THE RELATION BETWEEN IDIOSYNCRATIC VOLATILITY AND RETURNS FOR U.S. MUTUAL FUNDS

Submitted by

Kevin Skogström Lundgren[1]

Department of Economics

In partial fulfilment of the requirements

For the Degree of Finance: Master of Science

Lund University

Lund, Sweden

January 2016

Advisor: Frederik Lundtofte

[1] kevin.skogstrom@gmail.com

Abstract

Theoretically the relation between returns and idiosyncratic volatility should be non-existent or positive. Many empirical studies confirm this but Ang, Hodrick, Xing and Zhang (2006) contest the conventional view and find a negative relationship for a sample of U.S. firms. I contribute to the field by investigating the relation for a sample of U.S. mutual funds. The sample consist of a total of 10 917 equity mutual funds, the funds are divided in to four different classes depending on equity focus (growth, value, small cap and large cap). Data were collected for the period 1995 to 2015. Ang et al. (2006) relate returns with lagged idiosyncratic volatility making the implicit assumption that idiosyncratic volatility can be described as a random walk. But as Fu (2009) I find that this is not true and use an AR(2) model to estimate idiosyncratic volatility, inspired by Chua, Choong Tze, Jeremy Goh, and Zhe Zhang (2010). The idiosyncratic volatility is estimated relative to the Carhart (1997) four-factor mode and divided in to an expected and unexpected part as in Chua et al. The relation is examined using Fama-MacBeth (1973) regressions with both gross return and the Carhart alpha as dependent variables. The results suggest a positive relation only when using the Carhart alpha as dependent variable and all the control variables. Otherwise the results are inconclusive. As an additional robustness test I perform portfolio sorting on EIV without control variables and get results that suggest a negative relation.

Keywords: Idiosyncratic volatility, Mutual funds, Carhart four-factor model, ARIMA model, Carhart four-factor alpha

Table of Contents

1. Introduction

According to standard asset pricing theory most notably the CAPM (capital asset pricing model) investors hold diversified portfolios, i.e. the idiosyncratic volatility has been diversified away. It can be proven that no investor will be rewarded for taking on idiosyncratic risk. Only systematic risk is rewarded in a world where idiosyncratic volatility can successfully be diversified away. However, the assumption that investors can hold diversified portfolios may not be realistic. Theoretical models extending the CAPM model such as Merton (1987), Levy (1978) and Malkiel & Xu (2002) assumes that investors may not be able to hold well diversified portfolios and thus they will be compensated for taking on idiosyncratic volatility. Further, a number of studies document this positive relationship empirically (e.g. Fu, 2009, Eiling, 2013 and Chua, Goh & Zhang, 2010).

Ang, Hodrick, Xing and Zang (2006) contest this by showing that lagged idiosyncratic volatility is negatively related with returns for a sample of U.S. firms. Their findings have not come without criticism, a notable example is Fu (2009). He criticises Ang et al.'s use of lagged idiosyncratic volatility since it does not follow a random walk according to him, which is an implicit assumption of Ang et al. (2006). There are however studies that confirm Ang et al. (2006) and more studies that contest them.

The question that I pose in this study is: What is the relation between conditional idiosyncratic volatility and return for mutual funds?

The purpose of this study is to extend the research in the field to also cover mutual funds. Given the appeal of Fu (2009) criticism of Ang et al. (2006) I will follow a similar approach developed by Chua, Goh & Zhang (2010) closely. Chua et al. (2010) estimate their conditional idiosyncratic risk using an ARIMA type model making it easier to estimate conditional idiosyncratic volatility than Fu's GARCH type model. The study has solely focused on open-ended equity mutual funds registered in the U.S. and with a focus on U.S. equity covering the years 1995 to 2015. I hope to contribute to the research field by giving a better understanding of how idiosyncratic volatility relates to return/performance for mutual funds. The mutual fund industry is by no doubt a big industry and being able to better understand risk-pricing properties for that industry would indeed valuable. My study differs from Chua et al. in a number of ways; most important is of course the fact that I use mutual funds and not firms. I also use a different factor model for the estimation of idiosyncratic volatility, namely the Carhart four-factor model while Chua et al. use the Fama-French three-factor model. When investigating the relation between idiosyncratic volatility and return through Fama-MacBeth regressions I use both gross returns and risk-adjusted returns (Carhart four-factor alpha) this is different from Chua et al. and other studies employing Fama-MacBeth regressions.

All the models in this study have been estimated using Stata version 13. Stata codes are available from the author upon request.

The remainder of this study is outlined as follows. In Chapter 2 I present a theoretical framework from which the role and estimation of idiosyncratic volatility will be better understood. In Chapter 3 I discuss previous research in the area. In Chapter 4 I present the methodologies used to study the relationship empirically. Chapter 5 presents the data and how variables have been constructed. Chapter 6 presents my results and an analysis of them. Chapter 7 conclude my main findings and also provide suggestions for future research.

2. Theoretical Background

This chapter present the theory on which the study relies on. I first introduce the CAPM model and explain why idiosyncratic risk should not be priced in a CAPM world. I then present an extension allowing for idiosyncratic risk to be priced. I then present two pricing models that has performed better in studies than CAPM (Fama-French three-factor model and Carhart four-factor) and motivate my choice of the Carhart model to estimate idiosyncratic risk.

2.1 CAPM (Capital Asset Pricing Model)

CAPM was developed independently by Sharpe (1964) and Lintner (1965). The two most important assumptions underpinning the model is that i) the investors has mean-variance preferences and ii) investors share homogenous beliefs. The mean-variance preferences stems from Markowitz (1952) article about portfolio selection. It is an assumption that an investor cares only about the variance (risk) and the expected returns of an asset when selecting a portfolio. Given a set of assets that can be selected by the investor and assumptions i) and ii) we arrive at the Sharpe-Lintner version of the CAPM after a series of mathematical derivations:

$$E\left[\tilde{r}_q\right] = r_f + (E[\tilde{r}_m] - r_f)\frac{\sigma_{qm}}{\sigma_m^2} \tag{1}$$

Where, $E\left[\tilde{r}_q\right]$ = e*xpected return of asset q*, r_f = *risk-free rate*, $(E[\tilde{r}_m] - r_f)$ = *expected market excess return*, σ_{qm} = *covariance between asset q and the market portfolio m* and σ_m^2 = *variance of the market portfolio*.

One of the most important consequences of the CAPM for this study is that risk can be divided in to two parts, systematic and idiosyncratic volatility (risk). Systematic risk in the above equation is represented by σ_{qm} and the higher the systematic risk the higher the expected returns. But there is no term that represents the idiosyncratic volatility. Hence idiosyncratic volatility is not priced. Idiosyncratic volatility can be such things as the risk that the CEO of a company dies and diversifying will practically eliminate the effect of this type of risk, which is intuitive. Adding an asset with idiosyncratic volatility to an already well-diversified portfolio means that the idiosyncratic volatility is diversified away. Hence adding such an asset to a well-diversified portfolio does not generate a higher expected return.

Although adding an asset that correlates highly with the market portfolio raise the risk of the whole portfolio and thus the asset should be rewarded with a higher expected return.

2.2 Extending CAPM – Allowing for a Positive Relation Between Idiosyncratic Volatility and Return

Merton (1987) presents a theoretical motivation for the existence of a positive relation between idiosyncratic volatility and returns. His model is an extension of the normal CAPM model but with the additional assumption of incomplete information in the financial markets. What this basically means is that investors only have sufficient information about a limited number of stocks. The main assumption poised by Merton is that a stock will only be included in the investors' portfolio if he has "enough" information about the stock. Since it is assumed that investors do not have information about all the stocks in the market this must mean that investors in general are under-diversified. The main motivation for this assumption is the fact that real world investors only have a limited number of stocks in their portfolios even though they can invest in considerably more financial assets. Merton´s version of the CAPM can be stated as follows:

$$E[\tilde{r}_i] = r_f + \left(E[\tilde{r}_m] - r_f\right)\frac{\sigma_{im}}{\sigma_m^2} + \lambda\left(\frac{x_i}{q_i}\right)\sigma_i^2 \qquad (2)$$

The first two terms on the right hand side corresponds to the normal CAPM (see eq. 1). λ is the cost of risk, x_i is the ratio of the stock's value to the whole market, q_i is the fraction of the investor population who have information about asset i and σ_i^2 is the idiosyncratic volatility of stock *i.* From the equation it is obvious that idiosyncratic risk can affect the expected return of an asset positively.

Other theoretical studies that suggest a positive relationship are Levy (1978) and Malkiel and Xu (2002). These articles will however not be covered in this study.

2.3 Fama-French Three-Factor Model

The Fama-French three-factor model, henceforth FF3F (Fama & French, 1993; Fama & French, 1995 and Fama & French 1996) is a pricing model that contests the CAPM. The model has its background in Fama & French (1992) where different variables were tested to see which ones could describe the cross-section of returns the best. They discovered that market equity (ME) and book-to-market ratio described the cross-section of returns the best, using a sample of U.S. non-financial firms. Fama & French (1993) concludes that factors based on ME and book-to-market ratio is best used in conjunction with the market factor. The models is stated as:

$$E[r_i] - r_f = b_i\left(E[r_m] - r_f\right) + s_i(E[SMB]) + h_i(E[HML]) \qquad (3)$$

The model basically says that the excess return of a stock is equal to its sensitivity to three factors. $\left(E[r_m] - r_f\right)$ is the market factor and is defined as the expected return of broad market index minus the risk-free rate, $(E[SMB])$ is the so called SMB (Small Minus Big) factor and is defined as the difference between a portfolio of small capital stocks and a portfolio of large capital stock. The SMB factor is thus based on the notion that stocks with high market equity perform worse than stocks with low market equity. The last factor $(E[HML])$ is the HML (High Minus Low) factor and is defined as the difference between a portfolio consisting of high book-to-market stocks and a portfolio consisting of low book-to-market stocks. The HML factor construction comes from the idea that value stocks (low book-to-market ratio) should have a higher return compared to growth stocks (high book-to-market ratio).

