	0.919
C	AR12	OLS	0.020	0.101	0.133
		LASSO	0.016	0.089	0.119
		Ridge	0.016	0.089	0.118
		ElNet	0.016	0.089	0.119
D	PCA	OLS	0.525	0.119	0.246
		LASSO	0.044	0.091	0.132
		Ridge	0.015	0.088	0.116
		ElNet	0.033	0.090	0.126
E	PCA + AR12	OLS	12.642	0.169	0.543
		LASSO	0.015	0.087	0.116
		Ridge	0.015	0.087	0.115
		ElNet	0.015	0.087	0.116

Table 4: This table summarizes the accuracy of the forecast models. Each panel corresponds to a model specification, and in these panels four models are estimated on the variables. The reported measures are the Mean Squared Prediction Error (MSPE), Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). Lower values imply higher accuracy, higher values imply lower accuracy.

4.2.2 Directional accuracy

The directional accuracy of the models is assessed by the Pesaran and Timmermann (1992) statistic, which are stated in Table 5. That statistic is usually regarded with an one-sided test, where the null hypothesis is that of better accuracy than a random guess based on the sample proportion of the occurrence of positive and negative returns. As the statistics are not at all positive, the choice is made to state the models with a two-sided test. The significance of the

PT statistic is not calculated for the average of all models (Panel A), as it only summarizes the remainder of the table.

Table 5 shows that only the 9 anomaly, 9 anomaly with AR, and the PCA models (Panels A, B, D) estimated with OLS have insignificant PT statistics. All other models show (highly) significant negative PT statistics, which means that the models are significantly worse in forecasting the sign of future returns than the sample average. This shows that the models cannot forecast the sign well and the significance is in a sense even special as the models are significantly worse than a guess.

			% p<0.05	% p<0.1	PT	Sign.
A	Average		4.78	9.25		
B	9 Anomalies	OLS	6.16	12.30	-0.08	
		LASSO	6.27	10.70	-2.15	**
		Ridge	5.45	10.00	-3.50	***
		ElNet	5.42	10.30	-2.05	**
C	9 Anomalies + AR12	OLS	5.50	10.20	-1.40	
		LASSO	5.22	9.16	-3.73	***
		Ridge	3.66	7.45	-5.36	***
		ElNet	4.25	7.72	-3.58	***
D	AR12	OLS	4.41	8.36	-3.25	***
		LASSO	3.32	8.35	-4.16	***
		Ridge	3.85	7.82	-5.34	***
		ElNet	4.70	8.16	-4.64	***
E	PCA	OLS	5.62	10.30	-0.48	
		LASSO	4.23	10.60	-2.09	**
		Ridge	5.17	7.82	-3.83	***
		ElNet	4.02	9.03	-2.81	***
F	PCA + AR12	OLS	4.35	9.57	-2.27	**
		LASSO	4.53	9.07	-3.14	***
		Ridge	5.01	9.50	-2.74	***
		ElNet	4.46	8.53	-4.58	***

Table 5: This table states the result for the directional accuracy of the forecast models. Panel A reports the average over all statistics from Panel B to F, and each Panel B to F reports the directional accuracy for the forecast models in all specifications. The first two columns state the proportion of significant results in the individual tests, and the third and most important column states the joint PT statistic for all forecasts. Significance at the 10%, 5% and 1% level for a two sided test is indicated by *, **, and *** respectively.

4.2.3 Diebold Mariano statistics

To examine the forecasts, and draw conclusions about which model works best in forecasting returns the Diebold and Mariano (1995) statistics are calculated for all forecasts mutually. This means that for each model i, the *average* loss differential is compared against model j for all

$i, j = 1, \ldots, N$ and $i \neq j$. The models that are tested are in the columns, the models that they are tested against are in the rows. Therefor a positive significant statistic means that the model in the column is significantly more accurate than the model in the row. The models are tested one-sided, while also testing on significantly worse accuracy would clutter the table (and the table is symmetric so a positive significant result for model A against model B would imply a significant negative result for model B against model A in a two-sided test). Many coefficients that compare a certain model with much error with a model with little error lie very close to each other which is due to the large forecast error of the former.

