Equity Risks and the Cross-Section of Stock Returns

By
GENG Yan

A Dissertation
Presented to the HEC Montreal
In Partial Fulfillment
Of the Requirements for the Degree

Master of Science in Administration (Finance)
M.Sc. (Finance)

August 2017

Abstract

Since its inception in the 1970s, the capital asset pricing model (CAPM) has dominated the analysis of stock returns. Under CAPM, a stock's beta constitutes its most important risk as it measures the correlation between the fluctuation of security return and return on a market portfolio. Beta coefficients are used not only theoretically but also empirically to estimate expected return. Recently, scholars have focused on the predictive ability of beta measures in empirical tests.

Building on CAPM, this study aims to predict returns one month ahead using various beta measures. We focus on stock returns of mature companies listed on the New York Stock Exchange (NYSE), from January 1965 to December 2016 (624 months in total.) Chapter 1 presents the main objective of this study. Chapter 2 reviews related literature. Chapters 3 and 4 expound data and methodology. Chapter 5 features seven measures of beta coefficients and their inception through daily or monthly returns over different window lengths. We predict future returns using three types of predictive models based on cross-sectional regressions

The study also focuses on the analysis of models based on univariate and multivariate tests. Results indicate that predicting return is dependent on beta coefficients using returns in different observation periods. An analysis of model performance at the industry level shows that industries that are stable, obtain better forecasts.

Key words: risk factors; beta coefficients; return prediction

Sommaire

La théorie de CAPM est proposée dans les années 1970. Puisqu'il est devenu la théorie la plus importante pour analyser les rendements des stocks. Dans la théorie de CAPM, la bêta d'une sécurité est sa caractéristique de risque la plus importante, car elle mesure la corrélation entre la fluctuation du rendement de la sécurité individuelle et le rendement du portefeuille du marché. Les coefficients bêta sont utilisés non seulement dans les implications théoriques, mais aussi dans l'analyse empirique pour estimer le rendement espéré d'un stock. Ces dernières années, plus de travaux ont porté sur la capacité prédictive des mesures de bêta dans les tests empiriques.

Sur la base de la théorie CAPM, le but de cette étude est de prévoir les rendements espérés des stocks à un mois avant. Dans cet article, nous nous concentrons sur la prédiction des rendements des sociétés matures cotées à la Bourse de New York (NYSE). La période d'échantillonnage est comprise entre janvier 1965 et décembre 2016, ce qui correspond à une taille d'échantillon de 624 mois au total. Le premier chapitre présente l'objectif principal de cette étude. Le deuxième chapitre examine la littérature connexe. Ensuite, les chapitres 3 et 4 présentent les données et la méthodologie empirique. Au chapitre 5, qui est le chapitre le plus important, nous construisons sept mesures de coefficients bêta, en utilisant des rendements quotidiens ou mensuels sur différentes longueurs de fenêtres. Les résultats de régression indiquent que les coefficients bêta construits à partir de fenêtres d'observation à long terme (en utilisant des rendements de plus d'un an) sont liés aux performances passées. Pour prédire les rendements futurs des stocks, nous construisons trois types de modèles prédictifs, basés sur des régressions transversales. Les régressions transversales de Fama-MacBeth suggèrent que les prévisions de rendements construits à partir du modèle en utilisant la valeur moyenne des paramètres sur les différentes mesures de bêta pour chaque entreprise ont la meilleure capacité prédictive.

Ensuite, l'étude se concentre sur une analyse des modèles basées sur des tests univariés et multivariés. Le résultat indique que la prédiction des rendements dépend des coefficients bêta en utilisant des rendements dans différentes périodes d'observation (de 21 jours à 5 ans), mais ne dépend pas des rendements anormaux antérieurs. Une analyse de la performance du modèle au niveau de l'industrie montre que les industries stables et moins volatiles obtiennent de meilleures prévisions, telles que l'industrie de la construction et l'industrie des transports, de la communication, de l'électricité.

Mots clés : facteurs de risque ; Coefficients bêta ; Rendements de la prédiction

Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor, Professor Mathieu Fournier, for his help and support along with his valuable guidance at every stage of this thesis. I am also thankful for his kindness and patience. His keen and vigorous academic observations not only enhanced this thesis but also gave me insight for my future studies.

At the same time, I would like to express my appreciation for the tireless efforts made by Mr. Mohamed Jabir in the LACED lab with regard to the preparation of the data for this thesis. I would also like to thank Brenda Prai and Shuxin Li for all their kindness and help in correcting my paper and providing valuable comments and suggestions.

Furthermore, I extend my thanks to my literature teacher, who gave me pertinent knowledge and information from the world of literature for this paper. Last but not least, I would like to thank all my family and friends for their encouragement and support.

Contents

I. Introduction

In capital markets, the study of investment risk has always been the focus of attention in academia and in the industry. With the establishment and development of modern portfolio theory, economists have proposed a series of theoretical and practical methods to measure risk. Among the various risk measures proposed, a stock's beta is one of the most widely used risk indicators of an investment.

As early as 1952, Markowizt pioneered the quantification of risk measurement and its relation to expected returns. On the basis of this theory, William Sharpe founded the capital asset pricing model (CAPM) in 1964. The CAPM model became the most important equilibrium model in modern capital market theory partly due to its parsimony and intuitive appeal. This model underlines the link between an asset's risk, as measured by a security co-movement with the market portfolio, and its expected return. In recent years, the theory of capital market equilibrium based on the CAPM model has become the mainstream theory in the field of financial economics; it has also deeply influenced the research and practice of asset pricing. In the CAPM model, asset risk is divided into two parts: systemic risk and idiosyncratic risk. In the model, beta is considered as the quantitative measure of a stock systemic risk exposure. Beta captures the exposure of an asset (securities) to the fluctuations of the overall market portfolio returns. The essence of the CAPM model is to summarize systematic risk exposure of a stock to its beta coefficient (β). In theory, non-systematic risk should not be compensated as it can be diversified away. In contrast, systematic risk cannot be diversified and thus should be compensated. Because the beta coefficient summarizes the systematic risk exposure of a given security, it provides the basic information of asset selection and systematic risk management. For these reasons, the beta coefficient is the most widely used indicator of risk measurement.

The results of several empirical tests from various studies (Sharp, 1964; Black Scholes, 1972; Blum and Friend, 1973; Fama and MacBeth, 1973) support the CAPM theory. These early results provide evidence that the CAPM model holds sway and that stock beta is positively correlated to the average returns of these stocks. However, in 1977, Roll questioned the empirical test by asserting that the market portfolios used to test the CAPM theory are not valid and that the standard market index used to test this theory is not valid to test the model. Thus, the result from past empirical tests could not be used to verify the theory. After the 1980s, various studies (Roll, 1977; Reinganum, 1981; Shapiro, 1986; Fama and French, 1993) posed the problem of the validity of CAPM based on various empirical analysis: these showed weak support of the CAPM theory.

The trend in recent years has been to divide studies on beta into various strands of literature. One strand focuses on the study of the variability of beta-coefficient over time. The main results of these studies indicate the dependence of beta estimation on the method used to estimate it. Other studies have focused on studying the predictability of beta for stock returns. Because beta estimates depend on the length of the estimation window used to estimate it as well as on the frequency of the returns (i.e., monthly versus daily), this study analyzes the way various estimates of stock conditional betas can be combined to help improve return predictability in the cross-section.

Several previous works set out to estimate beta dynamically and conditionally through the lance of time-series CAPM regressions. In our study, we focus mainly on the following questions: How can we combine various estimates of firm betas to improve on-return forecasting in the cross-section? Which estimates of beta have the most predictability for stock returns? Armed with time-series of betas estimated from different window lengths and return frequencies, how can one optimally combine these estimates for forecasting purposes?

In our study, we construct seven beta estimates every month and for every firm from CAPM time-series regressions using respectively the past 21 daily returns, 63 daily returns, 252 daily returns, 2-year monthly returns, 3-year monthly returns, 4-year monthly returns, and 5-year monthly returns. Based on these beta estimates, we then run monthly predictive cross-sectional regressions of the stocks' next month returns on betas to estimate the loadings on these betas. Armed with the monthly loadings on betas (i.e., monthly prices of beta risk) and the monthly estimates of betas, we then study various predictive models for next month returns on stock that combine the monthly beta estimates and the past loadings on these betas. Through the lance of Fama-MacBeth cross-sectional regressions, we provide evidence that a predictive model, whichtakes the average of past loadings on betas times the betas estimated on the last month, outperforms the other predictive models we have considered in order to predict the cross-section of next month returns out-of-sample. The model we develop delivers a high adjusted R-squared, which is encouraging.

To better understand the information content of the various beta estimates, our analysis also relies on univariate cross-sectional regressions on each of the seven estimates of beta we use to benchmark the performance of the full model which combine the seven estimates for forecasting purposes. We also consider the predictive ability of the model when the intercept (α) is omitted. Our results suggest that most of the predictive ability of the model is contained in the betas and lagged prices of beta risk but not in the constant (α) for common stocks on NYSE. When considering various industries separately to assess the model forecast out-of-sample performance, we find that the model performs particularly well for industries that are relatively stable and less volatile (ex. construction, transportation, communications, electric, gas, and sanitation.)