It is easy to see that FF3F equation has the same form as the CAPM although with two additional factors. So it means that the systematic risk is not just explained by the market factor but also by the SMB and HML factor. No asset pricing theory was used as a foundation when developing the FF3F model but rather just empirical observations that ME and book-to-market improves the explanation of expected returns. Fama & French (1996) gives three suggestions of interpretations of their model. The first is that the FF3F model adheres to a rational multifactor asset-pricing model. The second is that only investor irrationality separates the FF3F model as better than CAPM. Finally that the CAPM do hold but data issues result in the FF3F being more suitable. If we instead focus on multifactor asset-pricing models as a justification for the model there are two theories of interest; Intertemporal Capital Asset Pricing Model (ICAPM) developed by Merton (1973) and Arbitrage Pricing Theory (APT) developed by Ross (1976). In short the ICAPM is the continuous time version of the CAPM and allows for multiple factors to explain the expected returns. The model is centred on the idea that investment opportunities change and that investors may want to change their holdings in light of changing opportunities, which the static CAPM does not allow. The normal CAPM model is actually a special case of the more general ICAPM. In ICAPM the only fixed factor is the market factor as in CAPM, but investors are also compensated for risk in state variables (such as GDP etc.). The assumptions of the ICAPM differs from the CAPM on several points most notably portfolios need not to be well diversified and formally there are not assumption on homogenous beliefs, although Merton (1973) points out that investors “principally” are homogenous in beliefs. The model is consistent with expected utility maximization although not in form mean-variance analysis as in CAPM. The assumption of equilibrium markets is also prevalent in ICAPM.

APT derives an asset pricing structure similar to the ICAPM although its underlying theoretical framework is different from that of both the CAPM and ICAPM. The APT is centred on the idea of arbitrage; if expected price diverge from the actual the market should correct this erasing any arbitrage opportunities. As such it is not an equilibrium theory as ICAPM and CAPM. The theory requires investors to have homogeneous beliefs about the expected returns (Ross 1973).

By assuming that the returns follow a discrete-time version of the ICAPM Petkova (2006) suggest that the FF3F model can be thought of as version of ICAPM. Even if it is not apparent what theoretical asset pricing foundation the FF3F model stands on it should be apparent that because it is formulated as a linear model similar to CAPM the idiosyncratic risks should not be priced in this setting either.

2.4 Carhart Model

Another asset pricing model that can be viewed as an extension of the FF3F model is the Carhart model introduced by Carhart (1997). The Carhart model in addition to the factors in the FF3F model contains a momentum factor (MOM). The momentum factor is constructed as the difference of a portfolio consisting of assets with historical high return and a portfolio of assets with a historical low return. The logic behind the MOM factor is the notion of mutual fund persistence, i.e. funds that have performed well in the past will continue to do so in the future. To my best recollection there have been no studies investigating whether the Carhart model can be described by an asset-pricing model such as the ICAPM or APT. But since the model is to large part similar to FF3F model the earlier discussion on the FF3F should be similar for the Carhart model.

Because of multifactor models better performance than the CAPM model and the commonality to use the Carhart model when pricing mutual funds I have chosen to estimate the idiosyncratic risk relative to the Carhart model. Previous studies in this field mostly use the Fama-French three-factor model for extracting the idiosyncratic volatility.

3. Literature Review

In this chapter I present and discuss previous studies that have investigated the relation between idiosyncratic risk and returns. I have focused on studies published in leading academic journal with the exception of the study I replicate and a couple of studies that have partially investigated the problem using mutual funds as a sample.

As have been discussed in the introduction Ang et al. (2006) discovered that idiosyncratic risk is negatively related with returns. They use a sample consisting of U.S. firms listed on NYSE, NASDAQ and AMEX during the time period July 1963 to December 2000 and investigate the problem using a portfolio approach. The approach works by first employing a trading strategy where idiosyncratic risk is based on daily data from the previous month then the stocks is sorted in to quintile portfolios based on the estimated idiosyncratic volatility and then kept for one month, the portfolios are value weighted. Idiosyncratic volatility is estimated relative the Fama-French (1993) three factor model and is defined as the standard deviation of the factor regression residuals in the past month. They use the three factor model and not the CAPM because the failure of the former in empirical studies. Their basic results show that the portfolios with high idiosyncratic risk have a lower return than portfolios with low idiosyncratic risk. To check the robustness of these results they control for size, book-to-market ratio, leverage, liquidity, volume, turnover, bid-ask spreads, coskewness and dispersion in analysts forecasts. Even after controlling for these variables the negative relation between idiosyncratic volatility and returns persist. In addition the authors look at momentum, exposure to aggregate volatility risk, different formation and holding periods (time window used to estimate idiosyncratic volatility and the length the portfolio is kept before being updated) and different time subsamples. After these additional robustness checks they still concludes that there is a negative relation between idiosyncratic volatility and returns.

Ang et al. (2009) investigate if the puzzling findings from their 2006 study can be extended to an international sample. They look at data from 23 developed countries during a time period stretching as far back as 1980 and up to 2003. In contrast to Ang et al. (2006) Fama-MacBeth type regressions are estimated instead of using a portfolio approach. The negative relation is confirmed in this study.

Stambaugh et al. (2015) provides support for the existence of a negative relationship using the concept of arbitrage risk and arbitrage asymmetry. Arbitrage risk is defined as the risk that "deters arbitrage", in other words the risk that investors won't take advantage of mispricing in financial assets. Arbitrage asymmetry on the other hand is the concept that buying (going long) a financial asset is easier than selling it (going short). Stambaugh et al. argues that arbitrage risk may be approximated by idiosyncratic volatility. Stocks with high idiosyncratic volatility thus posses a high arbitrage risk, this have the consequence that mispricing will be high for these assets. If there is over-pricing (under-pricing) of the asset then it must mean that expected returns will be low (high), hence there is a negative (positive) relation between idiosyncratic volatility and returns. Arbitrage asymmetry helps to establish the negative effect in aggregate sine it is easier to go long in stocks and hence under-pricing will be eliminated to a greater extent. They support their argument with empirical evidence using a sample of US stocks. They estimate mispricing by using 11 return anomalies such as financial distress. For each anomaly they construct a ranking on the stocks and then they take the average ranking percentile for each anomaly as the mispricing. Then by a portfolio approach they sort stocks on the level of mispricing and idiosyncratic volatility (relative the FF3F model).

Fu (2009) criticise Ang et al. (2006) use of lagged idiosyncratic volatility. He argues that using lagged idiosyncratic volatility as the next months expected idiosyncratic volatility is an implicit assumption that the idiosyncratic volatility follows a random walk (see the methodology for a deeper explanation). He test this by first graphically inspect autocorrelation at different lags and then performing a Dickey-Fuller test on the existence of a unit root in the idiosyncratic volatility series. He finds that idiosyncratic volatility cannot be explained by random walk and hence it is wrong to relate lagged idiosyncratic volatility with realised returns. Hence there is a need to estimate expected idiosyncratic risk; this was done using an EGARCH model. The mean equation in conjunction with the EGARCH model is the Fama-French three-factor model. The results from the Fama-MacBeth regressions are stark, idiosyncratic volatility is significantly positively related to returns both when excluding and including control variables. Further the lagged idiosyncratic risk is significantly negatively related with returns. By replicating Ang et al. (2006) in detail Fu argues that the negative relation comes from what is known as return reversals. High idiosyncratic volatilities and high returns tend to coincide and the return in the following month tends to reverse creating the appearance that high idiosyncratic stocks have low returns, when using lagged idiosyncratic volatility.

As in Fu (2009) Huang, Liu, Rhee & Zhang (2010) argue that return reversals are the explanation for the negative relation found in Ang et al. (2006). Not including the previous months return in the return-idiosyncratic volatility cross-sectional regressions will bias the coefficient for idiosyncratic volatility downwards. Further they argue that the method used to estimate idiosyncratic risk matter. Specifically idiosyncratic risk estimation based on daily data rather than monthly will enhance the bias. This is manifested in their cross-sectional regression results for a sample of U.S. stocks. For lagged ARIMA based idiosyncratic risk the coefficient switch sign from negative to positive after including the previous month returns. For EGARCH estimated idiosyncratic volatility the sign is positive with and without the previous months return although the value of the coefficient increases. Hence idiosyncratic volatility estimated from monthly data such as in the EGARCH specification do not experience such a severe bias.

Eiling (2013) also investigates the relation between returns and idiosyncratic volatility. Following a portfolio approach and employing an EGARCH model as in Fu (2009) to estimate conditional idiosyncratic risk she finds a positive relation between idiosyncratic risk and returns (risk adjusted returns). She argues that the positive relation is in part due to human-capital heterogeneity. The idea is that stocks with a high exposure to human capital should affect its returns. Because human capital is heterogeneous Eiling divides aggregated human capital in to several industry specific parts, hence getting industry specific human capital. She then extends the CAPM including the industry specific human capital variables defined as labour income growth rates. The model is estimated for a number of portfolios formed on size and book-to-market ratio. She finds that the model including industry specific capital performs much better than the standard CAPM. Hence the omission of human capital variables in pricing equation explains part of the idiosyncratic risk premium.

Guo, Kassa & Ferguson (2014) however criticise estimating EGARCH models using in-sample data (notably Fu, 2009) since it results in a look-ahead bias. A common technique to estimate an EGARCH model is to use the maximum likelihood method. Even if the model specification only includes lagged idiosyncratic volatility and return data the resulting log likelihood function will include contemporaneous returns. Hence the month t expected idiosyncratic volatility is estimated from parameters containing month t returns. If the month t return is positive then the bias will correlate positively and vice versa for negative month t returns. Since it has been showed that stock returns are positively skewed with more extreme positive returns than negative the bias will make the return-idiosyncratic volatility relation appear positive. By creating a "truly" out-of-sample idiosyncratic volatility they are able to show that after controlling for this bias there is no significant positive relation between expected idiosyncratic volatility and returns. It is however important to underline that the result do not confirm either Ang et al. (2006 & 2009) or the standard asset pricing models. It can be that the relation is indeed positive but estimating the EGACRH model as in e.g. Fu (2009) results in a bias that makes the results unreliable.