The results for the comparison of the accuracy of the models is stated in Table 6 and Table 7. The results show that almost all models (and specifications) significantly outperform the 9 anomaly model estimated with OLS. The same holds for the model that includes the anomalies and 12 autoregressive lags. For those two models the shrinkage methods are not able to beat any of the remaining specifications based on AR12 and PCA.

The AR12 model is significantly better than the PCA and AR with PCA model in every specification. This conclusion follows from the highly significant statistics in that part of the table. The AR model overall improves over the two models based on the anomalies and the anomalies and autoregressive lags but not significantly like the PCA based models. Only the autoregressive model estimated with OLS performs somewhat worse with less convincing results than the shrinkage forecasts.

The PCA specification improves significantly over the autoregressive with PCA specification, and the autoregressive with PCA specification improves only over the some OLS estimates.

The results in the DM tables have showed that the anomaly based models do not perform well compared to the other models. If PCA is applied to the anomaly data these forecasts improve but the forecasts still struggle to beat the autoregressive forecasts. Including autoregressive lags in the PCA model does not improve over the specification that includes only PCs. Thus, the autoregressive model is the most accurate model to forecast the returns with in this spectrum of models, and including shrinkage methods leads to significant improvements over the PCA models although the improvements are not significant for the autoregressive models.

Concluding, the AR variable combined with shrinkage methods performs best in compared to the other models. These results are striking because the model that includes AR lags and PCA contains more information but lacks to perform. The AR model with LASSO shrinkage has been tested by Nardi and Rinaldo (2011) and showed positive results. The analysis in this paper confirms and extends the results from Nardi and Rinaldo (2011) while other methods shrinkage methods tend to perform well (in comparison to the models that are considered here).

	Yf.9.OLS	Yf.9.LASSO	Yf.9.Ridge	Yf.9.ELnet	Yf.9.AR.OLS	Yf.9.AR.LASSO	Yf.9.AR.Ridge	Yf.9.AR.ELnet
Yf.9.OLS		2.01 **	2.01 **	2.01 **	1.39	2.02 **	2.02 **	2.02 **
Yf.9.LASSO	-2.01		1.12	1.17	-1.81	1.36	1.26	1.32
Yf.9.Ridge	-2.01	-1.12		-1.04	-1.77	-0.60	1.54	1.23
Yf.9.ELnet	-2.01	-1.17	1.04		-1.78	1.65 *	1.31	1.50
Yf.9.AR.OLS	-1.39	1.81 *	1.77 *	1.78 *		1.79 *	1.76 *	1.78 *
Yf.9.AR.LASSO	-2.02	-1.36	0.60	-1.65	-1.79		1.16	1.21
Yf.9.AR.Ridge	-2.02	-1.26	-1.54	-1.31	-1.76	-1.16		-1.13
Yf.9.AR.ELnet	-2.02	-1.32	-1.23	-1.50	-1.78	-1.21	1.13	
Yf.AR.OLS	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.AR.LASSO	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.AR.Ridge	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.AR.ELnet	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.pca.OLS	-2.02	-1.28	-1.54	-1.34	-1.76	-1.21	-1.53	-1.22
Yf.pca.LASSO	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.pca.Ridge	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.pca.ELnet	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.54	-1.22
Yf.AR.pca.OLS	-2.01	-1.24	-1.42	-1.27	-1.75	-1.14	-1.00	-1.11
Yf.AR.pca.LASSO	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.53	-1.22
Yf.AR.pca.Ridge	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.53	-1.22
Yf.AR.pca.ELnet	-2.02	-1.28	-1.54	-1.34	-1.76	-1.22	-1.53	-1.22

Table 6: This table gives the DM statistics for all models against each other. Significance means that the model in the column is better than the model in the row, with regard to forecast error. Significance on a one-sided test is denoted by *, ** and *** on the 10%, 5% and 1% level respectively.