1.1 The CAPM model

All the models are based on a set of underlying assumptions. The CAPM model focuses mainly on the relationship between the expected return of individual assets relative to the market portfolio. The assumptions of the CAPM model are the following (Zhou, 2013)[1]:

1. All investors are risk averse, and all kinds of investors will use the expected return of assets and the variance of assets to measure the benefits and risks.

2. Capital markets pose no obstacles and impose no transaction costs. Furthermore, with the number of asset transactions being easily broken down, any investor can buy any assets in accordance with market prices, whereas all financial instruments can also be listed and entered into the market for open trading. Thus, investors can, according to their own wishes, buy any products in the financial markets.

3. Investors during the investment decision-making process usually consider single income and risk, their investment period being the same.

4. All investors get their information from the market, which leads us to conclusion that their view on risk and income are the same.

5. Investors can be unrestricted to borrow and lend at risk-free rates.

---

[1] Zhou (2013): "Based on mobile platform of mobile group customer marketing strategy optimization."

6. There is no tax on consideration for investors in securities trading or asset selection, and there is no personal income tax. The dividend or capital gains obtained from investments will not affect the investment decision-making process.

7. All investors are price recipients: they can only passively accept the market price.

Obviously, the above assumptions reduce the complex capital market into a simple and perfectly competitive market, with every investor faced with the same effective set of variables. Thus, with its shortcomings, CAPM is not a perfect model. First, CAPM assumptions are not present in the real capital market. For example, transaction costs, taxes and consulting fees must exist when trading on the capital market, which indicate that the market is incomplete and does not meet complete market assumptions. In other words, the borrowing rate in real life is greater than the loan interest rate; in contrast, CAPM assumes that the borrowing rate is equal to the risk-free rate. Second, investors are concerned about future changes in the value of securities, and beta coefficient is measured in past returns. Moreover, in practice, market portfolios and risk-free assets do not necessarily exist.

Because the CAPM model is problematic, several previous works have abandoned some assumptions and added influencing factors into the model in order to make it more suitable for empirical tests such as the Intertemporal Capital Asset Pricing Model (ICAPM) proposed by Robert Merton (1973) and Fama-French three-factor model proposed by Fama and French in 1996. The former assumes that security returns are normally distributed over multiple time periods, whereas the latter considers the size of firms and the M/E factor as explanatory variables in the CAPM model.

1.2 Conditional beta

Asset pricing tests based on CAPM often assume constant betas. Several works consider dynamic conditional beta which is estimated by regressions using rolling-windows of a given length and return frequency.

In calculating beta, various frequencies of return can be considered (ex. daily, weekly, monthly, quarterly, semi-annually or annually.) Several studies have found that the estimates of beta are dependent on the frequency of the return used to estimate it, which is referred to as "interval effects."

Jensen (1969) in his "Risk, Capital Asset Pricing and Portfolio Evaluation" article argue that the estimated betas are independent of the return frequency and length of the estimation window. More recently, many empirical works (Fama, 1970; Levhavi and Levy, 1977; Schwartz and Whitcomb, 1977; Saniga, McInish and Gouldey, 1981; Handa, 1989; Kothari, 1995) have argued otherwise and found that the length of the observation periods of returns did affect the estimated beta.

Hwawaini (1983) estimated 21 company stocks from January 1970 to December 1973 with returns in different observation windows (one-month returns, three-week returns, etc.) to calculate beta. The author found that the estimation windows have a huge influence on beta. Hwawaini explains that the movement of the securities' price may not be synchronized with the market tendency, which may explain why various estimation windows result in differences in betas.

1.3 Study overview

The focus of this study is to construct a general predictive model based on the CAPM model that combines estimates of betas obtained from various rolling-window lengths and return frequencies. This study attempts to predict stock returns of mature companies listed on the New York Stock Exchange (NYSE) while considering the NYSE as a market index. The total number of companies to be analyzed is 3829. The sample period for observation is from January 1965 to December 2016; there are thus 624 months for analyzing.

First, we construct seven beta coefficients and study their properties. At the end of each month, we compute the betas for each stock listed on NYSE, which represents the short-run risk factors using daily returns from the past 21 days (21d), 36 days (36d) and 252 days (252d), and the long-run risk factors using monthly returns of the past two years (24m), three years (36m), four years (48m) and five years (60m.) Based on time-series regressions, we obtain seven betas on the last date of each month. To better understand the relation and auto-correlation of the betas estimated from various estimation window lengths and return frequencies, we use an autoregressive model (AR (1)) to test the serial correlation of these estimates in the time-series. The results indicate that the risk factors with long-run observation windows, such as $\beta^{252d}, \beta^{24m}, \beta^{36m}, \beta^{48m}, \beta^{60m}$ being calculated by more than oneyear returns, are more likely to depend on past performance of conditional factor loadings than on other elements.

Second, we intend to predict forward stock return one-month ahead based on the information available as of date t. We regress the cross-section of next month returns on the lagged beta estimates based on the last 60 months up to month t. From these predictive regressions, we obtain the monthly intercept estimates α as well as the loading on the betas, γ. Based on the monthly α and γ, three types of prediction models for next month stock returns are constructed and tested based on Fama-MacBeth cross-sectional regressions. Note that for each model, the model forecast at time t rely only on the information available up to time t. Our tests are thus out-of-sample. Our results suggest that Model II, for which we take a simple average of the time-varying alphas (α) and the seven gammas (γ) to predict returns in the next month (at t+1), obtains the best performance based on adjusted R-Squared.

Third, we also consider univariate analysis and multivariate analysis of the predictive ability of the betas., Our initial results indicate that there exist important serial correlations among risk factors (i.e., the betas.) The results of univariate analysis further indicate that the forecast ability increases with the length of observation windows. For instance, the prediction model with the individual beta calculated with five-year monthly returns is the most pertinent for future realized returns. Multivariate analysis shows that the best predictive model formed from the combination of the seven risk factors (i.e., the betas) constructed from different window length and without the constant factor (i.e., α), obtains the highest adjusted R-Squared.

Fourth, to analyze the predictive ability of the models for each industry, we have classified securities into eight industries. Our results indicate that the prediction of the model is more appropriate for industries that are relatively stable such as those in the field of construction transportation, communications, electric, gas and sanitation.

Ultimately, the aim of this study is twofold: first, conducting an in-depth analysis of the difference between the betas according to both their time-variations and their predictive ability; second, we develop a new way to combine various beta estimates to predict returns out-of-sample. Our main result suggests that valuable information is embedded in the past betas of various frequencies when forecasting the cross-section of stock returns.

The rest of the paper is organized as follows: Section 2 examines the previous literature; Section 3 illustrates the data set analyzed in our study; Section 4, the most important part in our study, introduces the main variables and methodology used to implement each model in order to forecast future stock returns in the cross-section; Section 5 presents the results; finally, Section 6 summarizes the work and offers some conclusions.

II. Literature Review

Sharpe (1964), Lintner (1965), and Mossin (1966) have developed individually the model of the capital asset pricing (CAPM model.) The main prediction of this theory is that the expected return on a security should be positively related to its exposure to the market portfolio as captured by its beta.

Over time, the CAPM model has become critical for financial analysts and investors, because of its intuitive appeal and applications. Indeed, estimates of betas are required in many financial applications from portfolio selection to systematic risk management. That being said, the model shows its flaws when applyied in empirical tests. Perhaps the biggest drawback of the CAPM is the weak link between beta and stock future returns as documented in various studies.

In recent years, researchers have undertaken many studies to find a better way of predicting stock returns. The result is three main approaches to developing the CAPM model. The first is a focus on improving the model estimation so that CAPM achieves a better fit in the cross-section and data test. Second, many researchers have worked on improving the accuracy of beta estimation. Third, following this line, researchers have considered accounting for time-varying betas through the lance of recursive regression models to allow for conditional betas estimates.

2.1 CAPM model

In some early works, Sharp (1964) first tested the relation between stock returns and risk factors. By using the average annual returns of 34 mutual funds during ten years from 1954 to 1963, he demonstrated empirically that the coefficient of correlation of annual returns and standard deviation is relatively high and about 0.8. This indicates a positive relation between betas and average returns.

Fama-Macbeth (1973) use common stock listed on the NYSEduring the period from 1926 to 1968. To assess the relation between systematic risk and returns, they proposed a new testing method, a two-step regression[2]. The result of the Fama-Macbeth regression indicates that stock returns are positively related to market risk factors. Moreover, this study provides further evidence that the CAPM model holds in the cross-section.

Black, Jensen, and Scholes (1972) conducted more formal empirical tests to assess the model's performance. Their results indicate that while there is a positive cross-sectional relation between beta and average returns, the relation between beta and average return remains too flat. Stambaugh offered similar evidence (1982): through the use of time-series regressions, he showed that the intercepts of excess asset return on the excess market return are positive for assets with low systematic risk (beta) and negative for assets with high systematic risk (beta). Thus, even if some study finds that CAPM holds unconditionally, others provide evidence that the relation between beta and average returns is relatively weak.

Rosenberg, Reid, and Lanstein (1985) collected empirical data to assess the model's empirical validity and practicality; unfortunately, they followed a CAPM model methodology that was used in previous studies and was found to be no longer appropriate.

[2] Fama–MacBeth regression: The method works with multiple assets across time (panel data.) The parameters are estimated in two steps: First regress each asset against the proposed risk factors to determine that asset's beta for that risk factor. Then regress all asset returns for a fixed time period against the estimated betas to determine the risk premium for each factor.