Chua, Goh & Zhang (2010) (who's methods I replicate) extend Fu (2009) by also arguing for the use of expected idiosyncratic volatility. Their sample consists of firms listed on NYSE, AMEX and NASDAQ during the period 1963 to 2003. They however contest other studies use of realised returns as a proxy for the expected returns (e.g. Ang et al., 2006 and Fu, 2009). Chua et al. argue that the realised returns are a poor proxy because unexpected return is often the biggest part of total return. Hence using the total realised return might obscure the true relation. They suggest a way to circumvent this problem by splitting realised idiosyncratic volatility in to an expected and unexpected part. The unexpected idiosyncratic volatility they argue is highly correlated with unexpected returns and thus acts as a control variable. Further they measure the idiosyncratic volatility relative to the Fama-French three-factor model as many of the other studies. They estimate expected idiosyncratic volatility by fitting a AR(2) model for each individual firm. The results show that when controlling for unexpected idiosyncratic volatility the relation between expected idiosyncratic volatility and realised returns are positive. This also holds true when controlling for other variables such as size and book-to-market.

Bali & Cakici (2008) finds contradicting evidence of the relation between idiosyncratic volatility and returns. They argue that the relation is sensitive to: i) using daily or monthly data in the estimation of idiosyncratic volatility, ii) the way assets are weighted when calculating portfolio returns (e.g. value-weighted and equally weighted), iii) breakpoints used to divide stocks in to different portfolios, specifically breakpoints including all the CRSP stocks, breakpoints including only NYSE stocks and forming portfolios in such a way that each portfolio contains 20% of the total market share, and iv) sorting out certain stocks depending on their size, price and liquidity. For all tests they follow a portfolio approach and they measure idiosyncratic volatility relative to the CAPM and FF3F model throughout the study. They are able to replicate the main findings of Ang et al. (2006) although taking in to consideration the above issues there are no robust findings.

As have been obvious from the discussion above previous research in this field has mainly focused on firms and not on mutual funds as this study has. Although there have been a few studies investigating idiosyncratic volatility in mutual funds and also its relation to returns. Falkenstein (1996) look at open-ended mutual funds and stocks in the U.S. and concludes that fund managers seem to have aversion for stocks with low idiosyncratic volatility. Thus one would not be surprised to find idiosyncratic risk in mutual funds even tough it is not a proof. Vidal-Garcia & Vidal (2014) for one thing investigates the idiosyncratic risk puzzle suggested by Ang et al. (2006) using a set of U.K. mutual funds. First of all they show that there is idiosyncratic volatility in mutual funds. Secondly they were not able to find any conclusive evidence on the relationship between idiosyncratic volatility and returns. Although their methodology do not follow standard procedure as the other more well cited studies in this field has used, e.g. they estimate the idiosyncratic risk against a panel factor model. Wagner & Winter (2013) test different extensions of two popular factor models (Fama-French Three factor model and the Carhart model) for a sample of European mutual funds and includes idiosyncratic volatility as a risk factor. They find that for many funds idiosyncratic risk affect returns negatively although there study is not a pure study on the relation between idiosyncratic risk and returns.

Previous research finds different evidence as to what the relation between idiosyncratic volatility and returns should be. All the articles provides interesting argument so it is difficult to say which one is the most correct. Although the appeal of using conditional idiosyncratic volatility instead of realised or lagged idiosyncratic volatility is according to me high. This since it is to naïve to assume that realised or lagged idiosyncratic volatility can be a good approximation of future idiosyncratic volatility and the fact that Fu (2009) disproves the use of lagged idiosyncratic volatility. Also it seems that when using conditional idiosyncratic volatility the result indicate a positive relation which is what in line with theoretical development. Findings of a negative relation or findings that are inconclusive should not be disregarded they are indeed also valuable. Also criticism of methodologies as in Guo et al. (2014) and Bali & Cakici (2008) should be adhered to. When it comes to the way one relates idiosyncratic volatility and returns I find that Fama-Macbeth regression are more appropriate than using a portfolio approach. This since i) it is easier to use multiple control variables at the same time and, ii) one get a specific value on the relation using all available data and not just a pattern.

4. Methodology

In this chapter I cover the methodologies used to investigate the problem posed in the introduction. I first cover how I estimated the idiosyncratic volatility. I then move on to explain how I tested for random walk in the idiosyncratic volatility series. After that I explain how I estimated conditional idiosyncratic volatility by using a AR(2) model. Lastly I cover how idiosyncratic volatility and return are related through Fama-MacBeth (1973) regressions, explain the control variables used and portfolio sorting on expected idiosyncratic volatility.

4.1 Estimation of Factor Equation

The first objective of this study was to estimate the monthly idiosyncratic volatility for each fund/month. From the discussion above the pricing model against which I estimate idiosyncratic risk is the Carhart (1997) model. Hence for each month using daily return observations I estimate the following regression model:

$$r_{i,d,m} - r_f = \alpha_{i,m} + \beta_{i,m}^{Mkt-Rf}(Mkt - rf)_{i,m} + \beta_{i,m}^{SMB} SMB_{i,m} + \beta_{i,m}^{HML} HML_{i,m} + \beta_{i,m}^{MOM} MOM_{i,.m} + \varepsilon_{i,d,m} \tag{4}$$

The regression were estimated for each fund *i* and month *m* using returns only for the given month. Thus the regression coefficients were updated each month. As in Fu (2009) I required a minimum of 15 return observations within each month for a regression to be fitted. The obvious reason for this is that too few observations would produce a statistically unstable estimation.

4.2 Estimation of Monthly Idiosyncratic Volatility

Next I estimated the monthly idiosyncratic volatility. The volatility measure was estimated in the same manner as Ang et al. (2006) and Chua et al. (2010), the equation looks like follows:

$$IV_{i,m} = \sum_{d=1}^{N_m} \varepsilon_{i,d,m}^2 \tag{5}$$

Hence the idiosyncratic volatility (IV) for firm *i* in month *m* is equal to the sum of the squared residual from the factor equation above (alternatively IV is equal to the RSS of the factor equation). There is however different ways to estimate the IV, Fu (2009) for example defines IV to be the standard deviation of the daily regression residuals for each month. It would also be possible take the square root of the above equation in order to define it as a standard deviation. Whether a different volatility measure would give a significantly different result in the end is a problem I leave to the reader. I doubt that there would be any significant changes to the results.

4.3 Test Whether IV Series Follows a Random Walk

As have been discussed one major criticism of Ang et al. (2006) was their use of one-month lagged idiosyncratic volatility. As Fu (2009) argues, it is only valid to use lagged idiosyncratic volatility if the IV series follow a random walk. Hence I first test for random walk in the IV series. To show why lagged idiosyncratic volatility can be used if it follows a random walk we can first assume that the IV series follow a random walk without a constant. Then the process will look like follows:

$$IV_{i,t} = IV_{i,t-1} + \eta_{i,t} \tag{6}$$

η_i is the residual term. If we assume that η_i has expected value of zero, and then if we take the expectation of equation 6 we will get the following:

$$E\left[IV_{i,t}\right] = E\left[IV_{i,t-1} + \eta_i\right] = E\left[IV_{i,t-1}\right] + E\left[\eta_{i,t}\right] = IV_{i,t-1} \tag{7}$$

The result from the last identity comes from the fact that the residuals expected value is zero and that the lagged IV has already been realised. Thus the best predictor of tomorrow is today's value. If lagged idiosyncratic volatility is to be used then it should follow a random walk. To test for random walk we will rewrite equation 6 and add a constant:

$$IV_{i,t+1} - IV_{i,t} = \gamma_{0,i} + \gamma_{1.i} IV_{i,t} + \eta_i \tag{8}$$

From the equation it is obvious that if the IV series follows a random walk $\gamma_{1.i}$ should be indistinguishable from zero. I test this by employing a Dickey-Fuller test. First I ran the regression (equation 6) for each mutual fund that had a minimum 24 monthly consecutive observations, the regression was performed on the whole series. Then I calculated t-statistics for each fund as follows:

$$t - statisic = \frac{\hat{\gamma}_{1,i}}{SE(\hat{\gamma}_{1,i})} \tag{9}$$

Where, $\hat{\gamma}_{1,i}$ = *estimated* $\gamma_{1.i}$ and $SE(\hat{\gamma}_{1,i})$ = *sample standard error*. The test statistic was then compared to the Fuller (1996) critical values for a unit root test with a constant on the 1% significance level.

4.4 Estimation of AR(2) Model

After concluding that the IV series do not follow a random walk in most cases there were sufficient grounds to model the IV series using a more sophisticated technique then simply using the lagged values. There are basically two time series techniques that can be used to model the idiosyncratic volatility, one is to use a GARCH type model and the other is to use to fit the series as an ARIMA type model. Fink, Fink & He (2012) finds that GARCH type models are superior to ARIMA type models. This may not be surprising considering GARCH model are more sophisticated. There are however some problems with using GARCH models. First it is important to understand that a GARCH model is estimated using maximum likelihood technique since it involves non-linear estimations. Maximum likelihood estimation as opposed to least squares estimations is not analytical but approximate. Coefficients are fitted by maximising a so-called log-likelihood function which is "connected" to the coefficients. An algorithm then changes the coefficients until the log-likelihood function has reached a local maximum. By maximising the function coefficients that best describe the data are found. But this function can sometimes be rather flat and thus there is a chance that no maximum will be reached before the algorithm stop iterating resulting in an error (non-convergence). Obviously it is possible to use a GARCH method but the risk of non-convergence and other practical difficulties make it infeasible for me to use a GARCH type model. So what is left is to use a ARIMA model and I choose to model the IV series as a AR(2) model as in Chua et al. (2010). Their choice to model the IV series as an AR(2) model is based on Aikike information criterion and the amount of serial autocorrelation. Hence there is no economic grounds on using the AR(2) model but the choice is purely atheoretical. The proper way in my case would be to test how the model should be specified using e.g. information criterion. I chose the AR(2) specification anyway for practical reason. For each mutual fund I fit the following AR(2) model:

$$IV_{i,m} = \phi_0 + \phi_1 IV_{i,m-1} + \phi_2 IV_{i,m-2} + \xi_{i,m} \tag{10}$$

Where i is the specific mutual fund and m is the current month. The model is fitted for funds with at least 35 monthly observations and expanding window approach is employed. The model is updated each month including all the previous observations up to the current month. The number of observations after each update is thus increasing. After the AR(2) model was estimated for all the funds the one-step ahead expected idiosyncratic volatility (EIV) was calculated as:

$$EIV_{i,m} = \phi_0 + \phi_1 IV_{i,m-1} + \phi_2 IV_{i,m-2} \tag{11}$$

Where $EIV_{i,m}$ is the expected idiosyncratic volatility for firm i at time m. As in Chua et al. (2010) I also calculate the unexpected idiosyncratic volatility (UIV). They argue that relating realized returns to idiosyncratic volatility as most studies do is not a good choice since a big part of realised returns consist of the unexpected part of returns. Because of difficulties in decomposing realised stock return in to a expected part and unexpected it might be better to include a variable that is highly correlated with unexpected returns such us unexpected idiosyncratic volatility (UIV). UIV is calculated as:

$$UIV_{i,m} = IV_{i,m} - EIV_{i,m} \tag{12}$$

Where $UIV_{i,m}$ is the unexpected idiosyncratic volatility for fund i in month m.