	Yf.AR.OLS	Yf.AR.LASSO	Yf.AR.Ridge	Yf.AR.ELnet	Yf.pca.OLS	Yf.pca.LASSO	Yf.pca.Ridge	Yf.pca.ELnet	Yf.AR.pca.OLS	Yf.AR.pca.LASSO	Yf.AR.pca.Ridge	Yf.AR.pca.ELnet
Yf.9.OLS	2.02 **	2.02 **	2.02 **	2.02 **	2.02 **	2.02 **	2.02 **	2.02 **	2.01 **	2.02 **	2.02 **	2.02 **
Yf.9.LASSO	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.24	1.28	1.28	1.28
Yf.9.Ridge	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.42	1.54	1.54	1.54
Yf.9.ELnet	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.27	1.34	1.34	1.34
Yf.9.AR.OLS	1.76 *	1.76 *	1.76 *	1.76 *	1.76 *	1.76 *	1.76 *	1.76 *	1.75 *	1.76 *	1.76 *	1.76 *
Yf.9.AR.LASSO	1.22	1.22	1.22	1.22	1.21	1.22	1.22	1.22	1.14	1.22	1.22	1.22
Yf.9.AR.Ridge	1.54	1.54	1.54	1.54	1.53	1.54	1.54	1.54	1.00	1.53	1.53	1.53
Yf.9.AR.ELnet	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.11	1.22	1.22	1.22
Yf.AR.OLS		1.16	1.19	1.15	-1.91	-1.11	-0.85	-0.91	-1.92	-1.92	-1.92	-1.90
Yf.AR.LASSO	-1.16		0.73	-0.54	-2.13	-2.99	-3.00	-3.05	-1.92	-2.60	-2.59	-2.59
Yf.AR.Ridge	-1.19	-0.73		-1.19	-2.14	-3.00	-3.01	-3.06	-1.92	-2.60	-2.59	-2.59
Yf.AR.ELnet	-1.15	0.54	1.19		-2.13	-2.95	-2.95	-3.00	-1.92	-2.58	-2.57	-2.58
Yf.pca.OLS	1.91 *	2.13 **	2.14 **	2.13 **		1.80 *	1.76 *	1.76 *	-1.88	1.04	1.02	1.05
Yf.pca.LASSO	1.11	2.99 ***	3.00 ***	2.95 ***	-1.80		0.94	0.82	-1.91	-1.59	-1.56	-1.58
Yf.pca.Ridge	0.85	3.00 ***	3.01 ***	2.95 ***	-1.76	-0.94		-0.55	-1.91	-1.83	-1.82	-1.81
Yf.pca.ELnet	0.91	3.05 ***	3.06 ***	3.00 ***	-1.76	-0.82	0.55		-1.91	-1.79	-1.77	-1.77
Yf.AR.pca.OLS	1.92 *	1.92 *	1.92 *	1.92 *	1.88 *	1.91 *	1.91 *	1.91 *		1.89 *	1.89 *	1.89 *
Yf.AR.pca.LASSO	1.92 *	2.60 ***	2.60 ***	2.58 ***	-1.04	1.59	1.83 *	1.79 *	-1.89		0.02	0.36
Yf.AR.pca.Ridge	1.92 *	2.59 ***	2.59 ***	2.57 **	-1.02	1.56	1.82 *	1.77 *	-1.89	-0.02		0.14
Yf.AR.pca.ELnet	1.90 *	2.59 ***	2.59 ***	2.58 **	-1.05	1.58	1.81 *	1.77 *	-1.89	-0.36	-0.14	

Table 7: This table gives the remainder of the results from Table 6.

