ERASMUS UNIVERSITY ROTTERDAM

ECONOMETRIC INSTITUTE

An Analysis of the Idiosyncratic Risk Puzzle

MASTER THESIS - QUANTITATIVE FINANCE

Author:

Jan-Willem Engelen

(335435)

Supervisor:

Erik Kole

Co-reader:

Karolina Scholtus

Rotterdam, August 2017

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Abstract

Understanding the relation between idiosyncratic risk and returns has been a highly debated subject among researchers in the past. Ang et al. (2006, 2009), among others, find evidence in support of the Idiosyncratic Risk Puzzle which postulates that idiosyncratic risk is negatively correlated with equity returns. This paper provides evidence supporting the negative relation between idiosyncratic risk and returns found in Ang et al. (2009) using the CAPM and three factor Fama French model within a similar Fama-MacBeth framework. However, when we sort the idiosyncratic risk values into quantiles, we do not observe a monotonous relation over the spectrum of idiosyncratic risk values. More specifically, we find a large spike in returns for the smallest 20% quantile in terms of size and for the lowest 20% in terms of idiosyncratic risk values on both an equally-weighted and value-weighted basis for the double-sort portfolio. And so, while we find evidence to support the findings of Ang et al. (2009) that there is a negative relation between idiosyncratic risk and returns, we believe that Fama-MacBeth regressions do not serve as a comprehensive analysis of the relation between idiosyncratic risk and returns. Ultimately, we find evidence that the negative relation identified by Ang et al. (2009) might simply be limited to a select few small-sized companies with low idiosyncratic risk values and relatively high returns. In addition, we have shown that similar conclusions are drawn when performing this exercise using a PCA approach, rarely applied to idiosyncratic risk in previous literature.

Keywords: Idiosyncratic Risk; Cross-section of Stock Returns; Portfolio Sorts; (PCA) Factor Model

1 Introduction

The total risk of an asset is comprised of two parts: systematic risk and idiosyncratic risk. Idiosyncratic risk is defined as the risk attributable to a specific asset, and so, is independent of systematic risk, the uncertainty inherent to the market. Classical economic theory leads us to expect high systematic risks to be coupled with high returns and low systematic risks to be coupled with low returns. However, recent papers such as those of Ang et al. (2006) and Ang et al. (2009) encounter a negative relation between idiosyncratic risk and stock returns where theory suggests that there should be no premium at all. These unexpected findings has led to what is known in the literature as *the idiosyncratic risk puzzle*.

In theory, idiosyncratic risk can be substantially mitigated or eliminated from a portfolio through hedging and adequate diversification as first shown by the works of Markowitz (1952) and Sharpe (1963). Our initial expectation may well be that idiosyncratic risk should not be a great concern for investors due to the fact that idiosyncratic risk is essentially diversifiable and, therefore, negligible. However, in practice, this premise does not always hold, for several reasons. Firstly, investors might not hold diversified portfolios due to, for example, stock liquidity constraints or even by choice. Moreover, an investor with relatively limited means in terms of wealth might be concerned by an increase in idiosyncratic risk as this will, in general, increase the number of securities an investor must hold to achieve full diversification as well as increase transaction costs. Lastly, the total profits of arbitrageurs and option traders are dependent on *total* risk, that is, systematic risk plus idiosyncratic risk, meaning that these traders will be interested in the building blocks that drive these options' values. Clearly, idiosyncratic risk is a factor which investors monitor - all the more reason that it is defined and estimated correctly.

The bulk of papers within the literature utilise asset pricing models, such as the Fama-French three factor model. These models have as an advantage that the results are more easily interpreted as its constituents are clearly defined market factors. Few papers have approached the idiosyncratic risk puzzle from a statistical angle where, although the results may be less precisely interpretable, this offers a profound manner in which to debunk the common misconception that stocks wielding lower idiosyncratic risk are expected to yield higher returns than stocks with higher idiosyncratic risk. The idea behind this thesis is to use statistical models as well as the commonly used Fama-French models in order to isolate idiosyncratic from systematic risk. Once we have isolated idiosyncratic risk, we search for a relation between idiosyncratic risk and returns. Finally, we try to identify whether this relation holds true over the entire sample of whether we attribute this to a select sub-sample.

In order to help us achieve our goal, we have identified four important steps. Firstly, we look to replicate the results found by Ang et al. (2006) using U.S. daily and monthly data. This is done by first calculating monthly idiosyncratic risk using daily data and the Fama-French three factor model. Next, these monthly idiosyncratic risk values are regressed over the cross-section of stocks in attempt to mimic the negative relation between idiosyncratic risk and returns found by Ang et al. (2006).

Secondly, we test whether this relation is positive when adopting a statistical approach, such as is found by Spiegel and Wang (2006) and Fu (2009) by applying PCA. The use of principal components allows us to approach the theoretical definition of systematic risk and, in turn, idiosyncratic risk. The aim is to demonstrate that the relation between idiosyncratic risk and returns is not always found to be negative, thereby reducing the robustness of the results found by Ang et al. (2009).

Thirdly, we take differences between the Fama-French and PCA residual values, and find the idiosyncratic risk using these differences. We have established that using PCA to find idiosyncratic risk allows us to approach the theoretical definition of idiosyncratic risk. By subtracting the residuals found using the Fama-French three factor model, we are left with the portion of which *is* considered idiosyncratic through PCA analysis, but *not* through asset-pricing models such as that of Fama and French.

Fourthly, we hope to gain further insights into these differences by using single- and double-sorted portfolios. We first sort company returns according to the idiosyncratic values found using e.g. the Fama-French three factor model. Relative to the Fama-MacBeth regression, this provides an alternative insight as to the relation between idiosyncratic risk and returns. We can then go one step further and perform a double sort by sorting on market capitalisation and idiosyncratic risk. This allows us to analyse how returns vary for similar-sized firms with differing idiosyncratic risk values. The added value of a double-sort is the ability to capture changes in the variation of idiosyncratic risk in the second dimension for each quantile according to size in the first dimension. For both single and double sort portfolios, we will perform an analysis on both an equally- and value-weighted basis in order to further account for the fact that several small cap firms might produce extreme return values relatively speaking as compared to larger firms. If this is indeed the case, we expect to see a noticeable discrepancy between the two sets of results.

We find evidence to support the findings of Ang et al. (2009) and observe a similar negative relation between idiosyncratic risk and returns using Fama-Macbeth regressions. The question this paper seeks to answer is whether this negative relation is inherent for all company returns across the dataset, or whether there is a smaller sub-set of firms that convey this negative relation. We believe to have found evidence of the latter shown in several steps including single- and double sort portfolios based on market capitalisation and idiosyncratic risk using an equally- and value-weighted approach for the Fama-French three factor model and a PCA model.

2 Literature Review

Idiosyncratic risk is the standard deviation of the residual returns beyond what investors expected given that period's systematic return. While this may provide us with a clear definition of idiosyncratic risk, the manner in which to distinguish systematic risk from idiosyncratic risk is not well defined in the literature. This entails that the definition of idiosyncratic risk is not tied down to one specific model, but can subjectively be described by a variety of models.

Merton (1987) shows that in the presence of market frictions where investors have limited access to information, investors in equity with high idiosyncratic risk expect to be compensated for holding imperfectly diversified portfolios suggesting a positive relation between idiosyncratic risk and stock returns. Jones and Rhodes-Kropf (2003) show that investors demand a premium for holding stocks characterised by high, non-diversifiable, idiosyncratic risk. Behavioural studies such as that of Barberis and Huang (2001) offer a different type of asset pricing model based on prospect theory, where investors are loss averse over the fluctuations of individual stocks that they own. Similarly, they find that stocks which have higher idiosyncratic risk should earn higher expected returns.

Lehmann (1990) studies the significance of residual risk using a statistical approach and finds a statistically significant positive coefficient on idiosyncratic risk over his full sample using U.S. monthly data. Lehmann (1990) postulates that the residuals from the single index market model contain factors which are associated with non-zero risk premia and offers two plausible explanation for such: (i) non-linearity of the residual risk effect and (ii) the inadequacy of the statistical procedures employed to measure it. Malkiel and Xu (2002) employ the Capital Asset Pricing Model (CAPM) as a basic, first model with which to estimate systematic risk and, in turn, idiosyncratic risk. They obtain estimates for idiosyncratic risk based on monthly data and document a significant positive relation between idiosyncratic risk and average returns. Lehmann (1990) and Malkiel and Xu (2002) both crucially identify that the market model residual variances based should partially reflect exposure to any omitted sources of systematic risk. While we will briefly investigate idiosyncratic risk values based on the CAPM, this paper builds on the framework set out by Ang et al. (2009) which revolves around the three factor Fama-French model. Crucially, it should be noted that Lehamann (1990) utilises monthly data to produce idiosyncratic risk values where as we will use daily values within each month to produce monthly idiosyncratic risk values.

Several papers which analyse the U.S. market such as that of and Wang (2006) control for the size and value factors of Fama and French (1993) under the conjecture that market makers employ vehicles to hedge these established risk factors by including them in their investment models. As this paper will also focus on the U.S. market, we will adopt the approach as employed by, among others, Spiegel and Wang (2006). We will use the standard CAPM as well as the Fama-French three factor model to represent the returns of the market.

Ang et al. (2006) separate systematic risk from idiosyncratic risk using the Fama-French three factor model based on U.S. daily data, and their paper serves as one of the cornerstones that our research will build on. The authors are some of the first to scrutinise idiosyncratic risk over the cross section of stock returns as opposed to the aggregate time series. Using the cross section of stock returns allows the authors to control for an array of cross sectional effects. They measure a negative relation between idiosyncratic risk and stock returns over the cross-section. They find that this phenomenon cannot be explained by the size and value factors of Fama and French (1993), the momentum effect of Jegadeesh and Titman (1993), or the effect of liquidity risk as defined by Pástor and Stambaugh (2003).

This finding could, of course, simply be a classic case of data snooping, as described by Lo and MacKinlay (1990), however, the results found by Ang et al. (2006) have been independently confirmed by Brown and Ferreira (2003), Jiang, Xu, and Yao (2005), Huang et al. (2006), Zhang (2006), and Bali and Cakici (2008) for the U.S. market.

The research done on the U.S. market by Ang et al. (2006) is extended to a broad sample of international markets by the same authors in Ang et al. (2009) in order to verify the same pattern observed in the U.S. cross section. They find that the negative relation between idiosyncratic risk and average returns is statistically significant for each of the largest (G7) equity markets (Canada, France, Germany, Italy, Japan, the United States, and the United Kingdom). Having shown the negative relation to be a global phenomenon, it is hard to justify the findings of Ang et al. (2006) as a small-sample problem.