4.5 Fama-MacBeth Regressions

The next step was to investigate the cross-sectional relation between idiosyncratic risk and return/risk-adjusted return. I did this by estimating Fama-MacBeth (1973) regressions according to equation 13.

$$Y_{i,m} = \varphi_{0t} + \sum_{k=1}^{K} y_{k,m} X_{k,i,m} + \eta_{i,m} \tag{13}$$

It is evident from the equation above that for each month m a cross sectional regression is estimated on k independent variables. When a cross-sectional regression has been fitted for all months in the sample the mean value of each coefficient is taken. In addition I use Newey-West (1987) adjusted standard errors for the estimation of the t-statistics. Newey-West standard errors are used when the errors in the model is thought to be heteroskedastic and autocorrelated. Before the regressions are run the maximum number of autocorrelation lags that the estimator can handle must be chosen. I used a simple technique presented in Greene (2012, p. 530) stated as $n^{1/4}$ where n is the number of time periods (239 in this case).

In this study the monthly gross return and the Carhart four-factor alpha has been used as the dependent variable. The alpha is defined as the intercept in equation 4. But instead of estimating the model using daily returns I use monthly in order to get the monthly alpha. Further I employ a rolling regression approach with a window of 36 monthly observations. The alpha also known as the abnormal return essentially gives the return above the return rewarded by taking on systematic risk or in other words the risk-adjusted returns. Hence it shows the skill of the fund manager.

4.6 Control Variables

Apart from using EIV and UIV as independent variables I also use the variables TER (total expense ratio), one month lagged TNA (total net assets), Age (months) as control variables. The idea is that these variables can explain returns in mutual funds and thus control for EIV and UIV. Fund managers argue that high fund costs (high TER) results in the fund being better managed and leads to higher returns, hence we would expect TER to be positively related to returns/alpha (Otten & Bams, 2002). Chen, Hong, Huang and Kubik (2004) argue that fund size affect performance of mutual funds when studying a sample of U.S. equity funds. There is no clear economical relation between fund returns and size but Chen et al. (2004) finds a negative relation after examining the sample empirically. They find that funds that invest a lot in small cap stocks experience a higher negative relation between size and returns and adheres it to illiquidity. In addition they argue that organisational diseconomies also is responsible for this negative relationship. In fund management soft information such as speaking with the CEO in a potential stock is important. This type of information can be hard to send up the hierarchy and thus for large funds soft information may be hard to process and hence they may perform worse than small funds since they can only focus on hard information (i.e. there are organisational diseconomies). Hence I would expect to find a negative relationship between size (TNA) and returns. When it comes to fund age the economical motivations to whether the relation should be positive or negative is more diffuse. Although as Chen et al. (2004) points out size and age may be partially correlated. For example one could expect that as a fund gets older it will be bigger and then suffer from the negative effects described above. Webster (2002) points out that age can be positively related to returns due to “accumulated experience”, “resources” and a better understanding of the market.

Otten & Bams (2002) estimate a one-period cross-sectional regression with the conditional four-factor alpha as the dependent variable for mutual funds in four different countries (France, Germany, Netherlands and the UK). They find that expense ratio is negatively related to alpha and for three countries this relation is significant. Fund assets are found to be positively related for all countries and significant. Age is negatively related for all countries although only significant for two countries. Among other variables Carhart (1997) also use

Expense Ratio and lagged TNA as independent variables in a Fama-MacBeth regression similar to the one used in this study. He finds that the coefficient for the expense ratio is statistically significant and negative and that the coefficient for lagged TNA is also negative but not significant.

There are more potential variables that could explain the cross-section of mutual fund returns but the ones I used in this study were the only available. More details about the construction of the control variables such is found in the data chapter of this essay.

4.7 Portfolio Sorting on EIV

As an extra robustness test I form portfolios sorted on EIV, I follow the method of Fu (2009) (see table 6 in Fu, 2009). For each month five quintile portfolios are formed on EIV, i.e. the mutual funds with the lowest 25% EIV are placed in portfolio one, the portfolio with the lowest EIV. The same procedure is repeated for the other quintiles. When all the portfolios had been formed for all months in the sample the time-series average of each monthly equally weighted portfolio return was estimated. The pooled mean or median (depending on the variable) of EIV, UIV, IV, TER, TNA and Age were calculated. I do not control for other variables affecting the returns as for example Ang et al. (2006). They use a double sorting approach and as such control for one variable at a time. It would of course be interesting from an academic standpoint to do this but I determined that it was out of this essay's scope.

5. Data

This chapter starts of by providing information on the sample, how it was chosen and retrieved. I then explain how the factors in the Carhart four-factor model are constructed. I finish by explaining how the control variables are constructed.

The data consists of a sample of mutual funds with a focus on equity investments registered in the United States. The sample includes a total of 10 917 mutual funds active some time during the period April 1995 to February 2015. The mutual funds are all denominated in local currency (USD). All the mutual funds have United States as their geographical focus. This since the study has used U.S. factors in the factor model employed; including mutual funds with a global focus would create false results since the factors are not global. From Bloomberg Terminal information about fund codes and fund names were downloaded. The sample was divided in to four sets depending on what kind of stocks each mutual fund in question focused on. The subsets are Growth (funds focusing on growth stocks, 3104 funds), Value (funds focusing on Value stocks, 2941 funds), Large Cap (funds focusing on large capitalization stocks, 3189 funds) and Small Cap (funds focusing on small capitalization, 1683 funds). The subsets were chosen on the basis that i) the individual subsets contained a reasonable amount of funds so as to be able to make statistical inference and as a whole make up a substantial part of the fund market in the U.S. for the sample period. There were other subsets of interest but with too few mutual funds in them. ii) I wanted to compare subsets that were mutually exclusive; comparing subsets where it is possible that funds are in multiple classes would distort the analysis especially since some subsets overlap more than others. The sample includes surviving and non-surviving funds thus omitting the problem of survivorship bias. Including only surviving funds in a sample often lead to overstated performance due to the fact that underperforming funds tend to die and fall out of the sample (Rohleder, Scholz & Wilkens, 2011).

From Datastream price information were downloaded and then converted in to simple returns. Due to the nature of the essay both daily and monthly price data were downloaded. The price data downloaded was a return index (RI) and assumes that dividends are reinvested by purchasing additional shares of a fund at the closing bid-price after the ex-dividend date. Price data were not available for all the mutual funds on the list downloaded from Bloomberg hence the number stated above does not reflect the true number of funds within each class. Further mutual funds that contained price series shorter than 36 months were dropped from the sample since those funds would not contribute much to the analysis.

5.1 Construction of Carhart Factors

Since I used the Carhart four-factor model to estimate the idiosyncratic volatility I needed to proxy the four factors. I have only focused on U.S. registered mutual funds and thus I have been able to use data supported by K. French website[2]. The SMB and HML factors from the website have been constructed in the same way as Fama & French (1993).

The factors are constructed as follows. First firms traded on NYSE on June for the year of interest are separated as either small size or big size firms by the median ME (market equity). Then firms are divided in to different growth styles depending on its book-to-market ratio, the separation points are the 70th BE/ME percentile and 30th BE/ME percentile. This creates 6 portfolios formed on book-to-market ratio and size (market equity). Thus SMB and HML factors are calculated as follows:

$$SMB = \frac{(Small\ value\ ret. + Small\ neutral\ ret. + Small\ growth\ ret.)}{3} - \frac{(Big\ value\ ret. + Big\ neutral\ ret. + Big\ growth\ ret.)}{3} \qquad (14)$$

And,

$$HML = \frac{(Small\ value\ ret. + Big\ value\ ret.)}{2} - \frac{(Small\ growth\ ret. - Big\ growth\ ret.)}{2} \qquad (15)$$

Where each term in the denominator of the equations indicates the returns of each portfolio. *Small* indicates that the portfolio consist of firms with a low ME, *big* the opposite of *small, value* means that the BE/ME for the firms in the portfolio is low, *growth* that the firms has high BE/ME and *neutral* that the BE/ME for the firms is in between *value* and *growth*.

The equity premium factor (R_m-R_f) is estimated by subtracting the risk-free return from the market return. The market return is proxied by the value-weighted return of U.S. firms that

[2] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. I would also like to thank Kenneth French for supporting this data library.

are available on CRSP and listed on NYSE, AMEX and NASDAQ. The risk free rate is proxied by the one-month Treasury-bill rate.