Several researchers expanded the original hypothesis to more general economic situations and came up with improved prototypes that built on the CAPM model. Robert Merton (1973) proposed the Intertemporal Capital Asset Pricing Model (ICAPM) which assumes that security returns are normally distributed over multiple time periods. Indeed, the ICAPM factors in investor participation over the long term. Fama-French (1992) further studied CAPM and found that its abnormal returns could be explained by other factors such as size and M/E. Thus, they added SMB and HML into the time-series regression and proposed a new model now known as the three-factor model.

Previous studies did plenty of empirical tests to compare the CAPM model and the Fama-French model. They focused on both multi-factor and single-factor regression as possible explanations and tried to discover the advantages of one model over the other. Together, these studies argue that accounting for missing factors is important for pricing purposes.

2.2 Conditional risk factor

Asset pricing tests often assume that betas are constant over time. The Fama–MacBeth procedure was applied to estimate and test unconditional asset pricing models. Unconditional beta tests have led to the rejection the CAPM even if it holds perfectly, period by period. Several works have proposed the dynamic conditional beta which is an approach to estimate regressions that accounts for time-varying parameters.

Robinson (1989) first studied the relationship between returns and risk factors under the assumptions of time-varying coefficients with seasonal patterns and locally stationary variables. However, the analysis is based only on an ordinary least squares estimation method for a single equation regression model; neither a two-step procedure nor a multi-equation model is considered in this study.

Andrew and Dennis (2010) focused on the analysis of time-varying factors and conditional long-run alphas and risk factors based on nonparametric methodology. They reject the null hypothesis that an asset's expected excess return is equal to zero after controlling for conditional betas.

Robert (2016) used non-nested tests and several approaches including a novel nested model to reject the null hypothesis that betas are constant by estimating regressions with time-varying parameters.

To further test conditional CAPM, Lewellen and Nagel (2006) proposed rolling-window regressions, where they use short-term observation windows, from one quarter to one-year rolling-window length and high frequency returns, to estimate the coefficients of parameters. The null hypothesis with the mean value of abnormal returns being zero has been rejected by the estimate of time-varying risk factors and pricing errors associated with the portfolios considered by this study.

Not directly related to the study of CAPM, Corsi (2004) proposed a new model to forecast the time series behavior for volatility. The purpose of his work is to obtain a relatively parsimonious conditional variance model which is easy to estimate and based on realized variances of various horizons and frequencies. While not directly related to CAPM, this study is the motivation for us to capture firm systematic risk exposure using various measures of beta constructed from return of different frequencies and various window estimation lengths.

2.3 Study of risk factors

Based on the characteristics of time-varying for risk factors, several studies focused on searching for other influencing factors which affect the performance of firms. Market exposure leads to different betas when measured across different return frequencies.

Thomas, Christopher and Jonathan (2013) studied the effect of the frequency on betas in empirical tests. They argue that there exists a clear relationship between the frequency dependence of betas and proxies for opacity. The results from their tests show that CAPM may be an appropriate asset pricing model at low frequencies but that additional elements must be factored in at high frequencies.

Zhang (2003) proposed a nonparametric measure of realized risk factor loadings in a multifactor pricing model by regressing intraday returns with the Fama-French three factor model. He classified portfolios in various industries to assess the effect of industry peculiarities and confirmed a relationship between returns estimated by risk factors and industry distribution. Newey-West test could be used on Fama-MacBeth cross-sectional regression to correct heteroscedasticity and autocorrelation in the residuals of these regressions.

III. Data

3.1 Sample

The choice of sample follows two principles: sufficient sample size and appropriate data frequency.

3.1.1 Selection of Market index

Our study focuses on predicting stock returns of companies listed on the New York Stock Exchange (NYSE.) In order to benchmark stock performance based on the market, the price and return data used in the analysis contain all common stock of publicly traded companies listed on the NYSE, which, when combined, should mirror the market portfolio. Consequently, the NYSE is considered as the market index in this study. Differentiating characteristics of the NYSE include stricter index rules as well as higher listing fees compared to other market exchanges. A company must have issued at least one million shares with a total worth of over $100 million and a revenue of over $10 million during the past three years to be listed on the NYSE. Considering its listing requirements, the NYSE contains the largest total market capitalization among all US exchanges. Thus, the NYSE has always listed companies with the largest total market capitalization.

3.1.2 Selection of study period

Data applied to the short-term analysis are daily returns from 1960.01.01 to 2016.12.31 and monthly returns of selected companies. The reasons for this application is due to two factors:

first, a large sample period ensures sufficient sample size and reflects variable market conditions like expansions and recessions; second, the use of daily and monthly returns should allow to capture the systematic risk of firm at different frequencies.

3.1.3 Selection of sample analyzed

The criteria used to filter the database to obtain our final sample also include delisted companies. This is to avoid a selection bias which could render inaccurate the results and ignore companies with less than 5 years of historical returns. The source of the latter filter is our estimation methodology since estimates of 5-year beta are required. We refer to Table 3.1 and Figure 3.1 for relevant information on the number of companies selected each year.

3.2 Data collection

To obtain sufficient observations for our statistical analysis, we set the analysis period from January 1965 to December 2016. Table 3.1 presents the summary the statistics for the number of stocks listed on the NYSE. The average number of companies for each year in our sample is about 1461 while the original total number of companies is 5280.

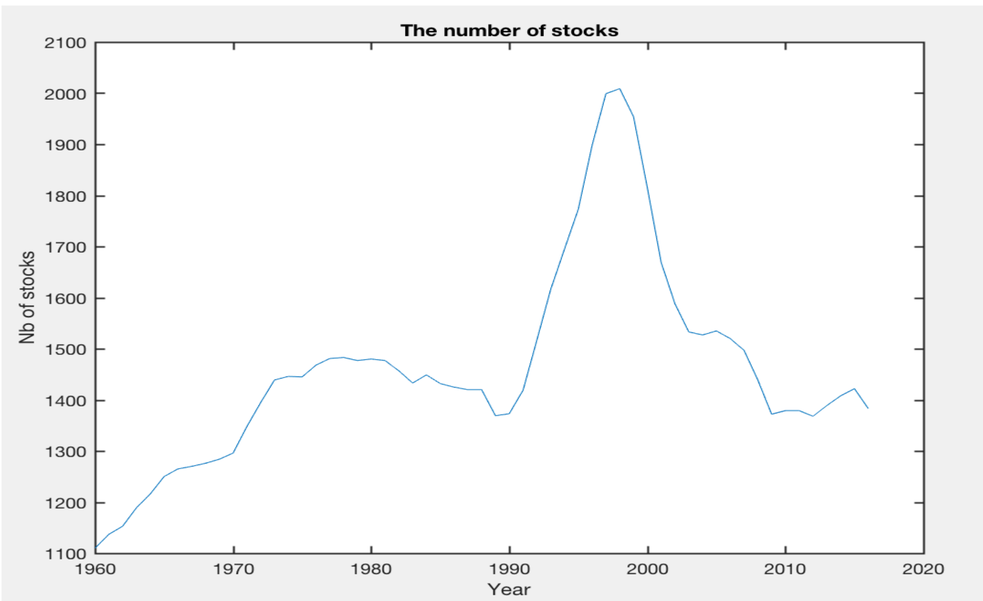
Table 3.1: Number of Stocks listed on the New York Stock Exchange

	Mean	Min	Max	Std. errors	Nb of observations
Nb of stocks	1461	1110	2009	14.029	57
[Date]		[1960]	[1998]		

Notes: The number of observations is 57 years due to the analysis period being from January1965 to December 2016. The standard deviation of the number of stocks is 14.029, which indicates the great variation among the number of stocks for each year.

From the graph in Figure 3.1, we can see that the number of stock listed on the NYSE before 1970 is around 1200, spiking during the 2000s and decreasing in the later years. Starting in 1990, the new listing of firms on the NYSE accelerates and reaches a historical high of 2009 in 1998, which means that more and more companies are expanding their operations to meet the NYSE listing requirements. The number of stocks then tends to be stable, hovering around 1400 every year.

Figure 3.1 The evolution of companies through time



3.2.1 Data sources

The CRSP (Center for Research in Security Prices) provides daily and monthly information including stock returns as well as index information (market returns), and the distribution and number of outstanding shares. Firm name codes and SIC codes are obtained from Compustat. We use the recently published CRSP/Compustat[3] merged dataset, which up to now has been viewed as being the most successful link between CRSP and Compustat.

3.2.1.1 Abnormal return/abnormal price

Price is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, then the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price. Thus, the price carrying the negative sign should be excluded; in addition, returns listed as "NaN" should be deleted for the same reason.

3.2.1.2 Industry classification

The industry classification is based on the newly available North American Industry Classification System (NAICS) which has been widely used in previous studies (Breeden, Gibbons, and Litzenberger 1989; Campbell and Mei 1993; Ang, Chen and Xing 2002, etc.) Industries are divided into 10 groups depending on the first two digits in the SIC code. Table 3.1 shows the standard classification of industry and the number of companies in each industry during our observation periods for this study.