Other, more recent, papers such as that of Spiegel and Wang (2006) employ an exponential GARCH model, a dynamic model which captures the time variation of stock variance. Using U.S. monthly data, they obtain a positive relation between idiosyncratic risk and returns but concede that their paper leaves open the question as to why Ang et al. (2005) contrasting results in that idiosyncratic risk is negatively correlated with returns in the daily data. This paper also does not investigate why the results of daily versus monthly datasets differ, but instead adopts papers written by Ang et al. (2006, 2009) as its theoretical and methodological framework by using daily data.

Malagon et al. (2013) adopt a different approach when analysing the idiosyncratic risk puzzle. The authors investigate whether investor-specific characteristics such as investment horizon offers differing conclusions as to the relation between idiosyncratic risk and returns. They employ Wavelet Multiresolution Analysis (WMRA) which allows for the decomposition of a time series into different time horizons, called time scales, each of which correspond to a particular frequency. In terms of the relation between idiosyncratic risk and returns, Malagon et al. (2013) find a positive one for long-run investors and a negative one for short-run investors, indicating that the puzzle disappears as the wavelet scale increases (long-term horizons). While interesting, this paper will not analyse different time horizons, but aims to retain Ang et al. (2006, 2009) as the foundation of its framework. More specifically, we calculate our idiosyncratic risk values on a monthly basis and, accordingly, sort our portfolios on a monthly basis only.

Chen and Petkova (2012) look to build on the conclusion made by Ang et al. (2009) that a missing risk factor is the most likely explanation for the Idiosyncratic Risk Puzzle. They postulate that when the two components of market variance, one related to stock variances and the other related to stock correlations, are disentangled, evidence shows that the correlation component of total market variance not priced in the cross section of returns, while the variance component is priced. This paper does not seek to find a missing risk factor, but does hope to provide further insight as to which levels of idiosyncratic risk demonstrate the Idiosyncratic Risk Puzzle, if not all by analysing different portfolios containing increasing levels of idiosyncratic risk.

In the literature up until now, research which has focused on U.S. data suggests that the correlation between idiosyncratic risk and stocks is ambiguous: the correlation is found to range between positive and negative for different data sets scrutinised under varying asset pricing and statistical models. Most researchers use monthly data and utilise either the classic CAPM or Fama-French three factor model. Ang et al. (2006) and Ang et al. (2009) are some of the first set of researchers to use U.S. daily data, for which they observe a negative relation. The reason past papers often choose to employ asset pricing models such as the CAPM and the three factor model is because these are far easier to interpret in economical terms as compared to a more statistical approach, such as Principal Component Analysis (PCA).

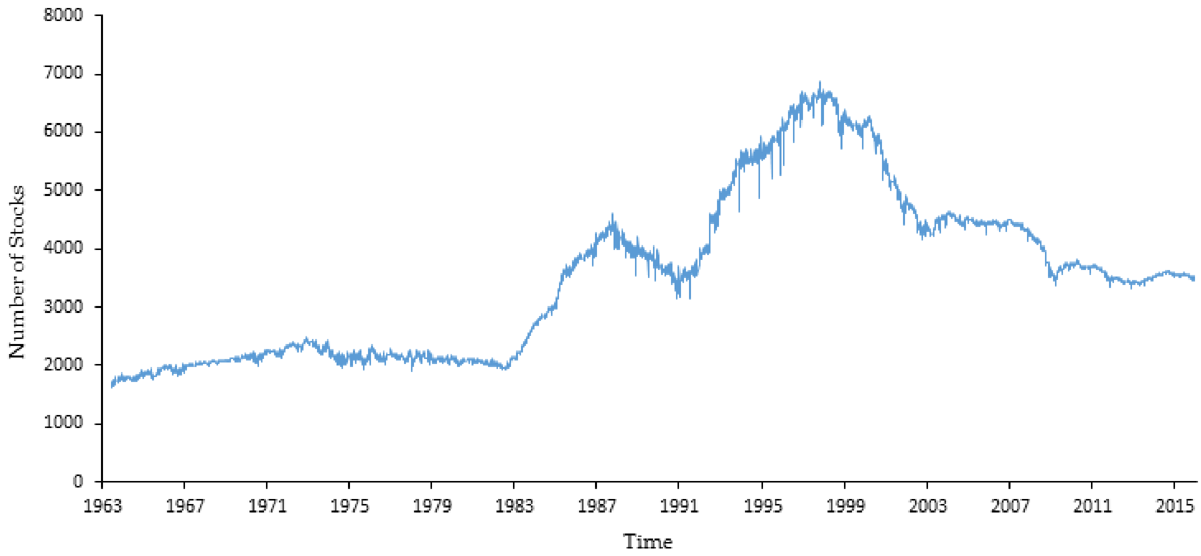
3 Data

The data set used in this paper mimics that of Ang et al. (2006) and Ang et al. (2009) for the U.S. market. The sample ranges from July 1st, 1963 to December 31st, 2015 and includes the NASDAQ, NYSE MKT (previously: AMEX) and NYSE indices. Stocks from the NASDAQ exchange are available from December 14th, 1972 while the data collected from the NYSE and NYSE MKT indices is available over the full range of the sample: July 1st, 1963 until December 31st, 2015.

The returns of each stock are collected using the CRSP/WRDS database on a daily basis. Specifically, returns of shares with share codes 10 and 11 are downloaded representing Ordinary Common Shares which have not and/or need not be further identified. These share codes are in accordance with data used by Fama and French (1993) as presented on their website. In addition, all prices under $1 and above $1000 are discarded from the data collection process. The returns are adjusted for split events which include stock splits, stock dividends, and other distributions with price factors such as spin-offs, stock distributions, and rights. The raw data is organised and structured using the statistical software package, Matlab, into a usable data set for our analysis. The structuring process renders 19,103 unique companies over the full time period, many of which, no longer exist at the present day. The number of stocks is depicted by Figure 1 and shows a gradual increase in the number of stocks across the three indices until its peak on October 28th, 1997: a lagged response corresponding to a more than 7% drop in the Dow Jones on October 27th, 1997. We observe several sharp troughs, noticeably between 1987 and 2000. These are predominately due to U.S. holidays such as Thanksgiving in November and Christmas in December. Exceptions are, for example, the blizzard of 1996 which swept over the east coast and temporarily closed stock exchanges.

The extracted returns of a single stock over time is defined as the percentage change in price from one day or month to the next. Excess Returns (R) are defined as those returns above the risk-free rate at a

Figure 1: Number of Stocks over Time



Shows the development of the number of stocks across the NASDAQ, NYSE MKT, and NYSE indices over the period July 1st, 1963 until December 31st, 2015.

certain point in time. The 1-month Treasury bill rate from Ibbotson Associates is used as a proxy for the risk-free rate, as proposed by Fama and French (1993). For the rest of this paper when we mention returns of any sort, we are referring to *excess* returns, be it excess stock returns or excess return on the market, unless explicitly specified. This is in accordance with Ang et al. (2006) and Ang et al. (2009).

The first three Fama-French factors Excess Market Return (EMR), Small Minus Big (SMB), and High Minus Low (HML) can be found on the Dartmouth University website and were obtained on a daily as well as on a monthly basis over the sample period July 1st, 1963 until December 31st, 2015. EMR is computed by subtracting the risk-free rate from the market return. SMB, or the size factor, is computed by subtracting the excess return of large companies, sorted by market capitalisation, from the excess returns of small companies. HML, or the value factor, is computed by subtracting the excess returns of companies with a high book-to-market ratio from the excess returns of companies with a low book-to-market ratio.

4 Methodology

The methodology set out in this paper is comprised of two main parts: The first part elaborates on the models used, both asset pricing as well as statistical models, and culminates with the calculation of idiosyncratic risk which uses the results of the aforementioned models to do so. Once idiosyncratic risk has been calculated, the second part of the methodology concerns itself with regressing idiosyncratic risk, testing the resulting coefficients, and employing single- and double sorts in order to further analyse the relation between market characteristics and idiosyncratic risk.

An important note regarding these two parts pertains to the use of notation for time. The first part uses daily observations, within each month, to calculate monthly idiosyncratic risk values. The second part regresses monthly returns on monthly explanatory values, including the monthly idiosyncratic risk values found in part one. In order to avoid any confusion caused by the notation of time, we use τ to denote *daily* observations, used throughout the entire methodology section, and t to denote *monthly* observations, predominately used in the second part of this section.

4.1 Part One: Calculating Idiosyncratic Risk

4.1.1 Asset Pricing Models

The papers written by Ang et al. (2006) and Ang et al. (2009) form the foundation of the methodological framework. These papers compute the idiosyncratic risk relative to the Fama-French three factor model which includes size and book-to-market risk. We define the basis of our asset pricing model as follows:

$$r_{i,\tau} = \alpha_i + \boldsymbol{\beta_i'} \boldsymbol{f_\tau} + \varepsilon_{i,\tau} \ , \tag{1}$$

where $r_{i,\tau}$ denotes the return of stock i at time τ, and $\boldsymbol{f_\tau}$ represents the vector of k factors. In the case of the Capital Asset Pricing Model (CAPM), k is equal to one and EMR_τ , the excess return on the market, is the only model factor. The Fama-French three factor model extends the CAPM by including a size and a value factor, SMB_τ and HML_τ, respectively. Equation (1) provides us with the resulting residual matrix $\varepsilon_{i,\tau}$, which is used to calculate idiosyncratic risk.

For this paper, we follow the example set by Ang et al. (2006) and Ang et al. (2009) and adopt the Fama-French three factor model as our main asset pricing model. In addition, we conduct the same preliminary calculations using the CAPM. This allows us to compare our results with those of past papers which also use the CAPM as well as to compare our results with the three factor model.

4.1.2 Statistical Models

Our statistical model uses Principal Component Analysis to describe a pre-defined proportion of the total variability in the data using K linear combinations of the original variables. More specifically, we analyse the daily idiosyncratic risk over the period of one month over the cross section of companies. For each month, we form an N_t x N_t correlation matrix. N_t new factors are then constructed from which K_t factors are selected such that the pre-defined threshold is attained and the resulting K_t factors explain 98% of the total variation in the data. Note here that we zoom-in on one particular month, and we regress over the days τ in that month and over the cross-section of active stocks N_t during a specific month t.

The factors are constructed in such that they are a linear combination of the original N_t variables and orthogonal to one another while simultaneously explaining the maximum amount of variance across the data set. These factors are independent of one another by design and each explains a certain unique proportion of the variation in the data. The factors are then used to make an N_t by N_t matrix of factor loadings where each column represents a factor and each row one of the original variables. The intersection of each row (asset) and column (factor) represents the loading, or sensitivity, of that asset on a particular factor.