The momentum factor (MOM) is estimated in a similar fashion as the SMB and HML factor. The portfolios are constructed using return data from stocks listed on NYSE, AMEX and NASDAQ. Although in this case portfolios are formed on size (ME) and 2 to 12 monthly prior returns. The size breakpoint is the median ME on NYSE stocks. The prior return breakpoint is the 30th and the 70th percentiles on NYSE stocks. MOM is then calculated as:

$$MOM = \frac{(Small\ High + Big\ High)}{2} - \frac{(Small\ Low + Big\ Low)}{2} \qquad (16)$$

Where *small* has the same meaning as above, *high* means that the firms in the portfolio has had a high previous returns, *low* means that the firms has had low previous returns. For more detailed information about the constructions of the factors please refer to Fama & French (1993) and K. French website.

5.2 Construction of Control Variables

Total Expense Ratio (TER)

TER is defined by Datastream as: "*The Total Expense Ratio (after waivers/reimbursements are subtracted, but before expense offsets/brokerage service arrangements are subtracted) as reported in the financial highlights in the annual report*".[3] Total Expense Ratio can be defined as a fund's costs divided by its assets.

Total Net Assets (TNA)

TNA is defined by Datastream as: "*The Total Net Assets (TNA) of the fund. TNA represents the total funds under management net of fees and expenses for a particular date. TNA is expressed in Millions*".[4]

Age

Age is defined as the number of months that the mutual fund has been in existence. There was no database available from which I could retrieve the age in months, I thus had to estimate the age myself. I estimated the age by first downloading price data for all the funds (from Datastream) and I chose to download data sufficiently long back in history to ensure that there

[3] http://extranet.datastream.com/data/Unit%20trusts/documents/Lipper%20on%20Datastream.pdf
[4] http://extranet.datastream.com/data/Unit%20trusts/documents/Lipper%20on%20Datastream.pdf

were only N/A (non-available) entries. I then assumed that until the first numerical entry appeared for a fund it was to be treated as “unborn” and after the last numerical entry I treated it as “dead”. Hence when the first numerical entry for a fund appeared its age was determined to one month and so on.

6. Results

This chapter presents the findings of the essay. I first document that the idiosyncratic volatility (IV) series, as a whole does not follow random walks. I then present statistic on the AR(2) model and summary statistic on the variables used. Then I discuss a potential multicollinearity problem and the rest of the chapter discuss the outcome of the Fama-MacBeth regression and the portfolio sorting on expected idiosyncratic risk.

4.1 The Process of the Idiosyncratic Volatility

Table 1 presents results on the Dickey-Fuller test performed on each mutual fund. In the table there is some summary statistic of γ_1 which is the coefficient for $IV_{i,t}$ in equation 8. The number in the parenthesis is the t-statistics of the coefficient and N represents the number of funds in each class (i.e. the number of IV series). The last column in the table presents the percentage of IV series that can be rejected as following a random walk (as specified in the methodology chapter). As is evident no fund class has a rejection percentage under 90% and the aggregated sample has a rejection percentage of about 98%. Hence we can safely conclude that in large idiosyncratic volatility cannot be described as following a random walk and using lagged idiosyncratic volatility as the next period expected idiosyncratic volatility is faulty. Fu (2009) also conduct this test and he finds that about 90% of IV series do not follow a random walk.

Table 1. Test if Idiosyncratic Volatility Follows a Random Walk.
This table reports the result from the Dickey-Fuller test, testing for random walk. The test is performed by regressing the first differenced idiosyncratic volatility against the current months idiosyncratic volatility for each mutual fund (see eq. 8). The statistics reported in the table are the mean, first quartile, median and third quartile of the coefficient estimates for all mutual funds within each class. The numbers in the parenthesis is the mean and quintile t-statistics. *N* is the total number of mutual funds in a given class. *RW rejected* states the percentage of all the mutual funds that have been rejected as a random walk after comparing the t-statistics (see eq. 9) with the critical values at the 1% level (see the lower section of the table).

		Mean	Q1	Median	Q3	N	RW rejected (%)
Growth	γ_1	-0,6592	-0,9254	-0,5981	-0,4452	3104	97,94
		(-12,58)	(-9,27)	(-7,00)	(-5,66)		
Value	γ_1	-0,7360	-1,0010	-0,7741	-0,4944	2941	90,44
		(-12,72)	(-10,63)	(-7,45)	(-5,86)		
Large Cap.	γ_1	-0,7368	-0,9941	-0,7694	-0,4943	3189	98,97
		(-10,44)	(-10,86)	(-7,57)	(-5,98)		
Small Cap.	γ_1	-0,7627	-1,0088	-0,8391	-0,5070	1683	98,57
		(-15,66)	(-11,12)	(-7,67)	(-6,12)		
Total	γ_1	-0,7185	-0,9898	-0,7185	-0,4780	10917	98,37
		(-12,47)	(-10,43)	(-7,38)	(-5,85)		

Dickey-Fuller Critical t-statistics (Fuller, 1996)	
No. of Obs.	t-statistic (1% level)
25	-3,75
50	-3,59
100	-3,50
250	-3,45

4.2 AR(2) Model Coefficients

Table 2 depicts the median regression coefficients from the AR(2) regression explained in the methodology chapter. As can be seen the coefficient for the first lagged IV is greater than the coefficient for the second lagged IV. Hence the last months IV affect the expected IV in a greater way than the IV two months ago. Another striking fact from the table is that funds in the Growth class have significantly higher coefficient estimates. The reason for this I leave to the reader. Chua et al. (2010) use a sample of U.S. firms, the median coefficients in their study are as follows: 0,0074, 0,2233 and 0,1297. My estimates are qualitatively similar to Chua et al. (2010), considering that the mutual funds are American and have equity focus it is not a big surprise.

Table 2. Median AR(2) Regression Coefficients.
This table displays median coefficients of the AR(2) models estimated on each individual mutual funds idiosyncratic volatility series. Each column presents the median coefficient estimate on all the funds within a class.

	ϕ_0	ϕ_1	ϕ_2
Growth	0,0002	0,2754	0,1105
Value	0,0001	0,1586	0,0568
Large Cap.	0,0001	0,1489	0,0597
Small Cap.	0,0001	0,1383	0,0586
Total	0,0001	0,1934	0,0738

4.3 Summary Statistics of Variables

In table 3 pooled summary statistic of variables used in the study is presented. The first thing to note is that there is idiosyncratic risk in mutual funds (see the IV row), Growth funds have the highest IV with a mean of 0,0961% and Large Cap fund have the lowest mean IV with 0,0183%. This can be compared with the aggregated sample (total) with a mean of IV of 0,0446%. One might argue that since mutual funds have a greater ability to diversify than individual investors the idiosyncratic risk should be slim to none. Although compared to studies that investigate firms the idiosyncratic risk is smaller for the mutual funds in my sample. Chua et al. (2010) for example has a mean IV of 2,79% so it seems that funds have lower idiosyncratic volatility then individual firms, which is logical. Further EIV, UIV and IV are for all the fund classes and in aggregate heavily skewed. EIV is positively skewed for all the funds classes except for Large Cap; UIV is positively skewed for all the classes except for Large Cap. Hence when it comes to Large Cap funds skewnes for EIV and UIV is reversed compared to the other classes. IV on the other hand is positively skewed for all the classes including the aggregate total class. The problem with highly skewed variables is that ordinary regression models are mean models. Since the mean is not a good way to describe the central tendency for highly skewed variables using them in a regression can produce sensitive results. Taking the logarithm of EIV and UIV would not be a great idea since for both and especially the UIV there are many negative values. Both for TNA and Age I took the natural logarithm to create make them more symmetrical and for TNA the coefficient estimate where abysmally small when the logarithm were not taken. Further I used the one month lagged TNA to avoid spurious correlation (Granger & Newbold, 1974).

Table 3. Summary Statistics of Variables For the Pooled Sample

This table provides pooled mean statistics for all the mutual funds in each class during April 1995 to February 2015. The data have been adjusted so as to match the length of EIV and UIV series, the main variables in the study, i.e. if TER have been reported for all years for a fund but the EIV series are reported only for 5 years the TER series is cut to match the EIV and UIV series. *Ret* is the monthly gross return (in percent). *Excess Ret* is the monthly excess return, i.e. the gross return minus the monthly 3-month T-Bill. *4F Alpha* is the alpha (intercept) using the Carhart model, to get the alpha I fitted the model using monthly data and a rolling window with 30 observations. *EIV* is the one-month ahead expected idiosyncratic volatility estimated according to eq. 11. *UIV* is the unexpected idiosyncratic volatility and was estimated according to eq. 12. *IV* is the idiosyncratic volatility relative to the Carhart model and estimated as in eq. 5. *TER* is the total expense ratio, the definition can be found in the methodology chapter and it was downloaded from Datastream. *TNA* is the total net asset of the fund and the logarithm was taken. It is also lagged with one month. Age is reported in months and the logarithm was taken due to high skewness.