---

[3] https://wrds-web.wharton.upenn.edu/wrds/index.cfm

Table 3.1: Standard industry classification

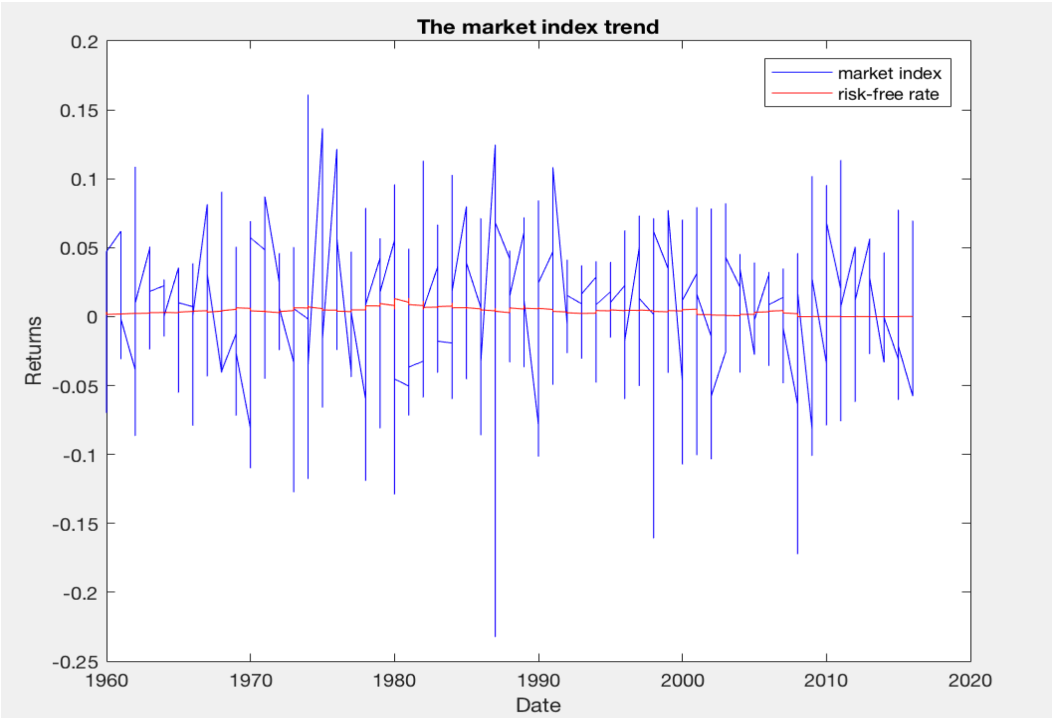
SIC code	Industry name	Nb of companies
<10	Agriculture, Forestry, and Fishing	18
10-14	Mining	262
15-17	Construction	80
20-39	Manufacture	1898
40-49	Transportation, Communication, Electric, Gas, Sanitary	481
50-51	Wholesale trade	384
52-59	Retail trade	365
60-67	Finance, Insurance, Real Estate	660
70-89	Service	483
90-99	Public administration	12

Notes: According to the North American Industry Classification System (NAICS), stocks listed on the NYSE could be classified into ten parts of industry.

3.2.2 Market information

Based on the data provided by the CRSP, Figure 3.2 presents the market index trend during our 56-year period. From the figure, we can see that the risk-free rates at monthly intervals are mostly constant, whereas the market index returns vary significantly through time.

Figure 3.2 The evolution of the market index through time



Note: The figure shows the evolution of market index and the risk-free rate during the period from January 1960 to December 2016. Data is collected from Fama-French factors and at one-month frequencies, while returns are recorded on the last day of each month.

3.2.2.1 The risk-free interest rate

Regarding the risk-free interest rate, we select the yield-to-maturity of one-month treasury bonds. The daily and monthly rates are both available in the database of the Fama-French factors provided by the Wharton research data service as they are used to calculate the market excess return in the database.

IV. Methodology

Theoretically, the beta coefficient is the key parameter in the standard CAPM model. As discussed in the second part, its stability has influence on the theoretical value of the CAPM model. In this section, we introduce seven constructed variables which represent the systematic risk factors constructed from different observation windows. These should account for a firm's exposure to the market index for various frequencies. Then, we assume that the future realized beta in the next month has relation with the seven constructed betas. We control its veracity by using OLS predictive regressions in Section 4.2. In Section 4.3, we turn our attention to predicting the forward return one-month ahead conditionally on the information available on month t. In this endeavor, we lay out three types of prediction models and develop general conditional estimators and their distributions by using cross-sectional regression. Section 4.3.5 develops a test to select the optimal prediction model for estimating the forward returns which account for autocorrelation and heteroscedasticity. Finally, we discuss further the conditional prediction of the models based on univariate and multivariate analysis while considering stocks industry by industry.

4.1 Variables

Sharp, Lintner (1965) used the excess returns on individual stocks to test the CAPM model. The model is as follows:

$$R_i - R_f = \beta_i (R_m - R_f) \tag{4.1.1}$$

in which, R_i is the daily return of stock i, R_m is the daily return of the market, R_f is daily risk-free rate. From the (4.1), β_i could be estimated by the slope of the market under the OLS method. In truth, this estimation method implies a hypothesis that the beta coefficient is a constant over the estimated period, and it does not change over time. Therefore, the beta coefficient obtained by this method is a static estimate.

4.1.1 Creating new beta-coefficients

In order to avoid the problems mentioned above, different methods have been used to define conditional betas.

Defining betas as the above-mentioned beta-coefficients estimated from regression over a given window length and return frequency, the proposed model can capture the effect of the length of period on the firm systematic risk when considering the past performance in the different observation periods. To simplify, for each stock, we consider the model with only seven betas components respectively $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ and β_7.

β_1: By using the past twenty-one days' daily return (21d), the estimation window is 21 days.

$$\beta_{i,t-1}^{21d} = \frac{cov(r_i, r_m | r_i, r_m \in (t-22, t-1)}{var(r_m)} \tag{4.1.2}$$

β_2: By using the past three-month' daily returns (63d), the estimation window is 63 days.

$$\beta_{i,t-1}^{63d} = \frac{cov(r_i, r_m | r_i, r_m \in (t-64, t-1)}{var(r_m)} \tag{4.1.3}$$

β_3: By using the past one-year's daily return (252d), the estimation window is 252 days.

$$\beta_{i,t-1}^{252d} = \frac{cov(r_i, r_m | r_i, r_m \in (t-253, t-1)}{var(r_m)} \tag{4.1.4}$$

β_4: By using the past two-year's monthly return (24m), the estimation window is 24 months.

$$\beta_{i,t-1}^{24m} = \frac{cov(r_i, r_m | r_i, r_m \in (t-25, t-1)}{var(r_m)} \tag{4.1.5}$$

β_5: By using the past three-year's monthly return (36m), the estimation window is 36 months.

$$\beta_{i,t-1}^{36m} = \frac{cov(r_i, r_m | r_i, r_m \in (t-37, t-1)}{var(r_m)} \tag{4.1.6}$$

β_6: By using the past four-year's monthly return (48m), the estimation window is 48 months.

$$\beta_{i,t-1}^{48m} = \frac{cov(r_i, r_m | r_i, r_m \in (t-49, t-1)}{var(r_m)} \tag{4.1.7}$$

β_7: By using the past five-year's monthly return (60m), the estimation window is 60 months.

$$\beta_{i,t-1}^{60m} = \frac{cov(r_i, r_m | r_i, r_m \in (t-61, t-1)}{var(r_m)} \tag{4.1.8}$$

Where r_i is the daily excess return of stock *i*; r_m is the daily excess return of the market. We exclude stock missing more than 12 days' records in a given month.

To estimate conditional betas at time t, we require at least 60 months of prior data on stock returns. This results in the loss of the 60 initial observations. To capture mispricing, we estimate the above equations with intercepts. All the seven beta-coefficients are recorded as the performance of month t_1 for stock *i*. At a monthly frequency from 1965.01.31 to 2016.12.31, we repeat the calculation for each stock and each month.

4.2 Constructing return predictions based on Beta

We assume a hierarchical process where at each level of the cascade, the future beta-coefficient depends on the beta-coefficient at different levels of the cascade (i.e., the next longer horizon beta-coefficient.) The model of the future systematic risk factors (beta-coefficient) at each level of the cascade (or time scale) is assumed to be a function of the risk factors experienced with different observation windows. In this study, we try to test for each company the relation of future betas with different observation windows and realized betas by AR (1), which corresponds respectively to 21 days, 63 days, 252 days, 24 months, 36 months, 48 months, 60 months. The model reads as follows:

$$E\left[\beta_{i,t+1}^{j}\right] = c_1 + \gamma_i^{21d}\beta_{i,t}^{21d} + \gamma_i^{63d}\beta_{i,t}^{63d} + \gamma_i^{252d}\beta_{i,t}^{252d} + \gamma_i^{24m}\beta_{i,t}^{24m} + \gamma_i^{36m}\beta_{i,t}^{36m} + \gamma_i^{48m}\beta_{i,t}^{48m} + \gamma_i^{60m}\beta_{i,t}^{60m} \qquad (4.2.1)$$

Equation (4.2.1) can be seen as a seven-factor stochastic beta-coefficients model, where j represents the observation periods of betas as defined above, and the factors are betas viewed at different frequencies by using the past realized returns.

4.3 The model for return prediction

We are also interested in the estimates of forward stock return one-month ahead of date t. To examine this data, we now present three types of models in the following sections.