The eigenvalue for a given factor measures the variance in all the variables which are accounted for by that factor and is computed as the sum of its squared factor loadings for all the variables in that column. The collection of eigenvalue forms the the 1 x N_t vector of eigenvalues, and the proportion of variance explained by a single factor is computed as the ratio of its eigenvalue to the sum of all eigenvalues. These proportions are ordered from largest to smallest and selected, one-by-one, until K_t principal components explain at least 98% of the variation in the data for a particular month t. For each month t, we regress the stock returns $r_{i,\tau}$ on the K_t monthly principal components $F_{K,\tau}$ where K denotes the K^{th} component.

$$r_{i,\tau} = \alpha_i + \boldsymbol{\beta_K F_{K,\tau}} + \varepsilon_{i,\tau} \tag{2}$$

In this manner, Equation (2) allows us to regress each time-series of daily stock returns on the selected K monthly principal components which again leaves us with the residual matrix $\varepsilon_{i,\tau}$. Note that Equation (2) is regressed separately for each month, t.

4.2 Calculating Idiosyncratic Risk

In order to calculate idiosyncratic risk, the resulting residual matrix $\varepsilon_{i,\tau}$ is transformed in accordance with Ang et al. (2009). That is, the idiosyncratic risk is found by taking the standard deviation of the residuals, for each stock, over the past month.

$$\sigma_{i,t} = \sqrt{Var(\varepsilon_{i,t,\tau})} \tag{3}$$

The values $\sigma_{i,t}$ are placed into a matrix comprised of N vectors of monthly idiosyncratic risk values for each stock i. To be clear, Equation (3) is regressed separately for each month, t, over all trading days τ and each stock i during that month.

4.3 Part Two: Regressing Idiosyncratic Risk

4.4 Fama-MacBeth

The relation between the idiosyncratic risk values and monthly stock returns are examined using a two-stage Fama-MacBeth regression. In the first stage, for each month, the cross-sectional firm returns are regressed onto idiosyncratic risk, factor loadings, and other control variables. An advantage of this method is that the cross-sectional regressions allow for controls of multiple factor loadings and other characteristics in a setting that retains power and has reduced noise as compared to single- and doubled-sorted portfolios.

$$r_{i,t} = \alpha_t + \gamma \boldsymbol{\sigma_{i,t-1}} + \lambda'_\beta \beta_{i,t} + \lambda'_z z_{i,t} + \varepsilon_{i,t}\ , \tag{4}$$

where t represents the current month and $t-1$ represents the previous month. The coefficients α_t, γ, β, and λ'_z are found using Ordinary Least Squares. $\boldsymbol{\sigma_{i,t-1}}$ denotes the one month lag value of idiosyncratic risk, known at the beginning of the current month, as used in Ang et al. (2009).

We also follow the example set by Ang et al. (2009) in that we use contemporaneous factor loadings estimated over the current month, t. This is because a factor model explains high average returns over a time period with contemporaneous high covariation in factor exposure over the same time period if the factor commands a positive risk premium. Although Ang et al. (2009) obtain almost identical results when using past factor loadings $\beta_{i,t-1}$, they choose to mimic the regressions run by Black et al. (1972), Fama and French (1992), and Jagannathan and Wang (1996), and so, choose to use contemporaneous factor loadings, $\beta_{i,t}$. Equation (4) is performed for each month t in our dataset which renders us vector of monthly γ coefficients.

In the second stage, we test the extracted vector of monthly γ coefficients. Specifically, we examine the sign and significance of the average value $\bar{\gamma}$, and we apply the Student's t-test to the vector of monthly γ coefficients.

$$t_{score} = \frac{\bar{\gamma}}{\sqrt{\widehat{Var}(\bar{\gamma})}} \tag{5}$$

Equation (5) gives us a Student's t-test for the average value of gamma, $\bar{\gamma}$. The result of this t-test renders a positive, negative, or insignificant relation between stock returns and idiosyncratic risk.

4.5 Differences between Idiosyncratic Risk

Ang et al. (2006) find that the idiosyncratic errors of a misspecified factor model could contain influence of missing factors. In this paper we reason that Principal Component Analysis approaches the correctly specified model while lending itself less to economical interpretation as compared to an Asset Pricing model such as the CAPM or the three factor Fama-French model. Next, we adopt the reasoning offered by Ang et al. (2006) that the negative relation between idiosyncratic risk and returns found in the literature may be due to the misspecification of the factor models used, and in particular, due to omitted variable bias.

In an attempt to assign a controlled and interpretable variable to this apparent misspecification of the usual asset pricing models, we take the difference between the Fama-French residuals, found by Equation (1), and the PCA residuals, found by Equation (2), at each point in time, τ.

$$\delta_{i,\tau} = \varepsilon^{3FF}_{i,\tau} - \varepsilon^{PCA}_{i,\tau} \,, \tag{6}$$

where the differences between each pair of daily residuals is denoted as $\delta_{i,\tau}$. We then use Equation (3) to calculate monthly idiosyncratic risk values for these differences in residual values, denoted $\sigma^{\delta}_{i,t}$.

Once the idiosyncratic values of these differences have been found, we employ Equation (4) and include a combination of market factors such as the three original Fama-French factors. In a similar fashion as before, we then implement the second stage of the Fama-MacBeth regression, namely, to test for a significant relation between the idiosyncratic errors and the monthly stock returns through a Student's t test shown by Equation (5).

4.6 Portfolio Sorts

The aim of portfolio sorting in this paper is to provide an alternative to the Fama-MacBeth in testing the relation between returns and idiosyncratic risk. Crucially, portfolio sorting allows us to sort the cross-section of monthly stock returns into buckets using a stock-specific characteristic such as idiosyncratic risk. The exercise of sorting these stocks into different buckets allows us to analyse the relation between returns and idiosyncratic risk for different levels of idiosyncratic risk as opposed to a Fama-MacBeth regression which provides us with a coefficient describing the cross-section as a whole. Furthermore, the exercise of double sorting allows us to first sort the cross-section of returns using one stock-specific characteristic such as market capitalisation and second by another characteristic, such as idiosyncratic risk. To this end, we are interested in, for example, the stability or fluctuations of portfolio returns and alpha values across the buckets of varying idiosyncratic risk size for a particular size of market capitalisation.

4.6.1 Single Sort

Regardless of the model used, we obtain idiosyncratic risk values using Equation (3) and create a matrix $\sigma_{i,t}$, the idiosyncratic risk of asset i at time t. For each month t, we then sort the number of stocks N_t into 5 quantiles based on their idiosyncratic risk values, from largest to smallest. Doing this for each month allows us to create 5 quantiles, each representing a portfolio characterised by idiosyncratic risk, over time. We can then average the returns of each quantile per month and regress these values using, for instance, the three Fama-French model given by Equation (1). To be completely clear, within each quantile, each month contains a set of companies with similar idiosyncratic risk values and corresponding returns. A regression over these companies results in average returns per month per quantile. These average returns can be regressed over time such that the resulting alpha measures the profitability of each quantile.

4.6.2 Double Sort

The double sort extends the single sort by sub-dividing each quantile into 5 sub-groups. Here, we first sort our portfolio of firms per month according to market capitalisation. In a similar manner, we now sort each of the 5 existing quantiles, on a monthly basis, according to the idiosyncratic risk values, found using a particular model e.g. 3FF or PCA, within each of the first five quantiles. As an example, for each month we divide all stocks N_t into 5 equally-sized quantiles according to their market capitalisation . Next, for the same month t, each quantile is sub-divided into 5 equally-sized portfolios according to the idiosyncratic risk values found using the 3FF model. This is then done for all months t of the sample, creating 5 x 5 portfolios over time. The average of each is then taken over the months to obtain 25 average monthly returns.

Similarly to the single-sorted portfolios, we are now interested in the average returns of these portfolios. If the results across the second sort portfolios are relatively 'stable' or unchanging for a particular quantile of market capitalisation where this was not the case for the full range of market capitalisation, we can infer the effect of idiosyncratic risk on returns to be tied to size, to a certain extent.

4.6.3 Value-weighted Sort

A value-weighted portfolio sort allows us to gauge the variability of IR values in each quantile for each month. The steps to obtain value-weighted portfolios are identical to those of the single- and double equally-weighted portfolio with the exception that we calculate a weighted average for each quantile, based on market capitalisation, instead of the arithmetic mean. In other words, take for example a single sort value-weighted portfolio: for each month we rank all idiosyncratic risk values from largest to smallest and divide these into, for example, 5 quantiles just as before. The difference is now that, within each quantile, we weight the firm returns using the market capitalisation of the corresponding firm at the beginning of the month, and we calculate the value-weighted average return of each quantile for that month. Average returns over time and alpha values for each quantile, be it for single- or double sorted portfolios, can then be calculated in the same manner as for equally-weighted portfolios. Here, we expect to see discrepancies between equally- and value-weighted results when the weights (i.e. market capitalisations) within a quantile are substantially different from those of the arithmetic mean, 1/N, N being the number of stocks in each quantile.

5 Results

5.1 Calculating Idiosyncratic Risk

We have established that idiosyncratic risk (IR) values depend on the models used to isolate these values. In order to provide a better understanding of how these values differ for the models we have used in this paper, we refer to Table 1 which provides an overview of summary statistics for each set of (IR) values corresponding to the different models used.

Table 1: Idiosyncratic Risk Summary Statistics

	Mean	Median	Max	Min	Std. Dev	Dispersion	Skewness	Kurtosis
CAPM	3.628	3.276	21.791	0.090	1.819	1.696	1.463	7.490
3FF	3.592	3.239	21.731	0.148	1.809	1.686	1.454	7.347
PCA98	0.040	0.022	5.692	0.000	0.095	0.035	27.395	1347.996
3FF - PCA98	3.588	3.237	21.731	0.148	1.808	1.685	1.451	7.315

Shows the summary statistics for the distribution of IR values for each model shown on the left.

Looking at the first two models, the CAPM and three Fama-French (3FF) model, of Table 1, we see that adding the factors High Minus Low (HML) and Small Minus Big (SMB) offers small improvements in terms of almost all metrics with the exception of the minimum value, which shifts slightly to the right and away from zero. However, these differences are minor at best, and it is safe to say that, based on Table 1, these models produce similar IR values. Their distributions are represented by the histogram for the 3FF model shown in Panel A of Figure 2, as the distributions are, visually, identical.