4.3 Portfolios

The results for the optimized portfolios are stated in Table 8. The portfolios are optimized with a weight restriction that makes sure that at least 25% of the investable universe of that period is included in the portfolio and the expected returns come form the forecast models in the last section. The covariance matrix is estimated by the shrinkage methodology of Ledoit and Wolf (2003a). Three benchmark indices are stated: "Index", "Index**" and equal weight ($1/N$). The statistics for Index are the statistics for the MSCI Emerging Markets Index, the statistics for Index** are the same value weighed index results, but with a trading filter as in the portfolio optimization procedure (see Chapter 3), and the equal weight portfolio is the portfolio that weighs all stocks that passed the trading filter equally. In addition, three optimized benchmark portfolios are stated which use the average return over the estimation period as a proxy for future returns.

The statistics for the benchmark portfolios show that the returns, volatility and Sharpe Ratio are high. The returns on the the three benchmark portfolios are highest for all regarded portfolios. The benchmark models in addition have high volatility and fairly high SRs compared to the optimized portfolios.

The model that uses the 9 anomalies to forecast returns was one of the worst performing models from the last section. The returns on the portfolios lag in compared to the benchmark portfolios. The portfolio that optimizes diversification based on elastic net forecasts performs best with a Sharpe Ratio of 0.7. The optimization for maximum SR gets the highest return with Ridge forecasts, but it does not get the highest SR.

Forecasting returns by using 9 anomalies and 12 AR lags showed to have limited power in the last section. The model performed better than the model that included only 9 anomalies, with Ridge forecasts. These forecasts have high SR and returns, and the returns for these portfolios lie very close together in comparison to the other forecast techniques in this model. The anomaly based optimizations show again that the forecasts did not perform well in the sense of accuracy, the returns are overall much lower than for the benchmarks.

Inspecting the forecast statistics in the last section led to the conclusion that the AR model performed best in the sense of accuracy both in a relative and an absolute perspective. This variable specification has the highest SRs across all portfolios for the minimum variance and maximum diversification portfolios. The maximum SR portfolio has the lowest SRs in this variable specification due to high volatility.

The PCA forecast specification performed well in the sense of accuracy although it was not better than the AR specification. The results from the portfolio optimizations show that OLS has very low returns and SR which is expected as OLS gave inaccurate forecasts. The ridge forecast performs best but it still does not come close to the benchmark indices. It does improve the returns and SR of the maximum SR benchmark portfolio.

Using 12 AR lags and PCA in a forecast model has showed to yield forecasts that are comparable to the AR and PCA specifications separately, although it is significantly worse than both models in the Diebold-Mariano test. The SR on the minimum variance portfolios for each model except OLS is high compared to the other models and the same holds for the maximum SR portfolio. Only the returns are lagging. The OLS model gives low returns (but also low volatility).

The benchmark portfolios are very hard to beat in the sense of return and Sharpe Ratio. Only the portfolios based on average returns s a proxy for future returns and the AR12 models get double digits returns for all three portfolios. The same holds for the Sharpe Ratios which are all pretty much all consistently above 0.6. The AR PCA model (with all estimation methods but OLS) also does fairly well in the sense of return and Sharpe Ratio. But the AR model shows the best statistics for all portfolios, it earns some of the highest returns and SR.

The optimized portfolios generally have lower volatility in absolute terms, but the portfolios with the lowest volatility generally show lowest returns (for example PCA-AR OLS, PCA OLS, 9 Anomalies LASSO and Ridge). This result is not restricted to the minimum variance portfolios but often occurs in the minimum variance portfolio and the maximum diversification portfolio. The maximum SR portfolio is not often able to maximize the SR out of sample. The maximum diversification portfolio is often close to the minimum variance portfolio.

These results show that low volatility investing does not always result in anomalous returns, the situation in which lower volatility does not translate automatically to lower returns. But these results can be driven by poor forecasting quality. The AR and historical average models show that the minimum variance portfolio can lower volatility and increase SR, which makes a case for low volatility investing.