Panel A: Growth							
Variables	Mean	S.D	Q1	Median	Q2	Skew.	N
Ret (%)	0,6771	5,6720	-2,1994	1,2696	0,0416	0,0142	295868
Excess Ret (%)	0,5505	5,6851	-2,3293	0,0111	0,0404	0,0063	295868
4F Alpha (%)	-0,1152	0,4895	-0,3300	-0,1021	0,0012	10,2840	294744
EIV (%)	0,1405	1,2781	0,0120	0,0304	0,0013	34,7404	298163
UIV (%)	-0,0444	1,3325	-0,0434	-0,0090	-0,0009	-26,8364	297650
IV (%)	0,0961	0,5049	0,0059	0,0130	0,0567	65,0800	297650
TER	1,4054	0,5537	1	1,3	1,8	0,9948	223896
$\ln(TNA_{t-1})$	4,3601	2,3723	2,8736	4,5098	6,0002	-0,3391	224699
ln(Age) (months)	4,6283	0,5844	4,1744	4,6052	5,0106	0,3631	297893
Panel B: Value							
Variables	Mean	S.D	Q1	Median	Q2	Skew.	N
Ret (%)	0,6554	5,2276	-1,9092	1,3002	3,6391	1,9725	263401
Excess Ret (%)	0,5379	5,2401	-2,0471	1,1408	3,5313	1,9719	263401
4F Alpha (%)	-0,1511	0,4008	-0,3116	-0,1254	0,0406	1,3023	261846
EIV (%)	0,0351	0,5688	0,0072	0,0151	0,0329	48,2930	265078
UIV (%)	-0,0111	0,6518	-0,0194	-0,0056	-0,0011	-13,7441	264665
IV (%)	0,0240	0,3365	0,0032	0,0060	0,0126	129,0019	264700
TER	1,3604	0,5228	1	1,26	1,73	0,7711	210029
$\ln(TNA_{t-1})$	4,3434	2,3502	2,8449	4,4739	6,0027	-0,3301	210699
ln(Age) (months)	4,5866	0,5627	4,1431	4,5643	4,9698	0,3407	264891
Panel C: Large Cap.							
Variables	Mean	S.D	Q1	Median	Q2	Skew.	N
Ret (%)	0,6006	4,9997	-1,8211	1,1737	3,4884	4,1373	301335
Excess Ret (%)	0,4776	5,0132	-1,9585	1,0075	3,3758	4,1137	301335
4F Alpha (%)	-0,1411	0,3941	-0,2784	-0,1171	0,0094	10,5706	300229
EIV (%)	0,0405	0,6946	0,0046	0,0105	0,0264	-15,3521	303323
UIV (%)	-0,0223	0,7678	-0,0160	-0,0038	0,0006	21,5676	302876
IV (%)	0,0183	0,3477	0,0021	0,0043	0,0094	120,7862	303017
TER	1,2072	0,5697	0,8300	1,1300	1,5600	0,9049	234348
$\ln(TNA_{t-1})$	4,5149	2,5407	2,8900	4,6578	6,3430	-0,3134	235099
ln(Age) (months)	4,6395	0,5895	4,1897	4,6151	5,0304	0,3399	303224

(Continued)

Table 3. Summary Statistics of Variables For the Pooled Sample (Continued)

Panel D: Small Cap.							
Variables	Mean	S.D	Q1	Median	Q2	Skew.	N
Ret (%)	0,7264	6,0383	-2,5945	1,4596	4,4628	0,3272	150766
Excess Ret (%)	0,6074	6,0518	-2,7148	1,3179	4,3682	0,3271	150766
4F Alpha (%)	-0,1854	0,4656	-0,3873	-0,1404	0,0649	-0,0959	149825
EIV (%)	0,0420	0,5027	0,0106	0,0206	0,0431	34,6464	151711
UIV (%)	-0,0099	0,6980	-0,0253	-0,0079	-0,0015	28,1725	151563
IV (%)	0,0322	0,4949	0,0048	0,0090	0,0176	115,9098	151588
TER	1,4289	0,5400	1,0900	1,3400	1,7700	0,5876	117753
ln(TNA_{t-1})	4,0868	2,2059	2,7014	4,2584	5,6623	-0,4568	118146
ln(Age) (months)	4,5487	0,5286	4,1271	4,5433	4,9345	0,2318	151601
Panel E: Total							
Variables	Mean	S.D	Q1	Median	Q2	Skew.	N
Ret (%)	0,6560	5,4234	-2,0219	1,2683	3,8498	1,5009	1011370
Excess Ret (%)	0,5340	5,4365	-2,1602	1,1087	3,7390	1,4934	1011370
4F Alpha (%)	-0,1427	0,4371	-0,3153	-0,1189	0,0522	6,7941	1006644
EIV (%)	0,0686	0,8637	0,0077	0,0173	0,0433	35,3135	1018275
UIV (%)	-0,0240	0,9374	-0,0235	-0,0060	-0,0008	-18,5287	1016754
IV (%)	0,0446	0,4216	0,0034	0,0072	0,0171	98,0095	1017000
TER	1,3378	0,5555	0,9600	1,2500	1,7100	0,8022	786026
ln(TNA_{t-1})	4,3608	2,3981	2,8449	4,5020	6,0433	-0,3259	788643
ln(Age) (months)	4,6089	0,5733	4,1589	4,5951	4,9972	0,3456	1017609

4.4 The Correlation Between EIV and UIV (A Potential Multicollinearity Problem)

Table 4 presents the time series mean of the cross-sectional correlations between EIV and UIV. As can be seen for all the fund classes and in aggregate the mean correlation is highly negative. Hence there is an issue of near collinearity between EIV and UIV. The main problem of using collinear variables in a regression is that the coefficient estimate will not be precise and hence the standard errors will be large (low statistical significance) (Mansfield & Helms, 1982). The R^2 may also stay high or increase when adding a collinear variable but that is due to the fact that the model as a whole explains a lot of the variance (Brooks, 2012, p.172). Another consequence is that the estimated coefficients can become large and change values when including a collinear variable (Mansfield & Helms, 1982). If we just glance at table 5 and 6 for changes in t-statistics and changes in coefficient estimates there are some patterns. Looking at table 5 where only EIV and UIV are included there are mixed evidence, only for the Growth class is there an increase in the significance and no change of sign for EIV regardless of the dependent variable being the four-factor alpha or gross returns. For the rest of the classes including the aggregated Total there are mixed evidence, with increases and decreases of the significance and changes in the sign of EIV for both gross return and four-factor alpha. Including the control variables there are however more clear patterns, at least for the regressions using four-factor alpha as dependent variable. In table 6 it is evident that for all the classes there is an increase in significance and no change in sign for EIV when UIV is included in addition to the other variables and when the dependent variable is the four-factor alpha. When gross return is used as dependent variable there is mixed result regarding significance and signs of EIV when adding UIV. Chua et al. (2010) do not report any correlations of the variables included in the study hence there is difficult to say anything about the existence of multicollinearity in their study. Although adding UIV increases the significance of EIV (when also including the control variables). Two examples of remedies for the potential multicollinearity problem would be to use principal components analysis or ridge regressions. Although since the problem of low significance and changing signs of EIV is not big at least for the Fama-MacBeth regressions using the four-factor alpha as the dependent variable I will not use any of these methods. But it could be a suggestion for the future to employ one of these methods to see if the results change.

Table 4. Time-Series Mean of Cross Sectional Correlation for EIV and UIV.
This table displays statistics on the correlation between EIV and UIV. The correlation was estimated by first taking the cross-sectional correlation within a given class for each time period and then taking the average of these correlation coefficients.

	Growth	Value	Large Cap.	Small Cap.	Total
Corr(EIV, UIV)	-0,8738	-0,8331	-0,9293	-0,8724	-0,9121

4.5 Fama-MacBeth Output Without Control Variables

If we now focus on the actual outcome of the Fama-MacBeth regressions without any of the control variables the results are mixed as can be seen in table 8. Only for the Growth class the coefficients are significant at the 1% level regardless of the dependent variable. For Large Cap funds there are significant coefficients for the regressions using the four-factor alpha as the dependent variable. Although as opposed to Growth funds the coefficients for Large Cap funds are positive. In aggregate the coefficients are negative when using gross returns as the dependent variable although when using the four-factor alpha the sign switch from negative to positive when UIV is included. Hence it is difficult to establish any clear relation between returns and idiosyncratic volatility when just including EIV and UIV.

Table 5. Fama-MacBeth Regression on EIV and UIV.

This table displays the output from the Fama-MacBeth regression including only EIV and UIV during the period April 1995 to February 2015. The numbers without parenthesis indicate the regression coefficients and the numbers within parenthesis is the Newey-West adjusted t-statistics. The coefficients are averages of cross-sectional regressions that have been performed for each month for all the funds within a particular class. *Avg. R^2* is the time-series averages of the cross-sectional R^2. EIV and UIV values higher/lower than 1/-1 have been deleted to get rid of extreme observations. Model 1-2 use gross returns as the dependent variable and models 3-4 use the four-factor alpha as the dependent variable.

Model	Cons.	EIV	UIV	Avg. R^2
Panel A: Growth				
1	0,0062	-0,0405		0,0083
	(1,50)	(-1,12)		
2	0,0071*	-1,0194***	-1,0797***	0,0910
	(1,75)	(-2,61)	(-2,73)	
3	-0,0015***	-0,0062***		0,0101
	(-4,73)	(-0,35)		
4	-0,0014***	-0,2477***	-0,2605***	0,0292
	(-4,42)	(-3,53)	(-3,38)	
Panel B: Value				
1	0,0059	0,0059		0,0079
	(1,65)	(0,08)		
2	0,0067	-2,4732*	-2,5773*	0,1132
	(1,99)	(-1,75)	(-1,83)	
3	-0,0017***	0,0253		0,0702
	(-6,58)	(0,29)		
4	-0,0017***	0,1240	0,1216	0,0897
	(-6,80)	(0,43)	(0,37)	
Panel C: Large Cap.				
1	0,0049	-0,0422		0,0061
	(1,45)	(-1,22)		
2	0,0054	0,5072	0,5079	0,1052
	(1,61)	(0,28)	(0,28)	
3	-0,0017***	0,1562***		0,1016
	(-9,34)	(3,13)		
4	-0,0018***	0,4145**	0,2589	0,1140
	(-9,88)	(2,10)	(1,15)	
Panel D: Small Cap.				
1	0,0069*	-0,1480		0,0076
	(1,68)	(-1,46)		
2	0,0079**	-2,3918	-2,4126	0,1283
	(2,01)	(-1,57)	(-1,60)	
3	-0,0022***	-0,0711		0,0128
	(-6,62)	(-1,65)		
4	-0,0022***	-0,0074	0,0884	0,0317
	(-7,01)	(-0,03)	(0,30)	

(Continued)

Table 5. Fama-MacBeth Regression on EIV and UIV. (Continued)

Panel E: Total				
Model	Cons.	EIV	UIV	Avg. R^2
1	0,0058	-0,0466**		0,0046
	(1,60)	(-2,17)		
2	0,0065*	-1,2095**	1,2190**	0,0875
	(1,83)	(-2,29)	(-2,31)	
3	-0,0018***	0,12873***		0,0447
	(-7,12)	(3,37)		
4	-0,0017***	-0,1468*	-0,2893***	0,0601
	(-7,20)	(-1,73)	(-2,67)	

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

4.6 Fama-MacBeth Output Including Control Variables

For Fama-MacBeth regression including all the control variables explained in the methodology chapter the story is a bit different. In table 6 Fama-MacBeth regressions including the control variables are presented. It is evident that the regressions that perform best are the ones using the four-factor alpha as the dependent variable and including UIV. For all the regressions using the alpha and UIV the coefficient for EIV is significant at least at a 5% level and also positive. Hence using all the control variables and using the risk-adjusted returns I get results in line with other studies using conditional expected idiosyncratic volatility and employing a Fama-MacBeth methodology, such as Fu (2009) and Chua et al. (2010).