If we then take the preceding two models and compare these to the PCA98 model where PCA98 represents principal component analysis using a 98% variation threshold, we see substantial improvements in moving from a factor model to a statistical model. Our reasoning for choosing 98% is two-fold. Firstly, while Fama-French models only use a handful of defined factors, a model based on PCA of e.g. 4 or 5 factors would not be able to explain a high percentage of the variability in the data. On the other hand, while increasing the number of included principal components in a regression helps us to explain a higher percentage of the variability, we still require a small amount of defined idiosyncratic risk in order to properly perform our analysis. We expand on this reasoning in the following section. On average over time, the PCA98 model selects 20 principal components in order to explain at least 98% of the variability in the cross-section of companies each month. Besides the mean and minimum shifting dramatically towards zero, Panel B of Figure 2 shows that the skewness of the distribution increases substantially which is also reflected by the relation of the median to the mean as compared to the original factor models shown in Table 1. In observing the skewness we note that the maximum of the distribution has not reduced to the same extent as the other values of the distribution, also portrayed by Panel B of Figure 2, caused by several large outliers. Despite several extreme values, the kurtosis value of the distribution is high meaning many observations lie around the mean and farther out in the tails. Based on these metrics, we conclude that PCA is successfully able to identify a selection of common factors which substantially reduces the IR values in general and shifts these towards zero. This leads us to suspect that when subjecting these IR values to the Fama-MacBeth regression, that we are more likely to see a neutral (or insignificant) relation between PCA IR values and returns than factor model IR values and returns.

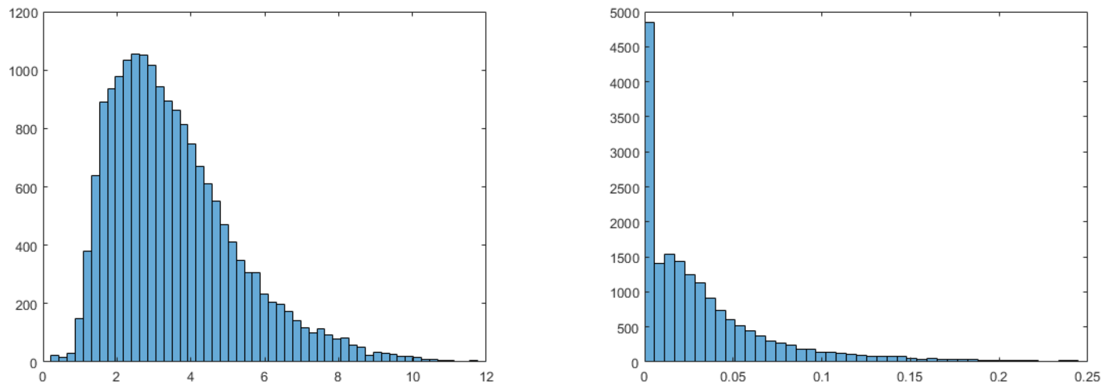
The differences between 3FF and PCA residuals does not produce very different results from their CAPM or three Fama-French factor counterparts. We attribute this to the observation that the stand-alone IR values of the 3FF model are substantially larger than those using PCA. Subtracting the latter from the former, in this case, does not result in drastically different IR values. For this reason, Panel A of Figure 2 serves as a visual representation of the IR 3FF - PCA model, as the values are quite similar to those of the IR 3FF model.

Based on preliminary analysis of the idiosyncratic risk values, we do not see significant differences between the factor models: CAPM, 3FF and 3FF - PCA. In addition, we note that utilising the PCA method to obtain IR values produces less extreme IR values than using factor models such as the CAPM or three factor model, which tells us that the PCA98 model produces lower risk values in general.

Figure 2: Distributions of Idiosyncratic Risk

Panel A: IR 3FF

Panel B: IR PCA (98%)



The panels above depict the distributions of idiosyncratic risk values for the (A) 3FF and (B) PCA 98% models.

5.2 Factor Analysis

In order to obtain a better understanding of the models used to calculate idiosyncratic risk, we must scrutinise the factors that make up our two main models: the three factor Fama-French model and our model based on principal component analysis. Table 2 shows the correlation coefficients found for the first ten principal components using PCA compared against well-known economic factors as defined by Fama-French including: Excess Return on the Market (EMR), Size (SMB), Value (HML), Robust Minus Weak (RMW), Conservative Minus Aggressive (CMA), Short-term Reversal (STR) and Momentum (MOM). The coefficients tell us that the first principal component is adequate proxy for the market factor, EMR, at 60%. While the coefficients describing all other relations between principal components and market factors fluctuate slightly, they are all deemed to be quite weak at less than 31% in terms of the absolute value for the first principal component and near-zero for all other principal components. Table 2 tells us that we see some correlation between the market factor and first principal component, but that the other individual principal components are not proxies for any of the other market components shown in the table.

Table 2: Factor Correlations

	Market	Size	Value	RMW	CMA	STR	MOM
PC 1	0.5977	0.1301	-0.2077	-0.2599	-0.2931	0.3087	-0.1183
PC 2	-0.0071	0.0266	-0.0101	-0.0175	-0.0141	-0.0026	-0.0365
PC 3	-0.0068	0.0050	0.0103	0.0096	0.0159	-0.0101	-0.0020
PC 4	-0.0079	0.0018	0.0269	-0.0068	0.0109	-0.0034	-0.0244
PC 5	-0.0165	0.0366	0.0058	-0.0236	0.0087	0.0076	-0.0066
PC 6	-0.0061	0.0064	0.0062	-0.0224	0.0035	-0.0181	-0.0176
PC 7	0.0031	0.0070	-0.0014	-0.0036	-0.0008	-0.0142	0.0034
PC 8	-0.0015	-0.0024	0.0018	-0.0051	-0.0018	0.0036	0.0003
PC 9	0.0029	-0.0009	0.0032	-0.0033	-0.0122	-0.0099	-0.0010
PC 10	-0.0026	0.0111	0.0061	-0.0072	0.0098	0.0045	-0.0085

The table above shows correlations between the time series of monthly values for principal components in the rows and market factors as columns. The market factors from left to right are Market, Size, Value, Robust Minus Weak (RMW), Conservative Minus Aggressive (CMA), Short-term Reversal (STR) and Momentum (MOM).

Table 3 demonstrates the average variation in the data explained by each principal component. As expected every next principal component explains a smaller amount of variation in the data compared to the preceding component. There are two main observations to be noted here. Firstly, the first principal component, which we saw to be somewhat correlated to the EMR market factor, does not explain more than 16% of the variation in the data, on average. Secondly, the first three principal components do not explain more than 30% of the data, on average, and the first ten don’t account for more than 65% of the variation in the data, on average. In order for a statistical model using PCA to account for at least 98% of the data, we need to incorporate the first twenty principal components, on average, due to the low percentage of variation explained by each principal component in this exercise. While the 98% threshold chosen is somewhat arbitrary, the reason it is chosen is two-fold. Firstly, each principal component has low added value in terms variation explained so we must incoporate several more factors as compared to the three factor Fama-French model in order to approach a small idiosyncratic value representing the theoretical definition of idiosyncratic risk. Secondly, a statistical model which would incorporate all principal components would essentially ignore idiosyncratic risk and render the approach useless.

Table 3: PCA: Average Variation Explained

	Variation Explained		Variation Explained
PC 1	15.5%	PC 12	4.1%
PC 2	7.5%	PC 13	3.9%
PC 3	6.4%	PC 14	3.8%
PC 4	5.9%	PC 15	3.6%
PC 5	5.5%	PC 16	3.5%
PC 6	5.3%	PC 17	3.3%
PC 7	5.0%	PC 18	3.1%
PC 8	4.8%	PC 19	2.6%
PC 9	4.6%	PC 20	1.8%
PC 10	4.4%	PC 21	0.9%
PC 11	4.3%	PC 22	0.2%

Principal Component Analysis is performed on a monthly basis to obtain a set of principal components each month. The table above shows the average variance explained in the data for each of the first 22 principal components over time.

In order to compare the ability of the principal components to capture the behaviour of the company data to that of the Fama-French models, we show the average R^2 and the standard deviation for both the CAPM as well as the three factor Fama-French factor in Table 4. The 3FF model explains slightly more variance in the data compared to the CAPM. The fact that the average R^2 value of both models are quite low does not surprise us, considering the fact that the first pincipal component, shown in Table 3, explains less than 15% of the variance in the data, on average. Principal components are statistically configured using company returns. It is, therefore, not surprising either that the first principal components do a better job at explaining variance in the data than Fama-French factors.

Table 4: Fama-French: Average R^2

	R^2	
	Average	Standard Deviation
CAPM	0.0654	0.0833
3FF	0.0877	0.0988

The table above shows the average R^2, or variance explained, by the CAPM and three factor Fama-French model as well as the respective standard deviations.

5.3 Fama-MacBeth Regressions

The Fama-MacBeth regressions allow us to scrutinise the relation between idiosyncratic risk and returns given a certain Fama-French or PCA factor model. Looking at Table 5, we see that the CAPM and three factor models seem to produce a significant negative relation given the t-scores -4.3 and -6.7 for the CAPM and 3FF respectively, lending support to Ang et al. (2009) findings. However, using a PCA approach, the t-score drops below the 95% threshold of 2, implying an insignificant relation between returns and idiosyncratic risk using a statistical approach. The 3FF - PCA98 differences approach yet again renders similar results to the CAPM and 3FF models. Table 5 shows Standard Errors and t-scores using both the standard OLS approach as well as using Newey-West corrections. The fact that the standard errors and t-scores only differ slightly tells us that the any intertemporal effect is minimal. While the significance level reduces slightly for each model, our conclusions remain the same.

Table 5: Idiosyncratic Risk Coefficients

		Standard OLS		Newey-West Correction	
	$\bar{\gamma}$	Standard Error	t-score	Standard Error	t-score
CAPM	-0.1577	0.0366	-4.3044	0.0468	-3.3726
3FF	-0.3366	0.0504	-6.6832	0.0611	-5.5100
PCA98	-0.0375	0.0765	-0.4906	0.0873	-0.4301
3FF - PCA98	-0.1975	0.0503	-6.6891	0.0610	-5.5178

Shows the Fama-MacBeth coefficient describing the relation between idiosyncratic risk and returns, denoted by $\bar{\gamma}$. Then, for each model on the left, the standard error and t-score using standard Ordinary Least Squares assumptions as well as when using a Newey-West estimator to correct for any possible dependency over time.

Following the observation made from Tables 1 and 5 regarding the CAPM and 3FF - PCA98 IR values, we conclude that both models do not provide material insights beyond what the 3FF model can offer us. From here on out, we will withhold our analysis surrounding the CAPM and 3FF - PCA98 model.

Results are available upon request.

The conclusions drawn on our observations so far surrounding the 3FF and PCA98 models differ from one another: The 3FF model supports a negative relation between returns and idiosyncratic risk while the PCA98 model does not imply a meaningful relation between the two. The latter finding is based on the t-score of the PCA 98% model in Table 5 and is visually supported by Panel B of Figure 2, which shows a large number of observations close to zero.

If the negative relation found using the 3FF model is due to, for instance, market tensions or the inability to diversify portfolios, then we would expect a monotone relation between idiosyncratic risk and returns for varying degrees of idiosyncratic risk. In other words, what this means is that we should expect to see the same change in returns in terms of direction as we move from one quintile of idiosyncratic risk value to the next throughout the entire distribution.