4.3.1 Cross-sectional regressions

First at all, we regress the cross-section of future realized stock returns from 1965 to 2016 on the seven betas defined above. From each month, we obtain one intercept and seven coefficients which can be interpreted as the prices of beta risk.

$$R_{t+1} = \alpha_t + \gamma_t^{21d}\beta_t^{21d} + \gamma_t^{63d}\beta_t^{63d} + \gamma_t^{252d}\beta_t^{252d} + \gamma_t^{24m}\beta_t^{24m} + \gamma_t^{36m}\beta_t^{36m} + \gamma_t^{48m}\beta_t^{48m} + \gamma_t^{60m}\beta_t^{60m} + \varepsilon_t \quad (4.3.1)$$

The subscript t refers to monthly measure, from 1 to 624 representing January 1965 to December 2016; C is from 1 to 3289 representing the proxy code of each stock; $\alpha_t = (\alpha_t^1, \alpha_t^2, \dots, \alpha_t^C)' \in \mathbb{R}^C$ is the vector of conditional alphas across stocks 1, …, C; $\beta_t = (\beta_t^1, \beta_t^2, \dots, \beta_t^C)' \in \mathbb{R}^{7*C}$ is the corresponding matrix of defined betas; the dependent variable is the realized return one-month ahead of the date t; the errors have been collected in the vector ε_t.

4.3.2 Model I

To estimate forward returns in the next month, we assume that multipliers (γ_s) have only been affected by the whole market and the time-varying. Thus, we suppose that in Model I the coefficients of risk factors for each company at date t are same as that from the cross-sectional regressions (4.3.1) on month t. Thus, using last month estimates of α_t and $\gamma_{j,t}$ for stock i from equation (4.3.1), we could write Model I as (4.3.2.1) and compute the root mean squared error for the model.

$$\hat{R}_{t,t+1}^{i,M_1} = \alpha_t + \gamma_t^{21d}\beta_{i,t}^{21d} + \gamma_t^{63d}\beta_{i,t}^{63d} + \gamma_t^{252d}\beta_{i,t}^{252d} + \gamma_t^{24m}\beta_{i,t}^{24m} + \gamma_t^{36m}\beta_{i,t}^{36m} + \gamma_t^{48m}\beta_{i,t}^{48m} + \gamma_t^{60m}\beta_{i,t}^{60m} \quad (4.3.2.1)$$

where $\hat{R}_{t,t+1}^{i,M_1}$ denotes the prediction of Model I for stock i's next month return, conditional on the information available as of month t.

4.3.3 Model II

The second model makes use of the average of alphas and gammas estimated. We collect the time-varying alphas (α) and gammas (γ) from (4.3.1) and note that the average value of alphas for stock i as $\overline{\alpha}^{i}$ (4.3.3.1) and that of seven gammas for stock i as $\overline{\gamma}_{i}^{21d}, \overline{\gamma}_{i}^{63d}, \overline{\gamma}_{i}^{252d}, \overline{\gamma}_{i}^{24m}, \overline{\gamma}_{i}^{36m}, \overline{\gamma}_{i}^{48m}, \overline{\gamma}_{i}^{60m}$ (4.3.2.2) respectively.

$$\overline{\alpha}_{i} = \frac{1}{T_i}\sum_{t=1}^{T_i} \alpha_{i,t} \tag{4.3.3.1}$$

$$\overline{\gamma}_{i}^{j} = \frac{1}{T^i}\sum_{t=1}^{T_i} \gamma_{i,t}^{j} \tag{4.3.3.2}$$

In the previous equations, T_i is the number of months listed on the stock exchange for company i during the period January 1965 to December 2016; t represents a given month.

By defining $\beta_{i,t}^{j}$, where j represents 21 days, 63 days, 252 days, 24 months, 36 months, 48 months, and 60 months, respectively, we can write Model II for i stock returns as (4.3.3.3)

$$\hat{R}_{t,t+1}^{i,M_2} = \overline{\alpha}^{i} + \overline{\gamma}_{1}^{i}\beta_{1}^{i,t} + \overline{\gamma}_{2}^{i}\beta_{2}^{i,t} + \overline{\gamma}_{3}^{i}\beta_{3}^{i,t} + \overline{\gamma}_{4}^{i}\beta_{4}^{i,t} + \overline{\gamma}_{5}^{i}\beta_{5}^{i,t} + \overline{\gamma}_{6}^{i}\beta_{6}^{i,t} + \overline{\gamma}_{7}^{i}\beta_{7}^{i,t} \tag{4.3.3.3}$$

4.3.4 Model III

The alphas (α) and gammas (γ) from (4.3.1) have been observed for stock i at time points $0 < t_1 < t_2 < \dots < t_n < T$. We have recorded vector α^{i} and β_{j}^{i} for stock i as $[\alpha_{t_1}, \alpha_{t_2}, \dots, \alpha_{t_n}, \ \alpha_{T}]$ and $[\gamma_{i,t_1}^{j}, \gamma_{i,t_2}^{j}, \gamma_{i,t_3}^{j}, \gamma_{i,t_4}^{j}, \gamma_{i,t_5}^{j}, \gamma_{i,t_6}^{j}, \ \gamma_{i,t_7}^{j}]$, respectively.

To check the characteristics for alphas and gammas for each stock over time, we conduct the following study. Suppose that each stock possesses coefficients of alphas and gammas varying over the time and with potential autocorrelation. Using the AR (1) model (4.3.4.1), current alphas and gammas for stock i could be estimated by themselves in the past (4.3.4.2).

$$\alpha_{i,t+1} = \omega_0 + \omega_1 \alpha_{i,t} + \epsilon_t \qquad (4.3.4.1)$$

$$\gamma_{i,t+1}^{j} = \delta_o^j + \delta_1^j \gamma_{i,t}^{j} + \epsilon_t \qquad (4.3.4.1)$$

$$\hat{\alpha}_{i,t+1} = \widehat{\omega}_0 + \widehat{\omega}_1 \alpha_{i,t} \qquad (4.3.4.2)$$

$$\hat{\gamma}_{j,t+1}^{i} = \hat{\delta}_o^j + \hat{\delta}_1^j \gamma_{i,t}^{j} \qquad (4.3.4.2)$$

where j is the observation windows, correspondent to 21 days, 63 days, 252 days, 24 months, 36 months, 48 months, and 60 months.

Suppose T_i is the number of month listing on the stock exchange for company i during the period January 1965 to December 2016, alpha-estimates and gamma-estimates for stock i could be calculated from time 1 to T. By defining $\beta_{i,t}^{j}$, where j is the observation windows, correspondent to 21 days, 63 days, 252 days, 24 months, 36 months, 48 months and 60 months, Model III could be written for i stock returns during period T as (4.3.4.3)

$$\hat{R}_{t,t+1}^{i,M_3} = \hat{\alpha}_i + \hat{\gamma}_{i,t}^{21d} \beta_{i,t}^{21d} + \hat{\gamma}_{i,t}^{63d} \beta_{i,t}^{63d} + \hat{\gamma}_{i,t}^{252d} \beta_{i,t}^{252d} + \hat{\gamma}_{i,t}^{24m} \beta_{i,t}^{24m} + \hat{\gamma}_{i,t}^{36m} \beta_{i,t}^{36m} + \hat{\gamma}_{i,t}^{48m} \beta_{i,t}^{48m} + \hat{\gamma}_{i,t}^{60m} \beta_{i,t}^{60m} \qquad (4.3.4.3)$$

where $\hat{\alpha}_i = \omega_0 + \omega_1 \alpha_{i,t}$ and ωs have been estimated by regression on the time-series of alphas. Note that the gamma hat is defined in a similar manner.

4.3.5 Model Fitting

4.3.5.1 Fama-MacBeth cross-sectional regression

To assess the models' predictive performance, we run cross-sectional Fama-MacBeth predictive regressions. During the period January 1965 to December 2016, on the last day of each month, we regress the cross-section of future realized returns on estimated returns from the three model predictions,

$$R_{t,t+1}^{i} = \phi_{0,t+1} + \phi_{1,t+1}\hat{R}_{t,t+1}^{i,M} + \varepsilon_{t+1} \quad (4.3.5.1)^{1}$$

where i indicates the company i, ranging from 1 to 3289; t is the month in the time series from 1 to 624; M is the type of model, $\hat{R}_{t,t+1}^{i,M}= [\hat{R}_{t,t+1}^{i,M_1}, \hat{R}_{t,t+1}^{i,M_2}, \hat{R}_{t,t+1}^{i,M_3}]$; $\hat{R}_{t,t+1}^{i,M}$ is the estimated returns from the three types of model; the independent variable is the future realized return on the last day of each month for all companies listed on the stock exchange, beginning with January 1965; ε_{t+1} is the residuals of regression. The t-statistics of significance of the coefficients of $\phi_{0,t+1}$ and $\phi_{1,t+1}$ are robust-tested by Newey et al. (1987) one lag (one month) correction. The adjusted R^2 is kept as an indicator of the quality of regression.

4.3.5.2 Newey-West t-statistics

The standard OLS regression are unbiased but statistical inference can be influenced due to the presence of autocorrelation in the residuals. As a result, we employ the Newey-West covariance correction to eliminate the bias caused by the serial

[1] Jacob Mincer & Victor Zarnowitz (1969): "The evolution of economic forecasts" Mincer and Zarnowitz proposed a relative accuracy analysis method in order to estimate a scientific economic forecast. Based on enough amount of empirical data, they proposed a model of comparison of predictions and the realizations to prove the tests validity.

correlations. To improve the measurement accuracy and to avoid strong seasonality, the lag of Newey-West correction should equal to one (month), corresponding to our returns prediction at monthly frequency.