The conclusion can be drawn that including including autoregressive lags is very important in forecasting returns compared to other forecast models. Even more, the sad conclusion must be drawn that advanced forecast models do not outperform a simple historical average as a proxy for future returns. Implementing covariance matrix shrinkage, and optimizing portfolios based on returns forecasts does not improve over a benchmark index.

		Index	Index**	1/N	Min var	Max div	Max Sh
Average return	r (%)	14.21	15.15	16.19	12.49	13.59	13.40
	vol (%)	20.49	21.23	22.09	11.81	14.10	16.43
	SR	0.56	0.59	0.61	0.83	0.78	0.65
9 Anom. OLS	r (%)				11.73	10.02	11.91
	vol (%)				13.25	12.08	18.87
	SR				0.69	0.61	0.49
9 Anom.LASSO	r (%)				8.57	7.02	11.60
	vol (%)				13.29	12.12	19.01
	SR				0.45	0.36	0.47
9 Anom. Ridge	r (%)				9.42	5.37	13.44
	vol (%)				13.09	10.85	15.57
	SR				0.52	0.25	0.69
9 Anom. ElNet	r (%)				8.22	10.50	8.85
	vol (%)				13.41	11.25	18.84
	SR				0.42	0.70	0.33
9 Anom.-AR12 OLS	r (%)				5.82	3.64	9.43
	vol (%)				12.79	11.77	18.14
	SR				0.25	0.09	0.37
9 Anom.-AR12 LASSO	r (%)				8.41	5.64	12.29
	vol (%)				11.54	11.20	16.20
	SR				0.50	0.27	0.60
9 Anom.-AR12 Ridge	r (%)				12.19	11.50	13.46
	vol (%)				13.42	11.23	14.92
	SR				0.71	0.79	0.73
9 Anom.-AR12 ElNet	r (%)				8.12	8.75	11.15
	vol (%)				11.49	10.64	17.69
	SR				0.48	0.57	0.48
AR12 OLS	r (%)				12.18	9.98	12.10
	vol (%)				14.94	15.32	19.77
	SR				0.64	0.48	0.48
AR12 LASSO	r (%)				11.14	8.95	11.70
	vol (%)				14.74	14.80	19.89
	SR				0.58	0.43	0.46
AR12 Ridge	r (%)				12.01	10.89	11.51
	vol (%)				14.96	15.72	19.67
	SR				0.63	0.52	0.45
AR12 ElNet	r (%)				11.51	8.72	12.11
	vol (%)				15.80	13.40	17.28
	SR				0.56	0.45	0.55
PCA OLS	r (%)				11.37	9.44	13.21
	vol (%)				14.91	13.91	17.10
	SR				0.59	0.49	0.62
PCA LASSO	r (%)				5.41	5.32	6.39
	vol (%)				11.49	11.91	15.18
	SR				0.24	0.23	0.25
PCA Ridge	r (%)				11.91	11.05	12.97
	vol (%)				15.08	14.32	16.86
	SR				0.61	0.59	0.61
PCA ElNet	r (%)				12.11	9.32	12.54
	vol (%)				14.94	16.74	19.27
	SR				0.63	0.40	0.51
PCA-AR12 OLS	r (%)				9.75	7.09	10.08
	vol (%)				12.66	11.28	18.16
	SR				0.56	0.39	0.41
PCA-AR12 LASSO	r (%)				9.53	6.40	10.66
	vol (%)				13.10	11.55	18.90
	SR				0.53	0.33	0.42
PCA-AR12 Ridge	r (%)				8.49	7.18	5.90
	vol (%)				12.85	11.87	15.20
	SR				0.46	0.38	0.21
PCA-AR12 ElNet	r (%)				12.12	12.69	13.81
	vol (%)				13.74	12.92	19.28
	SR				0.69	0.78	0.58

Table 8: This table summarizes the annualized return, volatility, and Sharpe Ratio for all models and all specifications (25 in total). In addition, the index return for the sample (”Index”), the index return with the trading filters (”Index**”), and an equal weight (1/N) strategy.