5.4 Single Sort Portfolios

5.4.1 Equally-weighted Portfolios

Single sort portfolios allow us to further scrutinise the relation between idiosyncratic risk and returns by dividing the idiosyncratic risk values into quintiles and then observing their average returns as well as their alpha values when regressed against three Fama-French factors. We administer this analysis using the three factor Fama-French model and the 98% PCA factor model. We exclude the CAPM model (and all other Fama-French models for that matter) from our results because this renders the same results as those of the three factor Fama-French model.

Table 6 shows the equally-weighted average returns and alpha values for the three factor Fama-French (3FF) and the PCA factor model (98%). Note that the the first quintile (Q1) contains the portfolio with the largest idiosyncratic risk values and the 5^{th} (Q5) with the smallest values. Furthermore, the last column in Table 6 displays the results for the full portfolio, i.e. where one bucket is used instead of five which displays the behaviour of the equally-weighted portfolio in general.

Using an equally-weighted 3FF (EW 3FF) model and looking at columns Q1 through Q5, the results show a negative relation between idiosyncratic risk and average excess returns as the average increases when idiosyncratic risk decreases from Q1 to Q5. The increase in average return and alpha values seems to be somewhat constant from Q1 to Q4, however, we see a large spike in moving into Q5. This tells us that, on average and according to the EW 3FF model, the lowest values of idiosyncratic risk are tied to firms with the highest average excess returns. Specifically, the Q1 and Q5 portfolio consisting of the highest and lowest idiosyncratic risk values produce annualised average excess returns of 3.4% and 41.9% respectively and alpha values of -2.8% and 28.5% respectively. We also notice that the negative signs tied to the alpha values of the first four quintiles signify that, on average, portfolios

Q1 to Q4 underperform for the 3FF model while Q5, consisting of firms with the lowest IR values, tends to outperform the market, on average, by 28.5%. The standard deviation of the average return increases from Q1 into Q5, showing that the variation in average return values in each portfolio widens as the idiosyncratic risk values decrease into Q5. The standard errors of the alphas on the other hand remain relatively constant between Q1 and Q4 but still show a substantial spike from 0.07 in Q4 to 0.18 in Q5 showing a large variation in values in the 5th quintile.

Observing these values for the equally-weighted PCA98 (EW PCA98) model and comparing these to the 3FF model, we notice no major changes in the standard deviations or standard errors of the average returns or alpha values between the two models in general. We note that the EW PCA98 model produces slightly larger increases in value from Q1 into Q4 as the relation in Q1 is characterised by a lower value and increases slightly more as compared to the 3FF model, however the differences between these models is small when using portfolios. Most importantly, the large spike in the 5th quintile still remains, signifying, on average, a large spike in average returns for firms with low idiosyncratic values. On average, the coefficient of the market factor, EMR, is close to one, which we would expect as our portfolio reflects all U.S. indices since 1963 and that the coefficients should reflect the market portfolio. We notice, however, that the sensitivity of the EMR coefficient increases into the 5th quintile implying that companies with lower idiosyncratic risk values are more sensitive to changes in the market as a whole. For both the EW 3FF and the EW PCA98 models we also see large spike in the SMB (Small Minus Big) coefficient. What is apparent is that in the 5th quintile, containing the 20% lowest idiosyncratic risk values, returns are more sensitive to the SMB factor than in other quintiles. The factor HML, representing the spread between value and growth stocks is small, in general, but slightly positive in general telling us that the stocks show some exposure to this factor. For the full portfolio we note that on an equally-weighted basis, the average (excess) return is given to be 1.06% on a monthly basis or 12.78% on a yearly basis.

Table 6: Equally-weighted Single Sorts (20% quintiles)

		Q1	Q2	Q3	Q4	Q5	Q5 - Q1	Full
EW 3FF	Avg Return	0.2802	0.4147	0.4753	0.7340	3.4905	3.2302	1.0646
	S.D. AR	*3.5445*	*4.7291*	*5.9141*	*7.3711*	*10.9105*	*9.1133*	*6.0160*
	Alphas	-0.2369	-0.2715	-0.2931	-0.1536	2.3766	2.6157	0.2736
	S.E. Alpha	*(0.0477)*	*(0.0482)*	*(0.0466)*	*(0.0675)*	*(0.1840)*	*(0.2158)*	*(0.0539)*
	EMR	0.7623	0.9721	1.1042	1.2303	1.4078	0.6448	1.0376
	S.E. EMR	*(0.0114)*	*(0.0115)*	*(0.0111)*	*(0.0161)*	*(0.0440)*	*(0.0515)*	*(0.0129)*
	SMB	0.1135	0.3648	0.7001	1.0664	1.9272	1.8123	0.8821
	S.E. SMB	*(0.0159)*	*(0.0160)*	*(0.0155)*	*(0.0224)*	*(0.0611)*	*(0.0717)*	*(0.0179)*
	HML	0.3055	0.2420	0.1062	-0.1041	-0.2565	-0.5607	0.1270
	S.E. HML	*(0.0172)*	*(0.0173)*	*(0.0167)*	*(0.0242)*	*(0.0660)*	*(0.0774)*	*(0.0194)*
	S.D. Resid	*1.1683*	*1.1776*	*1.1405*	*1.6470*	*4.4908*	*5.2646*	*1.3196*
EW PCA98	Avg Return	0.0656	0.3021	0.4657	0.8546	3.6922	3.6954	1.0646
	S.D. AR	*3.7621*	*4.7792*	*5.8373*	*7.2234*	*10.9031*	*8.9882*	*6.0160*
	Alphas	-0.4912	-0.3916	-0.2966	-0.0212	2.6237	3.1166	0.2736
	S.E. Alpha	*(0.0603)*	*(0.0497)*	*(0.0476)*	*(0.0665)*	*(0.1788)*	*(0.2188)*	*(0.0539)*
	EMR	0.7694	0.9800	1.1060	1.2166	1.4047	0.6347	1.0376
	S.E. EMR	*(0.0144)*	*(0.0119)*	*(0.0114)*	*(0.0159)*	*(0.0427)*	*(0.0523)*	*(0.0129)*
	SMB	0.1958	0.3764	0.6556	1.0292	1.9206	1.7245	0.8821
	S.E. SMB	*(0.0201)*	*(0.0165)*	*(0.0158)*	*(0.0221)*	*(0.0595)*	*(0.0728)*	*(0.0179)*
	HML	0.3094	0.2567	0.1143	-0.0640	-0.3188	-0.6266	0.1270
	S.E. HML	*(0.0216)*	*(0.0178)*	*(0.0171)*	*(0.0239)*	*(0.0642)*	*(0.0785)*	*(0.0194)*
	S.D. Resid	*1.4738*	*1.2160*	*1.1654*	*1.6238*	*4.3728*	*5.3395*	*1.3196*

For each model on the left, the monthly average return and alpha values for each quintile are shown. The standard deviation and standard error are shown for each under each number, respectively. In this table, higher quintiles represent lower idiosyncratic risk values. To the right of the five quintiles, the results are shown for the Q5 - Q1 trading strategy and the full portfolio.

Note that for the trading strategy Q5 - Q1 the coefficients are equal to the difference between the coefficients of Q5 and Q1 separately. Because the Q5 - Q1 trading portfolio mimics a portfolio long in low idiosyncratic risk and short in high idiosyncratic, this is to be expected. For all regression coefficients, the standard error has increased slightly reflecting the increased uncertainty in the coefficients obtained.

We also see that, for the full portfolio, the SMB coefficient is still substantially higher than zero and quite influential at 0.88. This sensitivity could be due to small cap firms with relatively high stock return values in the 5th quintile. As the results in Table 6 are based on equally-weighted (EW) portfolios, it could be the case that there are several smaller firms which produce relatively high return values, reflected by the positive size betas shown in Table 6. For this reason, we are interested in value-weighted (VW) portfolios and how these results then differ compared to the results of their equally-weighted counterparts. As we value-weight each bucket on a monthly basis, very different results signifies greater variation in size within each bucket.

5.4.2 Value-weighted Portfolios

Because we know the value-weighted portfolios are less influenced by small-cap firms, we can induce that there are several small-cap firms with high returns which skew the perceived relation between idiosyncratic risk and returns for the equally-weighted portfolio. We show the results for the value-weighted 3FF (VW 3FF) and PCA98 (VW PCA98) models in Table 7. Buckets Q4 ad Q5 contain the highest 40% of the idiosyncratic risk values and produce lower average returns for the VW portfolio compared to that of the EW portfolio. Where we see a substantial spike in average returns as well as alpha values for Q5 of the EW portfolio, average returns and alpha values of the VW portfolio in fact decrease slightly. The VW portfolios produce slightly higher average returns in the first three buckets containing the highest 60% of idiosyncratic risk values compared to the EW portfolio. These differences however are not substantial and are largely negligible. Because VW portfolios lend more weight to larger stocks within each bucket, it does not appear that small cap firms with high returns greatly influence our equally-weighted results for these first three buckets. Looking across the five quintiles of the VW portfolio we see that, in general, average returns decrease as idiosyncratic risk decreases, reflecting a slightly positive relation between the two. Furthermore, we also see that the SMB coefficient is slightly negative in Q1, increases over the quintiles and is slightly positive in Q5 telling us that average returns are less sensitive to the SMB coefficient for value-weighted portfolios.

When comparing average returns and alpha values for the full portfolio, i.e. when we do not sort the idiosyncratic risk values into separate buckets, we notice that the average returns and alpha values of the VW portfolios are lower than that of the EW portfolios in Table 6. The EW 3FF portfolio, for instance, produces an average return of 1.06% on a monthly basis or 12.78% annually while the VW 3FF portfolio shows a monthly average return of 0.51%, or 6.08% annually. Remembering that all returns shown in this report denote *excess* returns, the value-weighted average return is much more in line with what we might expect of the historical performance of U.S. indices in the last several decades. The VW alpha values have reduced substantially compared with the EW values, showing a near-zero alpha value of 0.01% for the VW portfolio compared to 0.27% of the EW portfolio.

In contrast to the equally-weighted portfolios, the trading strategy Q5 - Q1 of the value-weighted portfolio does not improve portfolio performance in terms of average returns or alpha. This is to be expected, as the spike observed in the 5th quintile of the equally-weighted portfolio is not featured for value-weighted results. Interestingly, the standard errors of the regression coefficients are only slightly different to that of the 5th quintile.

Lastly, we take a look at the full portfolio where the SMB coefficient is near-zero. This is to be expected for a value-weighted portfolio. Our results differ substantially based on the weight that is placed on each stock. Clearly, the relation between idiosyncratic risk and returns depends on the size of the companies in the portfolio and the method that is used to calculate average returns.