4.4 Further analysis

The model with highest R-square from the three types of prediction models could be selected and used for further analysis. In an attempt to mitigate the effects of the curse of dimensionality [5] that cause estimation errors, we used a low-dimensional combination of explicative variables when predicting returns. However, the number of possible combinations is extremely high, rendering any computation impossible. Another approach in deciding which variables to select involves using univariate and multivariate analysis and selecting the combination of betas that results in the highest R^2. Furthermore, we check the industry distributions of estimated returns by using the selected best combination to see if the performance could be affected by industry factor.

4.4.1 Univariate analysis

First at all, to test the effect of individual beta on the future realized return, we regress each beta defined above and the future realized return for the stock i as (4.4.1.1). Using the model with the highest R-square from the three types of prediction models, we compute t-statistics (cross-sectional regression and Newey-West) of significance of the coefficients and R^2 respectively:

$$R_{t,t+1} = \alpha_t + \gamma^j \beta_t^j + \varepsilon_t \qquad (4.4.1.1)$$

[5] "The curse of dimensionality," that is problems due to high-dimensional variables when analyzing and organizing data.

in which, the subscript t refers to monthly measure, from 1 to 624 representing January 1965 to December 2016; j is the observation windows, correspondent to 21 days, 63 days, 252 days, 24 months, 36 months, 48 months and 60 months; C is from 1 to 3289 representing the proxy code of each stock; $\alpha_t = (\alpha_t^1, \alpha_t^2, \dots, \alpha_t^C)' \in \mathbb{R}^C$ is the vector of conditional alphas across stocks 1, …, C; $\beta_t = (\beta_t^1, \beta_t^2, \dots, \beta_t^C)' \in \mathbb{R}^{7*C}$ is the corresponding matrix of defined betas; finally, he errors have been collected in the vector ε_t.

4.4.2 Multivariate analysis

To combine multiple variables, we analyze the results from univariate analysis and regress the future realized returns and the betas with similar results from the multivariate analysis. Also, depending on the definition of betas, we classify combinations as short-term factors, medium-term factors, long-term factors, and recorded R^2 to check the quality of regression.

4.4.3 Industry classification

We wish to explore the effect of industry factors on the quality of the models' forecasts. The industry classification is based on the newly available North American Industry Classification System (NAICS). Developed jointly by the U.S., Canada, and Mexico, it reflects the changing business activities in the last decade. Thus, we classify the companies in 10 industries as displayed in Table (4.1).

Table 4.1 Industry classification

SIC code	Industry name	Nb of companies	Average returns
<10	Agriculture, Forestry, and Fishing	18	0.0050
10-14	Mining	462	0.0115
15-17	Construction	317	0.0138
20-39	Manufacture	1898	0.0123
40-49	Transportation, Communication, Electric, Gas, Sanitary	481	0.0107
50-51	Wholesale trade	584	0.0117
52-59	Retail trade	365	0.0123
60-67	Finance, Insurance, Real Estate	660	0.0121
70-89	Service	483	0.0126
90-99	Public administration	12	0.0081
Total		5280	0.0110

Notes: The table reports the original data before filtering stocks by our criteria. According to the North American Industry Classification System (NAICS), stocks listed on the NYSE could be classified in ten parts of industry. "Nb of companies" represents the number of companies listed on the NYSE during January 1960 to December 2016. According to the industry classification, the average returns are calculated by the mean value of stock monthly returns in each industry.

To specify the effect of industry factor on the future return, we use the model with the highest R-square from the multivariate analysis and re-regress the future realized return and combined betas for each industry

$$R_{t,t+1}^{sic} = \alpha_t + \gamma_j \beta_{j,t}^{sic} + \varepsilon_t \qquad (4.4.1.1)$$

where sic represents respectively ten industries. Furthermore, the prediction model could be used to test the quality of regression and to check the influence of industry.

V. Empirical results

Following the above discussions, we first summarize the statistics of varying risk factors (i.e., time-varying betas) using different observation periods, then verify the forward betas one-month ahead of date with our assumptions supposed in section 4.2 to ensure beta in the following month is a function of the realized risk factors with different observation windows. With the verification completed, we predict the forward returns of the following months by using the cross-sectional regression on risk factors with different observation windows and by constructing a prediction model to estimate returns by Models I, II, and III (4.3.2, 4.3.3, 4.3.4). We also use the Fama-MacBeth cross-sectional regression test, the Newey-West test, the univariate analysis test, and the multivariate analysis test along with the study of industry classification effect on forward returns to select the most efficient prediction model.

5.1 Summary Statistics on risk factor

As mentioned in Section 4.1.1, by exploring a variety of window lengths and return horizons, we are able to define seven betas representing the effect of relative risks within different observation periods of stock returns. The result of statistics places the focus on time variations in betas and on gaining additional perspective for risk factors.

Table 5.1 reports summary statistics for betas within different observation windows. The mean values of risk factors are defined as: $\beta^{21d}, \beta^{63d}, \beta^{252d}, \beta^{24m}, \beta^{36m}, \beta^{48m}$, β^{60m} which are generally close to 1 and Table 5.1.1 indicates the potential collinearities among beta measures. Betas with short-term estimation windows, $\beta^{21d}$and β^{63d} have a mean value of 0.924 and 0.936 respectively. The values are closer and less than one, meaning the firms' systematic risks are similar to the short-term market index's risk. Betas with long-term observation windows, like β^{252d}, β^{24m}, β^{36m}, β^{48m}, β^{60m} have a mean value closer than one. These values, which are interpreted to show the average firms' systematic risks, are higher than the market index risk.

The statistics of kurtosis indicates the characteristics of fat tail for the risk factors we defined above. The kurtosis of the risk factors is much higher than that of a normal distribution and tends to decrease as the length of observation window increases. Thus, betas pdfs are leptokurtic with shapes dependent on the time scale, presenting a very slow convergence of the "central limit theorem" towards normal distribution. For the beta with an observation window of 21 days, the kurtosis is at its highest (12.83), presenting short-term risk factors that have more abnormal betas than predicted.

Table 5.1 Statistics for risk factors

	Min	Max	Mean	Median	SD	Skewness	Kurtosis	Nb of Obs	Obs window
β_t^{21d}	-6.13	3.233	0.924	0.852	0.987	-1.523	12.828	804785	21d
β_t^{63d}	-2.34	5.519	0.936	0.873	0.672	0.405	9.808	804785	63d
β_t^{252d}	-3.66	6.125	1.118	1.038	0.984	0.745	9.577	804785	252d
β_t^{24m}	-8.39	7.969	1.121	1.058	0.738	0.717	6.155	804785	24m
β_t^{36m}	-4.28	7.459	1.124	1.067	0.655	0.772	5.416	804785	36m
β_t^{48m}	-4.31	7.523	1.121	1.072	0.615	0.783	5.309	804785	48m
β_t^{60m}	-2.88	4.632	1.119	1.120	0.592	0.771	5.277	804785	60m

Notes: This table reports the statistics of risk factors with different observation windows. The first column represents the betas with different observation windows. At the end of each month, we calculate seven risk factors using daily returns during the past 21 days, 63 days and 252 days, and using monthly returns during the past 24 months, 36 months, 48 month and 60 months, respectively. The interval of risk factors is one month. We analyze the statistics of betas to study the feature of risk factors with different lengths of observation periods. The number of observations are 804785, which is sufficient enough for the sample dataset. The sample period for study is from January 1965 to December 2016.

Table 5.1.1 Correlation among risk factors

	β_t^{21d}	β_t^{63d}	β_t^{252d}	β_t^{24m}	β_t^{36m}	β_t^{48m}	β_t^{60m}
β_t^{21d}	1.000	0.681	0.224	0.270	0.281	0.279	0.275
β_t^{63d}	0.681	1.000	0.367	0.420	0.430	0.427	0.418
β_t^{252d}	0.224	0.367	1.000	0.739	0.644	0.591	0.551
β_t^{24m}	0.270	0.420	0.739	1.000	0.876	0.809	0.760
β_t^{36m}	0.281	0.430	0.644	0.876	1.000	0.926	0.877
β_t^{48m}	0.279	0.427	0.591	0.809	0.926	1.000	0.951
β_t^{60m}	0.275	0.418	0.551	0.760	0.877	0.951	1.000

Notes: This table reports the collinearity of risk factors with different observation windows. The first column represents the betas with different observation windows. At the end of each month, we calculate seven risk factors using daily returns during the past 21 days, 63 days and 252 days, and using monthly returns during the past 24 months, 36 months, 48 month and 60 months, respectively. The interval of risk factors is one month. We analyze the statistics of betas to study the feature of risk factors with different lengths of observation periods. The number of observations are 804785, which is sufficient enough for the sample dataset. The sample period for study is from January 1965 to December 2016.