5 Concluding remarks

5.1 Discussion

The results from the last chapter have showed that there are still stock market anomalies in emerging markets, which is showed by an analysis of hedge portfolios and two step analysis.

The formation of hedge portfolios is often considered beneficial as it corrects for market wide movements and has zero setup costs (when disregarding trading costs), and the results in this paper are sometimes mentioned in a long and a short leg. This is not entirely fair to other research, but it points into a direction of the source of an anomaly. Further research might focus on the exploration of optimal hedge portfolio specifications, in the sense that non-equal weighing schemes might be used or the portfolios could be based on other quantiles than the decile. Analyzing the hedge portfolio returns can add to the pronunciation of the results.

The two step analysis is based on the empirical methodology of Brennan et al. (1998) and Chordia et al. (2014), the authors adjust the stock returns for risk in the first step and regress the stock specific characteristics on the risk adjusted returns in the second step. This overcomes the problem of error in variable which is inherent to the original Fama and MacBeth (1973) approach in a large cross section. The usage of five principal components is based on literature but further research might implement a method to calculate emerging market wide Fama-French factors, or use a dynamic factor selection approach like Bai and Ng (2008).

The examination of the trend in the hedge portfolio returns yielded significant results for only three out of nine portfolios, and these returns might be undone by a risk adjusted approach. No significant attenuation in the returns seems infeasible because then the returns are persistent and robust to trading costs. Further research can adjust the holding period to assess the long-term profitability of the hedge portfolios. In the two step procedure a significant trend was hardly found, and not always with the expected direction. In this large cross section, the risk of poor data is always inherent and while all analyses are conducted with care and thought data quality is always a big risk in emerging markets.

Forecasting the returns with the regression and shrinkage methods has not yielded very good results. Forecasting returns is always difficult and further research might focus on in sample optimization of these models. The results have showed promising possibilities for shrinkage methods, which is an addition to the results from Nardi and Rinaldo (2011) and an opening for further research. The sample frequency might be one of the factors driving the results, perhaps changing to weekly observations forms a good trade-off between daily and monthly observations. Stock returns can also be researched on lower frequency with the usage of macroeconomic variables, this could open possibilities for the shrinkage methods to estimate such models robustly.

The portfolio optimization showed that it is very difficult to improve over the benchmark indices and the benchmark portfolios which are based on average historical returns. The implementation of forecast models and covariance matrix shrinkage seems to have introduced much uncertainty in the portfolios which translate to undesirable properties in comparison to the benchmark indices. It is nice to have a minimum variance portfolio but no investor will invest in a minimum variance emerging markets portfolio, while still the portfolio is expected to have much more risk (Bekaert et al., 2007). Implementing different covariance estimation techniques could perhaps improve the portfolios.

5.2 Conclusion

Based on the results in Section 4 the conclusion can be drawn that anomalies still have significant influence in emerging markets. Hedge portfolios based on asset growth, book-to-market value, illiquidity, issuance, idiosyncratic volatility, short-term reversal, momentum and turnover anomalies showed significant returns. Only the asset growth, book-to-market value, issuance and momentum portfolios show significant results after considering turnover and trading costs. An analysis of the trend in the book-to-market value portfolio shows that it has attenuated, and for the illiquidity and idiosyncratic volatility portfolio that it has accentuated in the sample period.

The two step approach yielded highly significant characteristic premiums for all anomalies. The trend in the characteristic premiums was only significant for the asset growth, momentum and size premium. These results imply that the effect of the characteristics often has not changed which implies that these premia are still priced in the market. Emerging markets show effects in the characteristics premiums that are not comparable to and explainable from developed market research. The inherently different characteristics of emerging markets will stay an interesting field of study for some time.

Some of the anomaly effects are showed to have accentuated or attenuated over time, while attenuation is expected from the literature accentuation is not. Positive significant hedge portfolio returns are still possible which would be impossible from an efficient market stance. These results show that emerging markets have included more information and have become more efficient in some aspects but progress is to be made.