Table 6. Fama-MacBeth Regressions on EIV, UIV and Control Variables.
This table displays the output from the Fama-MacBeth regression including EIV, UIV and the control variables using data from April 1995 to February 2015. The numbers without parenthesis indicate the regression coefficients and the numbers within parenthesis is the Newey-West adjusted t-statistics. The coefficients are averages of cross-sectional regressions that have been performed for each month for all the funds within a particular class. *Avg.* R^2 is the time-series averages of the cross-sectional R^2. EIV and UIV values higher/lower than 1/-1 have been deleted to get rid of extreme observations. Models 1-2 use gross returns as the dependent variable and models 3-4 use the four-factor alpha as the dependent variable.

Panel A: Growth							
Model	Cons.	EIV	UIV	TER	$\ln(TNA_{t-1})$	ln(Age)	Avg. R^2
1	0,0092**	0,1035		-0,0008	-0,0002	-0,00003	0,0407
	(2,43)	(0,96)		(-1,52)	(-1,18)	(-0,11)	
2	0,0085**	-0,5268	-0,6965	-0,0005*	-0,0002	0,0001	0,0896
	(2,19)	(-1,09)	(-1,47)	(-1,01)	(-1,18)	(0,36)	
3	0,0019***	0,1001***		-0,0007***	0,0005***	-0,0010***	0,0715
	(4,95)	(3,11)		(-8,02)	(7,66)	(-6,91)	
4	0,0018***	0,5360***	0,4988***	-0,0007***	0,0005***	-0,0010***	0,0926
	(4,86)	(5,00)	(3,96)	(-8,12)	(7,62)	(-6,74)	

(Continued)

Table 6. Fama-MacBeth Regressions on EIV, UIV and Control Variables. (Continued)

Model	Cons.	EIV	UIV	TER	$\ln(TNA_{t-1})$	ln(Age)	Avg. R^2
Panel B: Value							
1	0,0059	0,0993		-0,0003	-0,0001	0,0003	0,0276
	(1,56)	(1,51)		(-0,55)	(-0,63)	(1,32)	
2	-0,0069*	-0,6023	-0,6692	-0,0002	-0,0001	0,0002	0,0980
	(1,86)	(-0,42)	(-0,46)	(-0,38)	(-0,97)	(1,03)	
3	0,0021***	0,1155		-0,0007***	0,0004***	-0,0009***	0,1312
	(6,42)	(1,50)		(-9,37)	(7,64)	(-8,79)	
4	0,0021***	0,7983**	0,6963*	-0,0007***	0,0004***	-0,0009***	0,1514
	(6,50)	(2,21)	(1,78)	(-9,95)	(7,79)	(-9,17)	
Panel C: Large Cap.							
Model	Cons.	EIV	UIV	TER	$\ln(TNA_{t-1})$	ln(Age)	Avg. R^2
1	0,0040	-0,0398		-0,0007*	-0,0002	0,0007***	0,0217
	(1,21)	(-1,00)		(-1,77)	(-1,12)	(2,93)	
2	0,0046	2,9902	2,9864	-0,0007	-0,0002	0,0006***	0,0931
	(1,39)	(1,58)	(1,57)	(-1,93)	(-1,15)	(3,67)	
3	0,0007***	0,1685***		-0,0008***	0,0002***	-0,0005***	0,1546
	(4,01)	(3,33)		(-12,29)	(4,87)	(-5,00)	
4	0,0009***	1,2037***	1,0366***	-0,0009***	0,0002***	-0,0005***	0,1713
	(4,95)	(4,88)	(3,94)	(-13,19)	(4,90)	(-5,47)	
Panel D: Small Cap.							
Model	Cons.	EIV	UIV	TER	$\ln(TNA_{t-1})$	ln(Age)	Avg. R^2
1	0,0080**	-0,1324		-0,0007	-0,0002	0,0004	0,0323
	(2,00)	(-0,87)		(-1,49)	(-0,89)	(1,57)	
2	0,0082**	-1,2591	-1,1889	-0,0007	-0,0001	0,0003	0,1107
	(2,05)	(-0,71)	(-0,68)	(-1,61)	(-0,60)	(1,44)	
3	-0,0021	0,0209		-0,0001	0,0006***	-0,0009***	0,0886
	(-0,56)	(0,36)		(-1,20)	(8,85)	(-6,24)	
4	-0,0003	0,9316**	0,9343**	-0,0002**	0,0006***	-0,0009	0,1078
	(-0,86)	(2,49)	(2,35)	(-2,13)	(9,18)	(-6,38)	
Panel E: Total							
Model	Cons.	EIV	UIV	TER	$\ln(TNA_{t-1})$	ln(Age)	Avg. R^2
1	0,0060*	-0,0450		-0,0003	-0,0002	0,0047***	0,0291
	(1,73)	(-0,77)		(-0,61)	(-1,13)	(2,79)	
2	0,0063*	-0,4993	-0,5219	-0,0002	-0,0002	0,0004***	0,0813
	(1,80)	(-0,75)	(-0,79)	(-0,41)	(-1,16)	(2,76)	
3	0,0011***	0,1619***		-0,0006***	0,0004***	-0,0077***	0,1058
	(4,39)	(3,67)		(-7,53)	(7,06)	(-6,82)	
4	0,0011***	0,5124***	0,3556**	-0,0006***	0,0004***	-0,0008***	0,1214
	(4,52)	(4,09)	(2,37)	(-8,00)	(7,13)	(-6,88)	

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

4.7 Results of the Portfolio Sorting on EIV

Table 7 presents results on portfolios sorted on EIV. The important variables to study are the equally weighted returns (EWRET) and the four-factor alphas in the different portfolios. For each class except for the Large Cap class both EWRET and 4F Alpha increase between the low portfolio and portfolio 2 but then decrease for each higher portfolio. For the Large cap class EWRET and 4F alpha goes up and down between the portfolios. Table 7 is the portfolio version of the Fama-MacBeth regressions without control variables (also excluding UIV) presented in table 5. The big difference between the portfolio sorting and the Fama-MacBeth regression without any of the control variable is that there seem to be a more defined negative relation if the increase between portfolio one and two is ignored. Fu (2009) who also produce this portfolio sorting get a result that indicates a positive relation.

Table 7. Returns for Portfolios Formed on EIV.
This table presents summary statistics on portfolios formed on EIV and updated every month during April 1995 to February 2015. In every month mutual funds are sorted in to quintile portfolios based on its EIV the first portfolio contains the funds with the lowest idiosyncratic volatility and the last the funds with the highest idiosyncratic volatility. EWRET is the equally weighted portfolio returns. *4F Alpha* is the alpha (intercept) using the Carhart model, to get the alpha I fitted the model using monthly data and a rolling window with 30 observations. All the variables with *med* in parenthesis indicates that the pooled median has been calculated otherwise the pooled average has been calculated. Variables with *×100* indicate that the value have been multiplied by 100.

Panel A: Growth					
Variables	Portfolios formed on EIV				
	Low	2	3	4	High
EWRET (×100)	0,6303	0,6731	0,6174	0,5851	0,5406
4F Alphas (×100)	0,4911	0,5396	0,4784	0,4482	0,3960
EIV (med) (×100)	0,0066	0,0144	0,0290	0,0942	0,2824
UIV (med) (×100)	-0,0015	-0,0048	-0,0128	-0,0350	-0,0996
IV (med) (×100)	0,0049	0,0090	0,0119	0,0243	0,1592
TER	1,2573	1,4135	1,4354	1,5135	1,4677
TNA ($mil, med)	85,8	78	120,5	98,2	62,2
Age (months)	96,4943	107,8216	131,4756	147,6783	129,7393
Panel B: Value					
Variables	Portfolios formed on EIV				
	Low	2	3	4	High
EWRET (×100)	0,5921	0,6307	0,6089	0,5353	0,4978
4F Alphas (×100)	0,4066	0,4353	0,4105	0,3393	0,2943
EIV (med) (×100)	0,0041	0,0081	0,0138	0,0252	0,0578
UIV (med) (×100)	-0,0010	-0,0027	-0,0060	-0,0145	-0,0404
IV (med) (×100)	0,0031	0,0054	0,0067	0,0078	0,0095
TER	1,2034	1,3661	1,4053	1,4091	1,4469
TNA ($mil, med)	97,6	68,7	83,8	105,6	87,65
Age (months)	94,2281	98,2046	111,719	129,1098	146,7619
Panel C: Large Cap					
Variables	Portfolios formed on EIV				
	Low	2	3	4	High
EWRET (×100)	0,5170	0,4937	0,4938	0,4451	0,5162
4F Alphas (×100)	0,3456	0,3290	0,3388	0,2889	0,3588
EIV (med) (×100)	0,0023	0,0054	0,0097	0,0187	0,0501
UIV (med) (×100)	-0,0006	-0,0018	-0,0040	-0,0107	-0,0348
IV (med) (×100)	0,0015	0,0036	0,0053	0,0062	0,0075
TER	0,9489	1,2152	1,2809	1,3003	1,3290
TNA ($mil, med)	115,75	77,6	84,2	114,9	149,5
Age (months)	95,7969	100,2829	112,0683	138,9008	174,4698

(Continued)

Table 7. Returns for Portfolios Formed on EIV. (Continued)