Table 7: Value-weighted Single Sorts (20% quintiles)

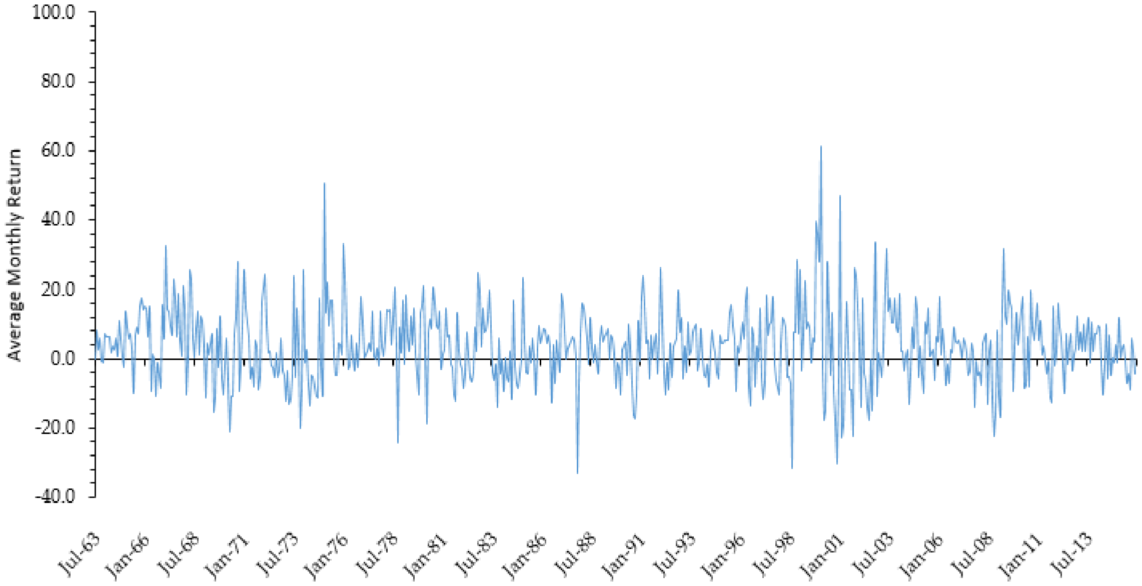
		Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q5 - Q1**	**Full**
VW 3FF	Avg Return	0.6375	0.5321	0.5425	0.3714	0.2809	-0.3566	0.5064
	S.D. AR	*4.3506*	*4.2781*	*5.4445*	*5.4686*	*5.8745*	*3.9179*	*4.4173*
	Alphas	0.2255	0.0551	0.0491	-0.1234	-0.3749	-0.6004	0.0129
	S.E. Alpha	*(0.0511)*	*(0.0524)*	*(0.0831)*	*(0.1056)*	*(0.1285)*	*(0.1488)*	*(0.0045)*
	EMR	0.9586	0.9474	1.0683	1.0390	1.1105	0.1519	0.9953
	S.E. EMR	*(0.0122)*	*(0.0125)*	*(0.0199)*	*(0.0252)*	*(0.0307)*	*(0.0356)*	*(0.0011)*
	SMB	-0.2050	-0.0909	0.1048	0.0784	0.1544	0.3594	-0.0126
	S.E. SMB	*(0.0170)*	*(0.0174)*	*(0.0277)*	*(0.0351)*	*(0.0428)*	*(0.0495)*	*(0.0015)*
	HML	-0.0435	0.0777	-0.2019	-0.1345	0.1767	0.2201	-0.0042
	S.E. HML	*(0.0184)*	*(0.0188)*	*(0.0299)*	*(0.0379)*	*(0.0462)*	*(0.0535)*	*(0.0016)*
	S.D. Resid	*1.2515*	*1.2831*	*2.0358*	*2.5853*	*3.1482*	*3.6442*	*0.1111*
VW PCA98	Avg Return	0.6131	0.5947	0.5293	0.5788	0.3827	-0.2304	0.5064
	S.D. AR	*4.5367*	*4.2368*	*4.4730*	*4.9593*	*5.8752*	*2.9004*	*4.4173*
	Alphas	0.1440	0.1758	0.1285	0.0394	-0.2265	-0.3705	0.0129
	S.E. Alpha	*(0.0416)*	*(0.0443)*	*(0.0753)*	*(0.0673)*	*(0.1061)*	*(0.1081)*	*(0.0045)*
	EMR	1.0003	0.9531	0.9228	1.0800	1.1279	0.1276	0.9953
	S.E. EMR	*(0.0099)*	*(0.0106)*	*(0.0180)*	*(0.0161)*	*(0.0254)*	*(0.0258)*	*(0.0011)*
	SMB	-0.0728	-0.2261	-0.1354	-0.0904	0.2184	0.2913	-0.0126
	S.E. SMB	*(0.0138)*	*(0.0147)*	*(0.0251)*	*(0.0224)*	*(0.0353)*	*(0.0360)*	*(0.0015)*
	HML	-0.0374	0.0011	-0.0769	0.0658	-0.0353	0.0020	-0.0042
	S.E. HML	*(0.0149)*	*(0.0159)*	*(0.0271)*	*(0.0242)*	*(0.0381)*	*(0.0389)*	*(0.0016)*
	S.D. Resid	*1.0187*	*1.0854*	*1.8449*	*1.6491*	*2.5976*	*2.6482*	*0.1111*

For each model on the left, the monthly average return and alpha values for each quintile are shown. The standard deviation and standard error are shown for each under each number, respectively. In this table, higher quintiles represent lower idiosyncratic risk values. To the right of the five quintiles, the results are shown for the Q5 - Q1 trading strategy and the full portfolio.

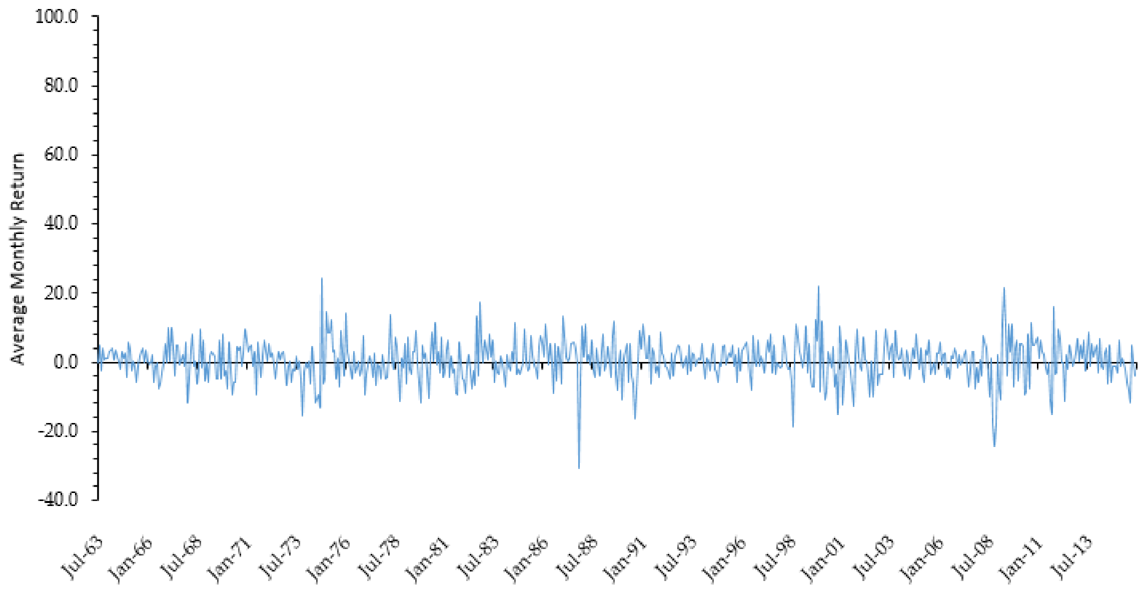
5.4.3 Disecting Average Returns

Figure 3: 5th Quintile over Time

Panel A: EW PCA (98%)



Panel B: VW PCA (98%)



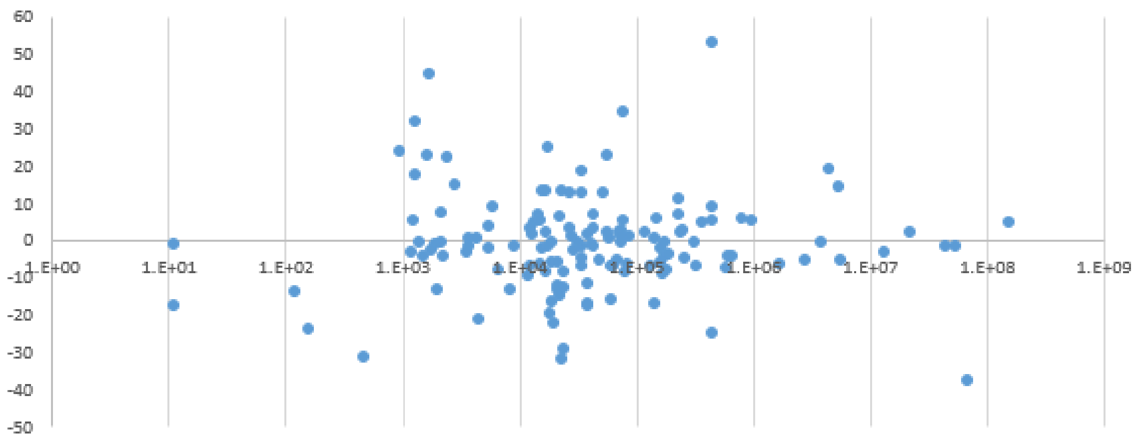
Shows the average annualised returns of the 5th quintile over time for both the EW PCA98 and VW PCA98 model in Panel A and B, respectively.

Figure 3 shows the average returns of the 5th quintile over time: Panel A shows the average monthly returns for the EW portfolio and Panel B shows the average monthly returns for the VW portfolio. The average returns per month for the EW PCA 98% portfolio in Panel A are far less stable than those of the VW PCA 98% portfolio shown in Panel B. Take for example the month of February 2000 which falls just before the so-called dot-com bubble of 2000 and 2002 and shows an average monthly return of 61%, or 734% annualised, over the cross section of low idiosyncratic stocks for the EW portfolio. Although still higher than the average month, in general, the VW portfolio produces an average monthly return of just under 12%, or 141% annualised.

However, we do see several extreme troughs (negative spikes) in years such as 1987, 2000 and 2008 for both portfolios. The effects of the October 1987 stock market crash, for instance, are clearly visible for both portfolios: -33% for the EW portfolio and -31% for the VW portfolio. Still, the fact that the two portfolios do not differ much during this month is not surprising as all firms, large and small, suffered large losses during the month of October 1987, or ”Black Monday”.