The results form the forecasts of individual stocks have showed that the anomaly variable specifications perform worst of all variable specifications. An improvement is made when model is estimated with the principal components from the anomaly variable matrix but the largest improvement comes from including autoregressive lags. The autoregressive, PCA, and autoregressive plus PCA specifications perform best in the sense of accuracy based on MSPE, MAE and RMSE. The autoregressive model forecasts based on shrinkage methods are significantly more accurate while considering the Diebold-Mariano test. All models perform worse than a random guess in the Pesaran-Timmerman test for directional accuracy.

Aggregating the information in the anomalies into a selection of PCs seems to reduce estimation error in the forecast models. Which makes sense as only the most important drivers in changes in the variable are included.

Including the individual stock forecasts in optimized portfolios does not yield improved portfolios when regarding returns, volatility or Sharpe ratio. Even more, the portfolios based on forecasts are often superseded by portfolios based on average returns. That being said, the optimized portfolios were generally not able to improve the benchmark indices on return, volatility or Sharpe ratio. Optimizing portfolios for out of sample performance is still not easy as many parameters have to be estimated which increases uncertainty.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, *5*(1), 31–56.

Amihud, Y. & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics*, *17*(2), 223–249.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259–299.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2009). High idiosyncratic volatility and low returns: international and further us evidence. *Journal of Financial Economics*, *91*(1), 1–23.

Bai, J. & Ng, S. (2008). *Large dimensional factor analysis*. Now Publishers Inc.

Baker, N. L. & Haugen, R. A. (2012). Low risk stocks outperform within all observable markets of the world. *Available at SSRN 2055431*.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, *9*(1), 3–18.

Barry, C. B., Goldreyer, E., Lockwood, L. & Rodriguez, M. (2002). Robustness of size and value effects in emerging equity markets, 1985–2000. *Emerging Markets Review*, *3*(1), 1–30.

Bekaert, G., Harvey, C. R. & Lundblad, C. (2007). Liquidity and expected returns: lessons from emerging markets. *Review of Financial Studies*, *20*(6), 1783–1831.

Blitz, D., Pang, J. & Van Vliet, P. (2013). The volatility effect in emerging markets. *Emerging Markets Review*, *16*, 31–45.

Blitz, D. & Van Vliet, P. (2007). The volatility effect: lower risk without lower return. *Journal of Portfolio Management*, 102–113.

Brennan, M. J., Chordia, T. & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, *49*(3), 345–373.

Chordia, T., Subrahmanyam, A. & Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, *58*(1), 41–58.

Choueifaty, Y. & Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, *35*(1), 40.

Chuhan, P. (1994). *Are institutional investors an important source of portfolio investment in emerging markets?* World Bank Publications.

Connor, G. & Korajczyk, R. A. (1988). Risk and return in an equilibrium apt: application of a new test methodology. *Journal of Financial Economics*, *21*(2), 255–289.

Cooper, M. J., Gulen, H. & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, *63*(4), 1609–1651.

Datar, V. T., Naik, N. Y. & Radcliffe, R. (1998). Liquidity and stock returns: an alternative test. *Journal of Financial Markets*, *1*(2), 203–219.

Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 253–263.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Fama, E. F. & MacBeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. *The Journal of Political Economy*, 607–636.

Griffin, J. M., Kelly, P. J. & Nardari, F. (2010). Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets. *Review of Financial Studies*, hhq044.

Griffin, J. M., Nardari, F. & Stulz, R. M. (2007). Do investors trade more when stocks have performed well? evidence from 46 countries. *Review of Financial studies*, *20*(3), 905–951.

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *Review of Financial studies*, *8*(3), 773–816.

Harvey, C. R., Liu, Y. & Zhu, H. (2014). ... and the cross-section of expected returns.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 881–898.

Jegadeesh, N. & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Kim, H. H. & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. *Journal of Econometrics*, *178*, 352–367.