5.2 Conditional betas prediction

Suppose that all the information could be obtained before the date t, as we assumed above, and that the firms used for analysis have been listed on the NYSE for at least 5 years. In this instance, we would be testing our assumption that the future systematic risk (beta-coefficient) at each month is the function of the risk factors experienced with different observation windows. The statistics of the in-sample forward betas one month ahead of the date t with different observation period of the model are shown in Table 5.2. We construct seven models in which dependent variables (y) being the estimated betas on the following month realized betas based on the past 21 days, 63 days, and one-year daily returns as well as two-, three-, four-, five-year monthly returns.

Table 5.2 shows that the constant factors (i.e., intercepts) are generally between 0.03

and 0.09. The adjusted R-square, which is considered as the indicator of the quality of the fit of the regression, indicates that the risk factors with long-run observation windows, such as $\beta^{252d}, \beta^{24m}$, β^{36m}, β^{48m}, β^{60m} calculated with returns more than one year in duration are more likely to be dependent on the past performance of conditional factor loadings. The adjusted R-square of regression for $\beta^{252d}, \beta^{24m}$, β^{36m}, β^{48m} and β^{60m} are higher than that of others, which are 89.4%, 89.7%, 89.0% and 87.7%, respectively.

Table 5.2 Conditional risk factors with observation windows prediction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α	0.0098*	0.0206*	0.0891	0.0513	0.0593	0.0343	0.0399
	(1.7457)	(1.722)	(1.5425)	(1.6161)	(1.6129)	(1.5303)	(1.5379)
β_t^{21d}	-0.0003	0.1528***	-0.0023	0.0010	0.0007	0.0004	0.0002
	(-0.0959)	(4.6366)	(-1.0279)	(1.0709)	(1.0768)	(1.068)	(1.0531)
β_t^{63d}	0.0006	0.5682***	0.0291*	0.0083	0.0047	0.0034	0.0021
	(0.0697)	(10.5059)	(1.7278)	(0.296)	(1.2249)	(1.2028)	(0.1637)
β_t^{252d}	-0.0005	0.0140	0.8897***	0.0161**	-0.0023	-0.0016	-0.0015
	(0.0141)	(0.0302)	(24.1636)	(1.9681)	(-0.0483)	(-0.0102)	(-1.0778)
β_t^{24m}	-0.0028	0.0049*	0.0115	0.9125***	0.0229**	0.0003	0.0006
	(-0.2162)	(1.7536)	(1.1085)	(24.3859)	(1.919)	(-0.054)	(1.0275)
β_t^{36m}	0.0023	0.0277	-0.0084*	0.0074	0.928***	0.0303*	0.0047
	(0.0755)	(1.1338)	(1.7115)	(1.1012)	(23.8297)	(1.8279)	(1.0033)
β_t^{48m}	0.0051	0.0032	0.0206	-0.0037	0.0060	0.9237***	0.0288*
	(0.0723)	(-0.0265)	(-1.0109)	(-1.0051)	(0.0515)	(22.415)	(1.7405)
β_t^{60m}	-0.0064	0.0093	(-0.0155)	0.0157	-0.0115	0.0141	0.9288***
	(-0.2512)	(1.083)	(-1.0344)	(-1.0471)	(-0.0815)	(1.113)	(29.7322)
Adj. R-square	0.6050	0.6510	0.8467	0.8939	0.8974	0.8897	0.8765

Table 5.2 Conditional risk factors with observation windows prediction

Notes: This table represents results of the risk factors using realized returns in the NYSE during the past observation periods in next month. At the end of each month, we compute the risk factors of $\beta^{21d}, \beta^{63d}, \beta^{252d}, \beta^{24m}, \beta^{36m}, \beta^{48m}, \beta^{60m}$ for each stock listed on the NYSE, which represents the short-run risk factors using daily returns during the past 21 days (21d), 36 days (36d), and 252 days (252d), along with the long-run risk factors using monthly returns during the past two years (24m), three years (36m), four years (48m), and five years (60m) and we recorded seven betas on the last date of each month. Assuming the individual risk factor with different observation period (at time t+1) could be related with both short-run and long-run risk factors at the previous month (at time t), seven tests are applied to verify this assumption. Because there is a time-series correlation between the dependent variable and explicative variables, we use the Autoregressive model (AR (1) model) to do the regression. Column (1) shows the regression result of systematic risk factor calculated by 21-day daily returns in the next month with betas we defined above at the month t. Column (2) shows the regression result of systematic risk factor calculated by 63-day daily returns in next month with betas we defined above at the month t. Column (3) shows the regression result of systematic risk factor calculated by 252-day daily returns in next month with betas we defined above at the month t. Column (4) shows the regression result of systematic risk factor calculated by 2-year monthly returns in the next month with betas we defined above at the month t. Column (5) shows the regression result of systematic risk factor calculated by 3-year monthly returns in the next month with betas we defined above at the month t. Column (6) shows the regression result of systematic risk factor calculated by 4-year monthly returns in the next month with betas we defined above at the month t. Column (7) shows the regression result of systematic risk factor calculated by 5-year monthly returns in the next month with betas we defined above at the month t. Adjusted R-square is considered as the indicator of the quality of regression. The dotted portion indicate the best results of appropriate models with good quality. The sample period for observation is from January 1965 to December 2016. The symbols of *, ** and *** represents significance of the result at the 90%, 95% and 99% levels respectively.

5.3 Return prediction models

In this section, to obtain the evolution of parameters with time based on monthly measurements, we first use the cross-sectional regression, then provide three types of prediction models to further estimate the forward monthly returns for each company and develop the tests to select the optimal results in Section 5.4.

5.3.1 Estimating parameters by cross-sectional regression

In total, we include 624 months in our analysis, from January 1965 to December 2016. The dependent variable of regression is the following month forward realized return. Thus, we obtain coefficients of abnormal returns (α) and systematic risk factors (β) each month by using the cross-sectional regression. Figure 5.1 indicates the evolution of constant parameters and coefficients of seven risk factors with different observation windows. From the results shown, the alphas and coefficients of risk factors have obvious serial correlation. The fluctuating tendency shows that the coefficients in the following month is related to past performance. Thus, the estimated returns should be corrected by Newey-West, due to the serial correlation among the coefficients.

The results shown in Table 5.3 are the statistics of cross-sectional regression with time-varying loadings. The value of coefficients for each risk factors are the mean value of 624 parameter estimates on each month. The parameters we obtained are relatively from Fama -MacBeth regressions with separately individual risk factors. The mean value of abnormal returns (α) is 0.0099, which means that the abnormal returns from January 1965 to December 2016 are around 1%. The signs of coefficients of β^{21d}, β^{252d}, β^{24m}, β^{60m} are negative, which demonstrates that the risk factors of the observation period in the past 21 days, one year, two years and five years are negatively related to the forward returns one-month ahead of the date t. However, the t-values are not very significant but β^{24m} is. The result of standard deviation reports the stability of the constructed variables.

Table 5.3: The statistics of cross-sectional regression

	Coef.	t-value	Std. errors	Min	Max	Nb of Obs
α	0.0099	1.7524	0.0762	-0.1453	0.1504	624
γ_t^{21d}	-0.0022	-1.0947	0.0671	-0.0923	0.1489	624
γ_t^{63d}	0.0043	1.0591	0.0806	-0.2494	0.1090	624
γ_t^{252d}	-0.0045	1.3321	0.0728	-0.2454	0.0715	624
γ_t^{24m}	-0.0028	-2.1950	0.0125	-0.3121	0.3239	624
γ_t^{36m}	0.0025	1.0835	0.1449	-0.3569	0.3284	624
γ_t^{48m}	0.0049	1.1644	0.1746	-0.2827	0.4931	624
γ_t^{60m}	-0.0065	-1.6320	0.1546	-0.2562	0.3179	624

Notes: Table 5.3 summarizes the statistics of the constant factors and coefficients corresponding to the seven risk factors with different observation windows along time-varying, which have been collected from the equation of (4.3.1). The dependent variable is all the realized forward returns one month ahead of the date t. At the end of each month, we compute the risk factors for each stock listed on the NYSE, which represents the short-run risk factors using daily returns during the past 21 days (21d), 36 days (36d), and 252 days (252d), and the long-run risk factors using monthly returns during the past two years (24m), three years (36m), four years (48m), and five years (60m), and we recorded seven betas on the last date of each month. Column "Coef." and "t-value" represent the estimates and t-statistics values of alphas and gammas from the equation of (4.3.1). Column "Std. errors" is the square root of the standard deviation of alphas and gammas for 624 months. Column "Min" and "Max" are the minimum value and maximum value of alphas and gammas for 624 months. The sample period for observation is from January 1965 to December 2016; there are thus 624 months for analyzing.

5.3.2 Constructing models' prediction

To further study the prediction of returns one-month ahead of the month t, we introduce three types of prediction models and check the significance of parameters in this section.

Model I is established by directly using the coefficients at the month t from the cross-sectional regression 4.3.1. We calculate the estimates of forward returns by multiplying the coefficients of parameters with the corresponding risk factors on each month. The result is same in Figure 5.1 and Table 5.3 as in the section above. From the statistics of risk factors, generally, the t-values of coefficients in this model are not very significant.

As discussed in the Section 4.3.3, Model II of prediction adopt the mean value of each company as its parameters. We calculate the average value of coefficients of α and risk factors with different observation periods and record them for each stock.