For months such as February 2000, we are interested in whether the difference in average monthly return between the EW and VW porfolios is due to one or a set of companies with explosive excess return values, and if these companies have small or large market caps. Figure 4 shows the cross section of monthly returns of companies in February 2000. Note that the x-axis is shown on a logarithmic scale. What we can see from Figure 4 is that there are several very small firms showing quite extreme monthly average returns, both positive and negative. The visual sample lends further evidence to the idea that several small market cap firms, exhibiting extreme average return values, heavily influence our EW portfolio results. In particular, Figure 4 allows us to induce that that the spike seen in Q5 of Table 6 for the EW portfolio can be attributed to small firms, with low idiosyncratic risk values and relatively high average monthly returns.

Figure 4: 5$^{\text{th}}$ Quintile - February 2000: PCA (98%)



Shows the annualised return values of the 20% lowest idiosyncratic risk values of firms for the month of February 2000. The Y-axis depicts average monthly returns as a percentage, and the X-axis displays market capitalisation in thousands of US$, shown on a logarithmic scale.

5.5 Double Sort Portfolios

The evidence so far using single sort portfolios suggests that small firms with low idiosyncratic risk produce relatively high returns and alpha values. In order to better access the relation between size, idiosyncratic risk, and the effect these have on returns we perform a double sort portfolio. In doing so, we first sort the monthly stock returns of the 3FF model into five quintiles according to size and, second, sort each of these portfolios into 5 sub-portfolios according to the IR values obtained using the 3FF model. We then repeat this exercise for the PCA98 model. The results for the EW and VW 3FF double sort portfolio are shown in Table 8 and Table 9, and the results for the EW and VW PCA98 double sort portfolio are shown in Table 10 and Table 11.

In analysing double sort portfolios, when we look at one quintile of the first sort, market capitalisation, and then look across the quintiles of the second sort, idiosyncratic risk, we look at whether the alpha values shown are stable or whether they fluctuate. If they are, for instance, stable then market capitalisation sufficiently captures the full effect on average returns and there is no longer enough variation in the data to sort in terms of the second criteria, idiosyncratic risk. For both models and both weighting methods, if we look at the 5th quintile of the first sort shown by the last two rows each Table 8, we see that the alpha values across quintiles of idiosyncratic risk remain relatively stable. What this tells us is that for large market capitalisation firms, average returns are not substantially influenced by different levels of idiosyncratic risk.

However, as we gradually increase market capitalisation, for both models and methods of weighting, we that the alpha values reflect a negative relation between average returns and idiosyncratic risk. In the most extreme case, the lowest quintile of market capitalisation, we see the same large spike in alpha values for the 5th quintile containing the lowest 20% of idiosyncratic risk values as we saw in the equally weighted portfolios. Clearly, the negative relation identified in previous sections is attributable to low market capitalisation firms where this effect gradually subsides for increasingly larger category of market capitalisation firms.

Comparing the EW and VW weighting methods for each model, there are no major differences in terms of alphas values and standard errors. This contrasts to the single sort portfolio in that the value weighted portfolio did not exhibit the same spike in the 5th quintile that the single sort equally-weighted portfolio did. This makes sense as, for the single sort value-weighted portfolio, less weight is placed on small capitalisation firms with high average returns and alpha values. The double sort exercise provides further evidence that there are several small capitalisation firms with low idiosyncratic risk values which provides high average returns and alpha values.

Table 8: Equally-weighted 3FF Double Sort: (i) Size and (ii) Idiosyncratic Risk

		Idiosyncratic Risk					
		Q1 = High	Q2	Q3	Q4	Q5 = Low	Q5 - Q1
Market Captalisation	Q1 = Low	-1.1893	-0.9940	-0.2870	0.9247	7.5551	8.7444
		(0.0929)	*(0.1147)*	*(0.1398)*	*(0.1769)*	*(0.3541)*	*(0.3542)*
	Q2	-0.8321	-0.7941	-0.5065	0.2527	2.2543	3.0864
		(0.0649)	*(0.0795)*	*(0.0851)*	*(0.1051)*	*(0.2071)*	*(0.2379)*
	Q3	-0.4493	-0.3666	-0.0741	0.2492	0.9402	1.3895
		(0.0613)	*(0.0669)*	*(0.0695)*	*(0.0714)*	*(0.1663)*	*(0.2031)*
	Q4	-0.2884	-0.1596	0.0110	0.2641	0.2794	0.5678
		(0.0593)	*(0.0632)*	*(0.0591)*	*(0.0669)*	*(0.1329)*	*(0.1687)*
	Q5 = High	-0.1085	0.0029	0.0801	0.1359	-0.1059	0.0026
		(0.0567)	*(0.0542)*	*(0.0505)*	*(0.0542)*	*(0.1253)*	*(0.1650)*

Shows the EW 3FF model sorted for (i) size and (ii) idiosyncratic risk, split into 5x5 quintiles. For each quintile, the top value denotes the alpha of the portfolio and the value below denotes the standard error of the same. The last column on the right depicts the Q5 - Q1 trading strategy.

Table 9: Value-weighted 3FF Double Sort: (i) Size and (ii) Idiosyncratic Risk

		Idiosyncratic Risk					
		Q1 = High	Q2	Q3	Q4	Q5 = Low	Q5 - Q1
Market Captalisation	Q1 = Low	-1.1729	-0.9333	-0.1927	0.9588	6.6093	7.7822
		(0.0912)	*(0.1122)*	*(0.1346)*	*(0.1749)*	*(0.3566)*	*(0.3641)*
	Q2	-0.8172	-0.7435	-0.4367	0.2835	2.1566	2.9738
		(0.0652)	*(0.0800)*	*(0.0865)*	*(0.1051)*	*(0.2066)*	*(0.2391)*
	Q3	-0.4292	-0.3165	-0.0404	0.2876	0.8861	1.3153
		(0.0615)	*(0.0669)*	*(0.0705)*	*(0.0725)*	*(0.1651)*	*(0.2017)*
	Q4	-0.2780	-0.1315	0.0468	0.2875	0.2156	0.4936
		(0.0594)	*(0.0640)*	*(0.0605)*	*(0.0681)*	*(0.1358)*	*(0.1705)*
	Q5 = High	0.0869	0.0656	0.1048	0.0851	-0.4247	-0.5116
		(0.0505)	*(0.0534)*	*(0.0487)*	*(0.0642)*	*(0.1437)*	*(0.1789)*

Shows the VW 3FF model sorted for (i) size and (ii) idiosyncratic risk, split into 5x5 quintiles. For each quintile, the top value denotes the alpha of the portfolio and the value below denotes the standard error of the same. The last column on the right depicts the Q5 - Q1 trading strategy.

Table 10: Equally-weighted PCA (98%) Double Sort: (i) Size and (ii) Idiosyncratic Risk

		Idiosyncratic Risk					
		Q1 = High	Q2	Q3	Q4	Q5 = Low	Q5 - Q1
Market Captalisation	Q1 = Low	-1.2311	-0.9922	-0.2781	1.0189	7.4903	8.7214
		(0.0994)	*(0.1201)*	*(0.1427)*	*(0.1750)*	*(0.3507)*	*(0.3553)*
	Q2	-1.0015	-0.8384	-0.4974	0.2218	2.4854	3.4869
		(0.0768)	*(0.0755)*	*(0.0850)*	*(0.1005)*	*(0.2081)*	*(0.2454)*
	Q3	-0.5983	-0.4073	-0.1889	0.1790	1.3159	1.9142
		(0.0689)	*(0.0645)*	*(0.0653)*	*(0.0788)*	*(0.1630)*	*(0.2061)*
	Q4	-0.4092	-0.2554	-0.0246	0.2114	0.5857	0.9949
		(0.0671)	*(0.0647)*	*(0.0574)*	*(0.0638)*	*(0.1398)*	*(0.1835)*
	Q5 = High	-0.2355	-0.1182	-0.0048	0.1345	0.2303	0.4658
		(0.0617)	*(0.0562)*	*(0.0495)*	*(0.0532)*	*(0.1280)*	*(0.1716)*

Shows the EW PCA98 model sorted for (i) size and (ii) idiosyncratic risk, split into 5x5 quintiles. For each quintile, the top value denotes the alpha of the portfolio and the value below denotes the standard error of the same. The last column on the right depicts the Q5 - Q1 trading strategy.

Table 11: Value-weighted PCA (98%) Double Sort: (i) Size and (ii) Idiosyncratic Risk

		Idiosyncratic Risk					
		Q1 = High	Q2	Q3	Q4	Q5 = Low	Q5 - Q1
Market Captalisation	Q1 = Low	-1.2160	-0.9469	-0.2505	1.0093	6.4800	7.6960
		(0.0968)	*(0.1158)*	*(0.1364)*	*(0.1717)*	*(0.3479)*	*(0.3600)*
	Q2	-0.9689	-0.8190	-0.4348	0.2554	2.3979	3.3667
		(0.0772)	*(0.0759)*	*(0.0854)*	*(0.1003)*	*(0.2080)*	*(0.2466)*
	Q3	-0.5763	-0.3468	-0.1574	0.2030	1.2800	1.8563
		(0.0687)	*(0.0642)*	*(0.0669)*	*(0.0786)*	*(0.1622)*	*(0.2050)*
	Q4	-0.3910	-0.2380	0.0172	0.2219	0.5415	0.9326
		(0.0661)	*(0.0649)*	*(0.0588)*	*(0.0653)*	*(0.1421)*	*(0.1837)*
	Q5 = High	-0.1369	0.0400	0.0392	0.1544	0.0364	0.1733
		(0.0606)	*(0.0510)*	*(0.0466)*	*(0.0631)*	*(0.1378)*	*(0.1836)*

Shows the VW PCA98 model sorted for (i) size and (ii) idiosyncratic risk, split into 5x5 quintiles. For each quintile, the top value denotes the alpha of the portfolio and the value below denotes the standard error of the same. The last column on the right depicts the Q5 - Q1 trading strategy.

The double sort portfolios shown in Tables 8 through 11 serve as an extension to our analysis of single sort portfolios in the previous section. The two main key findings are the following: (i) when first sorting for size and second for idiosyncratic risk the difference between the EW and VW results subsides as compared to the difference between EW and VW single sort results. What this tells us, is that because we have first sorted for size, the variation within each of these quintiles is limited in terms of firm size. Secondly, (ii) we have seen that for the EW and VW single sort portfolios, a spike exists when moving from Q4 to Q5, the latter portfolio containing firms with the lowest 20% idiosyncratic risk values. We have further established that, when first sorting for size and second for idiosyncratic risk using double sort portfolios that the variation within each first sort is limited in terms of market size i.e. that the firm sizes within each quintile are quite similar. We then saw that, for the 80% largest firms, this spike was either subdued or non-existent. However, the quintile containing the smallest 20% firms in terms of size showed a substantial spike when moving from the fourth to the 5^{th} quintile of idiosyncratic risk values, the latter containing firms with the lowest 20% in terms of idiosyncratic risk. What we can conclude from this section is that firms size and idiosyncratic risk are related in that there are a handful of small firms with high idiosyncratic risk values, however, this relation does not hold for the full data set where we do not see a clear relation between firm size and idiosyncratic risk.