Ledoit, O. & Wolf, M. (2003a). Honey, i shrunk the sample covariance matrix. *UPF Economics and Business Working Paper*, (691).

Ledoit, O. & Wolf, M. (2003b). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, *10*(5), 603–621.

Li, X., Sullivan, R. N. & Garcia-Feijóo, L. (2014). The limits to arbitrage and the low-volatility anomaly. *Financial analysts journal*, *70*(1), 52–63.

Lo, A. W. & MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, *45*(1), 181–211.

Nardi, Y. & Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis*, *102*(3), 528–549.

Pafka, S. & Kondor, I. (2002). Noisy covariance matrices and portfolio optimization. *The European Physical Journal B-Condensed Matter and Complex Systems*, *27*(2), 277–280.

Pesaran, M. H. & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, *10*(4), 461–465.

Pontiff, J. & Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, *63*(2), 921–945.

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk*. *The journal of finance*, *19*(3), 425–442.

A Appendix

A.1 Hedge portfolios

	Panel A					Panel B				
	r (%)	SD (%)	TO (%)	t-stat		r (%)	SD (%)	TO (%)	t-stat	
AG	0.85	6.25	17.2	2.11	**	-1.23	6.32	12.7	-3.01	***
BM	1.20	6.36	16.5	2.98	***	-0.86	5.57	13.8	-2.45	**
ILLIQ	-0.17	7.73	56.9	-0.35		0.81	6.01	44.9	2.12	**
ISSUE	-0.27	5.91	14.4	-0.71		0.25	5.58	15.7	0.70	
IVOL	-0.50	7.07	54.2	-1.11		0.39	3.65	63.2	1.70	*
R1	0.29	6.53	85.1	0.71		-0.65	7.19	84.5	-1.42	
R212	0.61	5.97	29.4	1.57		-0.57	7.60	31.9	-1.16	
SIZE	0.36	5.47	8.1	1.05		-0.03	4.97	22.7	-0.09	
TURN	0.66	7.40	32.3	1.41		-0.05	3.94	43.2	-0.21	

Table 9: SD denotes standard deviation, TO denotes portfolio turnover. Panel A: This panel gives the monthly average raw returns of the long anomaly decile portfolios. Significance for a two sided test with the null hypothesis that individual returns are zero at the 10%, 5% and 1% level is indicated by *, **, and *** respectively. Panel B: This panel gives the monthly average raw returns of the short anomaly decile portfolios. Significance for a two sided test with the null hypothesis that individual returns are zero at the 10%, 5% and 1% level is indicated by *, **, and *** respectively.

	Mean	SD(mean)	T-stat
AG	1.94	0.18	10.95
BM	1.91	0.22	8.51
ILLIQ	-0.40	0.41	-0.98
ISSUE	-0.38	0.15	-2.50
IVOL	-0.30	0.32	-0.92
R1	0.10	0.31	0.32
R212	0.87	0.38	2.27
SIZE	0.24	0.31	0.77
TURN	0.34	0.36	0.93

Table 10: This table states the mean returns with consideration of portfolio turnover and trading costs. SD(mean) is calculated by dividing the standard deviation by the square root of T (with $T = 249$).

	Panel A			Panel B		
	b	t-stat		b	t-stat	
AG	10.72	1.50		16.08	2.21	**
BM	7.12	1.00		14.29	2.19	**
ILLIQ	20.63	2.34	**	5.42	0.80	
ISSUE	11.20	1.64		9.41	1.48	
IVOL	26.59	3.39	***	4.70	1.13	
R1	15.14	2.05	**	17.46	2.13	**
R212	9.26	1.36		15.88	1.86	*
SIZE	7.65	1.21		15.62	2.89	***
TURN	6.92	0.81		9.55	2.16	**

Table 11: This table states the coefficients for the exponential trend model. Panel A: Trend in HPf returns long leg. Panel B: Trend in HPf returns short leg.