Panel B: Small Cap					
Variables	Portfolios formed on EIV				
	Low	2	3	4	High
EWRET (×100)	0,7564	0,7818	0,7097	0,6101	0,5659
4F Alphas (×100)	0,5693	0,6075	0,5477	0,4361	0,3892
EIV (med) (×100)	0,0065	0,0120	0,0197	0,0331	0,0796
UIV (med) (×100)	-0,0013	-0,0042	-0,0086	-0,0179	-0,0568
IV (med) (×100)	0,0049	0,0078	0,0100	0,0117	0,0133
TER	1,2551	1,4347	1,4744	1,4988	1,5276
TNA ($mil, med)	69,8	60,45	66,7	95,3	70,1
Age (months)	89,7259	94,0405	104,684	119,3762	138,0651

Panel B: Total					
Variables	Portfolios formed on EIV				
	Low	2	3	4	High
EWRET (×100)	0,5671	0,6221	0,6091	0,5420	0,5433
4F Alphas (×100)	0,3943	0,4555	0,4453	0,3782	0,3827
EIV (med) (×100)	0,0039	0,0089	0,0162	0,0312	0,1256
UIV (med) (×100)	-0,0009	-0,0030	-0,0070	-0,0178	-0,0587
IV (med) (×100)	0,0028	0,0058	0,0078	0,0097	0,0265
TER	1,1212	1,3396	1,3952	1,4053	1,4824
TNA ($mil, med)	98,6	71,3	84,1	116	84,9
Age (months)	95,0833	101,7458	111,4745	136,8708	147,7921

4.8 Some Remarks of the Results

The results thus indicate that there is a positive relation between at least risk-adjusted returns and idiosyncratic volatility when controlling for other variables that explains the cross-section of fund returns. Why risk-adjusted returns produce significant and positive coefficients for EIV is difficult to explain but evidently when removing the return from systematic risk factors a pattern appears. Previous studies get significant results using the gross returns but for some reason I do not get significant results using gross returns. No study in this field use risk-adjusted returns in their Fama-MacBeth regression so it is difficult to make any comparison. Although since both the risk-adjusted returns and the EIV stem from the same factor model (the Carhart model) there might be a “mechanical” relationship. This is of course a problem if it would be true but on the surface at lest this explanation seems far-fetched.

7. Conclusion

Here I discuss the main findings from the results and give my conclusions. I finish by giving some recommendations for future research.

In this study I seek to answer the question of the relation between conditional idiosyncratic volatility and returns for mutual funds. I first show that idiosyncratic volatility does not follow a random walk and thus dismiss Ang et al. (2006) use of lagged idiosyncratic volatility when investigating the relation. This is in line with Fu (2009) who also dismisses that idiosyncratic volatility follows a random walk for stocks. I then follow the methodologies of Chua et al. (2010) and use a AR(2) model to estimate the idiosyncratic volatility. As them I also divide idiosyncratic volatility in to two parts, Expected Idiosyncratic Volatility and Unexpected Idiosyncratic Volatility. The reason for this is that UIV essentially acts as a control variable controlling for noise in realised returns. The relation was investigated primarily through Fama-MacBeth regressions, using both gross returns and the four-factor alpha as dependent variables. Using only EIV and UIV as dependent variables there are a few cases in which the coefficient on EIV is significant. There are however a few instances where the coefficient on EIV is significant but the results are ambiguous. Using all the control variables (Total Expense Ratio, Total Net Assets and Age) the results are more clear at least when using the four-factor alpha as dependent variable. The results show a positive relation between adjusted returns and conditional idiosyncratic volatility. This is in line with both Fu (2009) and Chua et al. (2010). Hence there seems to be a positive relation when controlling for systematic risk. It is however peculiar that a clear positive relation only arises when including all the control variables and especially when using risk adjusted returns. Earlier studies usually give the same results both with and without the inclusion of control variables. However, no study that I am aware of uses risk adjusted returns as dependent variable. As discussed in the result chapter this may be due to some mechanical relationship between the Carhart alpha and the idiosyncratic volatility that is also estimated using the Carhart model. As a further robustness test I did portfolio sorting on EIV in line with table 6 in Fu (2009). This indicates a negative relationship if one disregards that between the lowest EIV portfolio and the second lowest there is an increase in equally weighed returns. No double sorting was performed so the risk-return relation was not controlled for any other effects. In conclusion my findings suggest that there is a positive relationship between conditional idiosyncratic volatility and returns if one disregards the ambiguous results produced when not using any control variables and when gross returns are used instead of risk adjusted returns. The negative relation suggested by the portfolio sorting is weaker evidence than the results from the Fama-MacBeth regression since the regressions take in to account all the control variables.

7.1 Future Research

To my best recollection this is the first study to investigate the relation between idiosyncratic volatility and returns for mutual funds in a proper way. There are however a few shortcomings in my study that could be corrected for. First even though the use of conditional idiosyncratic volatility is appealing it would be better to use a more sophisticated method such as an EGACRH model used by Fu (2009). Secondly the appropriateness of using risk-adjusted returns could be better investigated. The sample could also be extended to include mutual funds from other countries than the U.S.

8. References

Ang, A., Hodrick, R.J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance,* Vol. 61, No. 1, 259-299

Bali, T.G., & Cakici, N. (2008). Idiosyncratic Volatility and the Cross Section of Expected Returns. *Journal of Financial and Quantitative Analysis,* Vol. 43, No. 1, 29-58

Carhart, M.M. (1997). On Persistence in Mutual Fund Performance. *The Journal of finance*, Vol. 52, No. 1, 57-82

Chen, J., Hong, H., Huang, M., & Kubik, J.D. (2004). Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization. *American Economic Review,* Vol. 94, No. 5, 1276-1302

Brooks, C. (2008). *Introductory Econometrics for Finance* (2nd ed.). New York: Cambridge University Press

Chua, C.T., Goh, J., & Zhang, Z. (2010). Expected Volatility, Unexpected Volatility, and the Cross-Section of Stock Returns. *Journal of Financial Research,* Vol. 33, No. 2, 103-123

Eiling, E. (2013). Industry-Specific Human Capital, Idiosyncratic Risk, and the Cross-Section of Expected Stock Returns. *The Journal of Finance,* Vol. 68, No. 1, 43-84

Falkenstein, E.G. (1996). Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings. *The Journal of Finance,* Vol. 51, No. 1, 111-135

Fama, E.F., & French, K.R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance,* Vol. 47, No. 2, 427-465

Fama, E.F., & French, K.R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics,* Vol. 33, No. 1, 3-56

Fama, E.F., & French, K.R. (1995). Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance,* Vol. 50, No. 1, 131-155

Fama, E.F., & French, K.R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance,* Vol. 51, No. 1, 55-84

Fama, E.F., & MacBeth, J.D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy*, Vol. 81, No. 3, 607-636

Fink, J.D., Fink, K.E., & He, H. (2012). Expected Idiosyncratic Volatility Measures and Expected Returns. *Financial Management,* Vol. 41, No. 3, 519-553

Fu, F. (2009). Idiosyncratic Risk and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics,* Vol. 91, No. 1, 24-37

Fuller, W.A. (1996). *Introduction to Statistical Time Series* (2nd ed.). Canada:John Wiley & Sons, Inc.

Granger, C.W.J., & Newbold, P. (1974). Spurious Regressions in Econometrics." *Journal of Econometrics*, Vol. 2, No. 2, 111-120.

Guo, H., Kassa, H., & Ferguson, M.F. (2014). On the Relation Between EGARCH Idiosyncratic Volatility and Expected Stock Returns. *Journal of Financial and Quantitative Analysis,* Vol. 49, No. 1, 271-296

Huang, W., Liu, Q., Rhee, S.G., & Zhang, L. (2010). Return Reversals, Idiosyncratic Risk, and Expected Returns. *Review of Financial Studies,* Vol. 23, No. 1, 147-168

Levy, H. (1978). Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio. *The American Economic Review*, Vol. 68, No. 4, 643-658

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, Vol. 47, No. 1, 13-37

Malkiel, Burton G., & Yexiao Xu. (2002). Idiosyncratic Risk and Security Returns. Unpublished working paper, *University of Texas at Dallas*

Mansfield, E.R., & Helms, B.P. (1982). Detecting Multicollinearity. *The American Statistician,* Vol. 36, No. 3a, 158-160

Markowitz, H. (1952). Portfolio Selection. *The journal of Finance,* Vol. 7, No. 1, 77-91

Merton, R.C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica,* Vol. 41, No. 5, 867-887

Merton, R.C. (1987). A Simple Model of Capital Market Equilibrium With Incomplete Information. *The journal of finance*, Vol. 42, No. 3, 483-510

Newey, W.K., & West, K.D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica,* Vol. 55, No. 3, 703-708

Otten, R., & Bams, D. (2000). European Mutual Fund Performance. *European Financial Management Journal,* Forthcoming

Petkova, R. (2006). Do the Fama–French Factors Proxy for Innovations in Predictive Variables?. *The Journal of Finance,* Vol. 61, No. 2, 581-612

Rohleder, M., Scholz, H., & Wilkens, M. (2011). Survivorship Bias and Mutual Fund Performance: Relevance, Significance, and Methodical Differences. *Review of Finance,* Vol. 15, No. 2, 441-474

Ross, S.A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory,* Vol. 13, No. 3, 341-360

Sharpe, W.F. (1964). Capital asset prices: A Theory of Market Equilibrium Under Conditions of Risk. *The journal of Finance,* Vol. 19, No. 3, 425-442

Stambaugh, R.F., Yu, J., & Yuan, Y. (2015). Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. *The Journal of Finance,* Vol. 70, No. 5, 1903-1948

Vidal-García, J., & Vidal, M. (2014). Seasonality and Idiosyncratic Risk in Mutual Fund Performance. *European Journal of Operational Research,* Vol. 233, No. 3, 613-624

Wagner, N., & Winter, E. (2013). A New Family of Equity Style Indices and Mutual Fund Performance: Do Liquidity and Idiosyncratic Risk Matter?. *Journal of Empirical Finance,* Vol. 21, 69-85

Webster, D. (2002). Mutual Fund Performance and Fund Age. *Available at SSRN 1764543*

Greene, W.H. (2012). *Econometric Analysis* (7th. ed.). Edinburgh Gate, Harlow: Pearson Education Limited