6 Conclusion

The aim of this paper is to replicate and further scrutinise the findings of papers such as Ang et al. (2006) and (2009). Ang et al. (2009) find a significant negative relation between idiosyncratic risk and return while theory tells us that idiosyncratic risk should be inherently unpredictable by nature. We apply the CAPM and Fama-French three factor model as well as a PCA factor model to equally- and value-weighted datasets. These models are used within a Fama-MacBeth regression and are compared to one another. Next, we apply single and double sorted portfolios in terms of idiosyncratic risk in order to build on our findings of the Fama-MacBeth regressions, to identify any fluctuations in the relation between idiosyncratic risk and return, and to make observations on how size skews our perception of this relation.

We find that the CAPM and the three factor Fama-French model are equivalent in terms of the insights they provide when isolating idiosyncratic risk and describing its relation with returns. The PCA factor models clearly result in much smaller residuals resulting in far smaller idiosyncratic risk values in general, illustrated by the distributions in Figure 2. The models which use the differences in residuals to find idiosyncratic risk are not very different from the standard three factor Fama-French model, showing us that the idiosyncratic risk values of the PCA factor model are small and independent relative to the Fama-French factor models.

We are able to replicate the results found by Ang et al. (2009) in that the Fama-MacBeth regressions produce a significant negative relation between idiosyncratic risk and returns when using a three factor Fama-French model and an equally-weighted dataset. The PCA factor models produce an insignificant relation between IR values and returns using a 98% threshold of explained variance. This falls in line with our expectation that idiosyncratic risk values found using PCA analysis should approach the theoretical values of idiosyncratic risk which, in turn, should inherently be unpredictable by nature. While we observe a similar negative relation between idiosyncratic risk and returns, as found by Ang et al. (2009), using a three factor Fama-French model within a Fama-MacBeth framework, this tells us very little about the monotonicity of the relation. In order to do, we provide an extensive analysis using single and double sort portfolios.

The implications of the results stemming from our sorted portfolios are intriguing: For both the equally-weighted single sort portfolios, we see a relatively flat trend over the first four out of five quintiles of idiosyncratic risk values with slight, but minor, fluctuations over the highest 80% of idiosyncratic risk values. We do, however, see a large spike in the last quintile representing the lowest 20% of idiosyncratic risk values. This finding holds for all models tested including the three factor Fama-French model and the PCA factor model. When comparing these results to those of the value-weighted single sort portfolio, we see that the spike witnessed in the 5^{th} quintile of the equally-weighted single sort portfolio disappears as the small capitalisation firms receive less weight using a value-weighted method. These single-sort portfolios tell us that the negative relation observed between idiosyncratic risk and returns using a Fama-MacBeth regression is heavily concentrated in the 5^{th} quintile of these portfolio sorts. In other words, 80% of the dataset fails to show a clear negative relation between idiosyncratic risk and returns while the last 20% containing the lowest idiosyncratic risk values does. Then, when portfolio returns are value-weighted as opposed to equally-weighted, the strong negative relation witnessed in the 5^{th} disappears suggesting the negative relation observed between idiosyncratic risk and returns could be influenced by a number of small firms with low idiosyncratic risk values and relatively high returns.

After zooming in on the 5^{th} quintile of the PCA (98%) factor model over time, we see that the enormous annualised returns observed for the equally-weighted models can be attributed to a higher variation in the average level of portfolio returns as well as to large spikes at particular points in time such as the one identified in February 2000 right before the dot com bubble. When scrutinising February 2000 and looking at the returns of the firms with the lowest 20% idiosyncratic risk values, we again find that these results are not due to the entire sample set of companies all exhibiting large returns over the month of February 2000, but rather because there are several small-sized firms that skew this average due to their excessively high return values.

Our double sort portfolios sort company returns first by market capitalisation and second by idiosyncratic risk. The results of the equally-weighted and value-weighted double sort portfolios are similar telling us that there is not much variation in terms of size within each quintile sorted by market capitalisation. The 5^{th} quintile spike seen for the single sort portfolios is only witnessed for the first quintile containing the smallest 20% of firms in terms of size telling us that size and idiosyncratic risk do not depict a clear relation for the majority of the dataset, with the exception of the negative relation reflected by the spike in the 5th quintile containing small-sized firms with low idiosyncratic risk values.

Ang et al. (2009) find a negative relation between idiosyncratic risk and returns when controlling for size in their Fama-Macbeth regressions. This relation, however, is over the full portfolio of stocks and does not mean a size effect is no longer present. Through the use of double portfolio sorts, we find that the effect of firm size is not the same across all small firms and that the negative effect witnessed by Ang et al. (2009) can be largely attributed to very small firms with low idiosyncratic risk values which produce relatively high average returns.

In conclusion, this paper contributes to the literature by providing substantial evidence that the negative relation between idiosyncratic risk and returns found by recent papers including those of Ang et al. (2006) and (2009) is due large in part by a small set of small-sized firms with low idiosyncratic risk values and high average returns as opposed to an inherent negative relation supporting a case of an Idiosyncratic Risk Puzzle. In other words, while Ang et al. (2009) may have found a negative relation using Fama-MacBeth regressions, our portfolio sort exercise has uncovered a large spike in the lowest quintile of idiosyncratic risk for small-sized firms and found very little evidence of a meaningful relation in the remaining 80% of the dataset. This paper also demonstrates that the conclusions are similar when using a PCA (98%) model to isolate idiosyncratic risk instead of a Fama-French factor model.

While these findings appear robust for both the three factor Fama-French model as well as when using a PCA approach, it is important to acknowledge that the conclusions drawn in this paper have certain limitations. This paper uses Ang et al. (2006, 2009) as a foundation for its methodological framework, yet it restricts itself to the U.S. stock market. While Ang et al. (2009) conclude that the puzzle of low returns to high idiosyncratic volatility stocks is a global phenomenon, Koch et al. (2009) finds that the size effect is not pronounced in Germany, implying that double sorting would not lead to the same conclusions in Germany as in this paper.

The hope is that this paper will help provide insight to future researchers on the subject of idiosyncratic risk. More specifically, we hope to have given future researchers insight into the relation between size and idiosyncratic risk and, in turn, their effect on returns. Our analysis of the 5th quintile of idiosyncratic risk values could serve as a basis for further research which might seek to explain this seemingly massive spike in average returns for low idiosyncratic risk value companies. Several papers such as that written by Spiegel and Wang (2006) scrutinise the relation between liquidity, idiosyncratic risk and returns. We believe a further analysis of the relation between liquidty, idiosyncratic risk and returns within a double-sort framework would be an interesting avenue for further research.

Acknowledgements

This paper would not have been possible without the patience and guidance of Erik Kole and Karolina Scholtus. Having written the bulk of this thesis while working a full-time job, this paper needed a little more care and attention than Erik might be used to and, for that, I am grateful. I'd also like to say thank you to my parents for believing that even pigs can fly. They will be delighted to know that this here ends my academic career.

References

Ang, A., Chen, J., and Xing, Y. (2005). Downside Risk. (212).

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *Journal of Finance*, LXI(1):259–299.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence. *Journal of Financial Economics*, 91(1):1–23.

Ang, A., Liu, J., and Schwarz, K. (2010). Using Stocks or Portfolios in Tests of Factor Models. *Unpublished Working Paper, . . .* , (212):1–33.

Bali, T., Cakici, N., Yan, X., and Zhang, Z. (2005). Does Idiosyncratic Risk Really Matter? *Journal of Finance*, 60(2):905–929.

Brandt, M. W., Brav, A., Graham, J. R., and Kumar, A. (2010). The idiosyncratic volatility puzzle: Time trend or speculative episodes. *Review of Financial Studies*, 23(2):863–899.

Camphell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, 56(1):1.

Chen, Z. and Petkova, R. (2012). Does Idiosyncratic Volatility Proxy for Risk Exposure? *The Review of Financial Studies*, 25(9):2745–2787.

Cochrane, J. H. (2005). *Asset Pricing*.

Fama, E. F. and French, K. R. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, 2(47):427–465.

Fama, E. F. and 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. and MacBeth, J. D. (1973). Risk , Return , and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636.

Fu, F. (2009). Idiosyncratic Risk and the Cross-section of Expected Stock Returns. *Journal of Financial Economics*, 91(1):24–37.

Goyal, A. and Santa-Clara, P. (2003). Idiosyncratic Risk Matters. *Journal of Finance*, 58(3):975–1007.

Harry Markowitz (1952). Portfolio Selection. *Journal of Finance*, 7(1):77–91.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., and van Dijk, H. K. (2004). *Econometric Methods With Applications In Business And Economics*.

Herskovic, B., Kelly, B., Lustig, H., and Nieuwerburgh, S. V. (2015). The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications. *Journal of Financial Economics*, 119(2):1–53.

Kleibergen, F. (2010). Reality Checks for and of Factor Pricing. *UvA Econometrics*.

Kleibergen, F. and Zhan, Z. (2015). Unexplained Factors and their Effects on Second Pass R-squared's. *Journal of Econometrics*, 189(1):101–116.

Koch, S. (2009). The Idiosyncratic Risk Puzzle: Evidence from the German Stock Market.

Lehmann, B. N. (1990). Residual Risk Revisited. *Journal of Econometrics*, 45(1-2):71–97.

Malagon, J., Moreno, D., Rodríguez, R., Malagon, J., Moreno, D., and Rodríguez, R. (2012). Time Horizon Trading and the Idiosyncratic Risk Puzzle. *Quantitative Finance*, 15(2):327–343.

Pastor, L. and Stambaugh, R. F. (2001). Liquidity Risk and Expected Stock Returns.

Patton, A. J. and Timmermann, A. (2007). Portfolio Sorts and Tests of Cross-Sectional Patterns in Expected Returns. *Working Paper*, pages 1–44.

Patton, A. J. and Timmermann, A. (2010). Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts. *Journal of Financial Economics*, 98(3):605–625.

Plyakha, Y., Uppal, R., and Vilkov, G. (2014). Equal or Value Weighting? Implications for Asset-Pricing Tests. pages 1–64.

Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. *Management Science*, 9(2):277–293.

Spiegel, M. and Wang, X. (2005). Cross-sectional variation in stock returns: Liquidity and idiosyncratic risk. *Working Paper, Yale University*, pages 1–49.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*.

Xu, Y. and Malkiel, B. G. (2003). Investigating the Behavior of Idiosyncratic Volatility. *The Journal of Business*, 76(4):613–645.