Model III incorporates the autocorrelation which exists among the coefficients of alphas and gammas for each stock, meaning the coefficients of parameters varying over time show a relation with the past month performance. Thus, we use the AR (1) model (4.3.4.1) to get the parameters ($\widehat{\omega}_0$, $\widehat{\omega}_1$, $\hat{\delta}_0$, $\hat{\delta}_1$) used for estimating the coefficients of risk factors for the next month ($\hat{\alpha}^i_{t+1}$, $\hat{\gamma}^j_{i,t+1}$), where j represents the 21 days, 63 days, 252 days, 24 months, 36 months, 48 months, 60 months, respectively.

Figure 5.1 The evolution of parameters through time

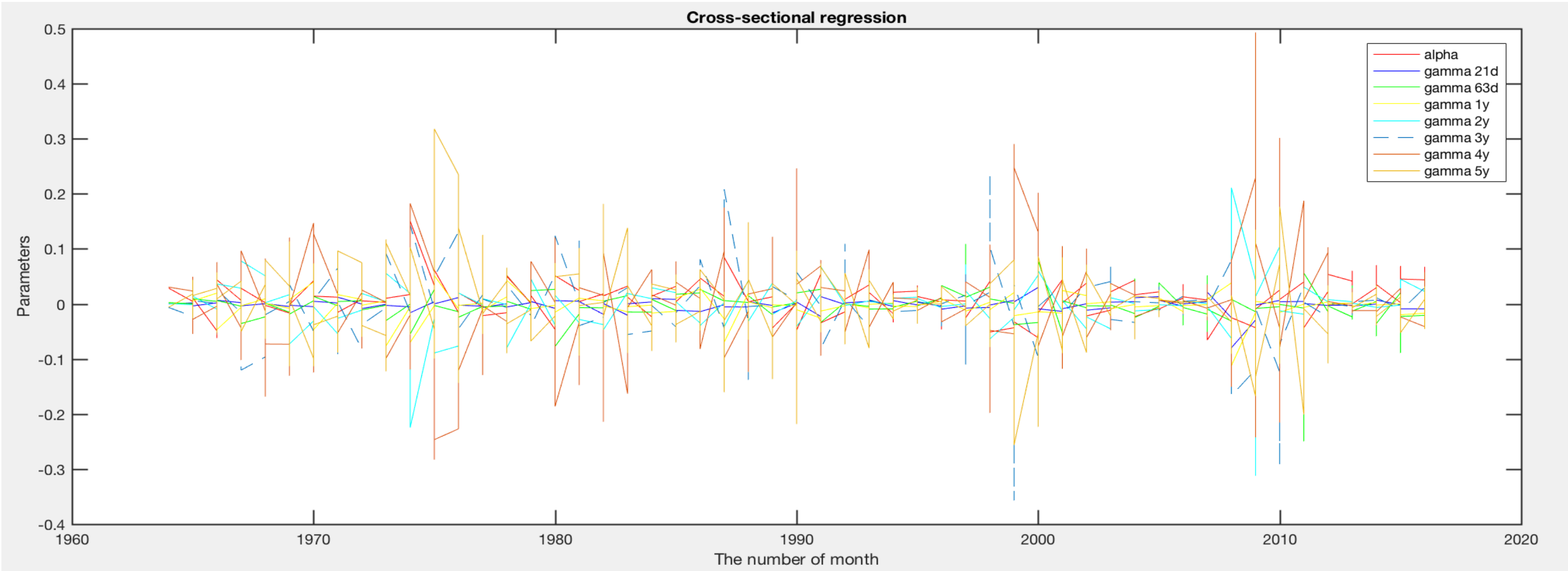


Figure 5.1 indicates the evolution of constant factors and coefficients corresponding to seven risk factors with different observation windows along time-varying, which have been collected from the equation of (4.3.1). The dependent variable is the realized forward returns one month ahead of the date t for common stocks listing on the NYSE. At the end of each month, we compute the risk factors for each stock listed on the NYSE, which represents the short-run risk factors using daily returns during the past 21 days (21d), 36 days (36d) and 252 days (252d), and the long-run risk factors using monthly returns during the past two years (24m), three years (36m), four years (48m), and five years (60m), and we recorded seven betas on the last date of each month. The line is red for the constant factors, blue for the coefficient of risk factors with 21 days, green for that with 63 days, yellow for that with 252 days, cyan with two years, dotted line with three years, magenta with four years, and orange with five years. The sample period for observation is from January 1965 to December 2016 --there are thus 624 months for analyzing.

5.4 Analyzing models' prediction

The Fama-MacBeth cross-sectional regression (4.3.5.1) supports our check for the models' predicative performance by regressing the realized forward returns one-month ahead of date with the returns estimated by the three types of prediction models. The dependent variable is the realized forward returns. Returns estimated from prediction models are regarded as explicative variable. In order to reduce bias caused by autocorrelation and heteroscedasticity, we apply the Newey-West test to correct the statistical values.

$$R_{t,t+1}^{i} = \phi_{0,t+1} + \phi_{1,t+1}\hat{R}_{t,t+1}^{i,M} + \varepsilon_{t+1} \qquad (4.3.5.1)^{1}$$

The time variance in Figure 5.2 demonstrates the evolutions of the constant factor ($\phi_{0,t+1}$) and the coefficient of forward returns estimated by prediction models ($\phi_{1,t+1}$). From the three panels of Figure 5.2, we can see the scale of y-axis [-15, +10] for Model II using the average parameters for each stock. The amplitude of variation is almost 10 times bigger than that of the Model I, directly using parameters at month t [-0.8, +1]. Model III used parameter estimators 5 times smaller than Model II at [-2, +1.5.] The movement patterns for Model II help to capture the time-varying market variation. Table 5.4 further explains the statistics of the three models. The intercepts of regressions show significance for Model I and Model II, but the coefficients of forward returns estimated by these two prediction models are not significant enough to reject the null hypothesis. Robustness analysis demonstrate a t-statistic value of -0.6139 for the

[1] Jacob Mincer & Victor Zarnowitz (1969): "The evolution of economic forecasts" Mincer and Zarnowitz proposed a relative accuracy analysis method in order to estimate a scientific economic forecast. Based on enough amount of empirical data, they proposed a model of comparison of predictions and the realizations to prove the tests validity.

coefficient of predicted returns in Model II, which is enough to accept the hypothesis null, that the coefficient of predicted return equal to one. The analysis concludes that the returns estimated by Model II have no relation with the realized forward returns. The adjusted R-square is considered an indicator of quality for the different models. Model II's adjusted R-square value of 31.3% is much higher than that of other prediction models. Thus, we can conclude that Model II is the optimum model and we should select the second model as the prediction model for our further study, which uses the average parameters from the cross-sectional as the coefficients for each stock.

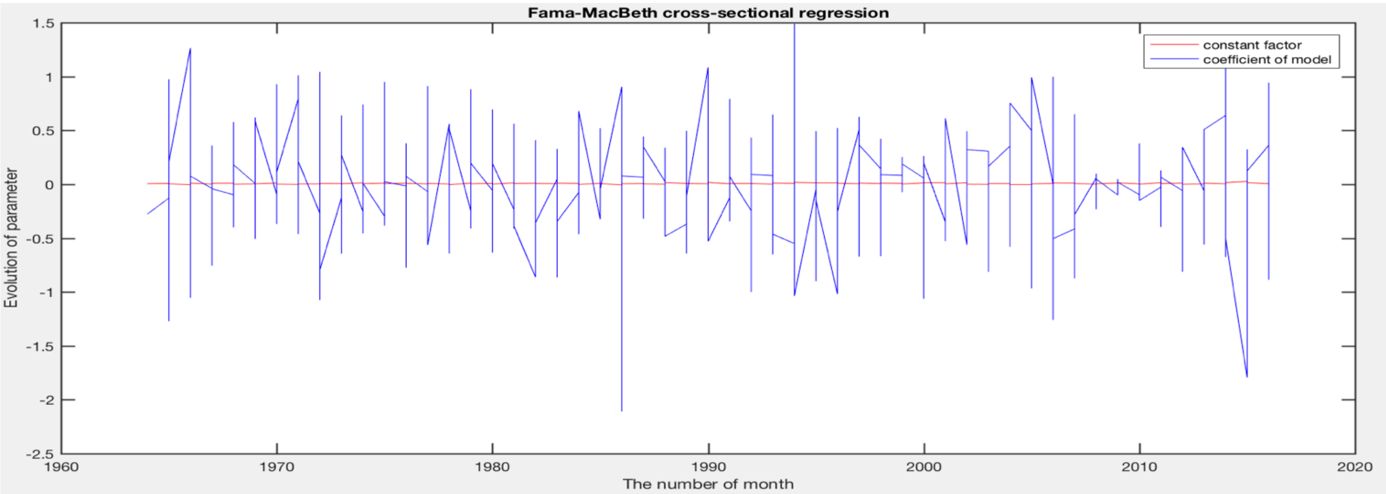
Table 5.4: Fama-MacBeth cross-sectional regression of prediction models

	Model I	**Model II**	**Model III**
$\emptyset_0$	0.0076***	0.0036*	0.0076***
t-stat_OLS	(25.8392)	(1.6455)	(38.329)
t-stat_NW	(25.7111)	(1.5698)	(36.549)
$\emptyset_1$	-0.0045*	0.5433	-0.0045***
t-stat_OLS	(1.8714)	(-0.6238)	(3.3351)
t-stat_NW	(1.8689)	(-0.6139)	(3.3290)
Adj. R-square	0.089	0.313	0.187

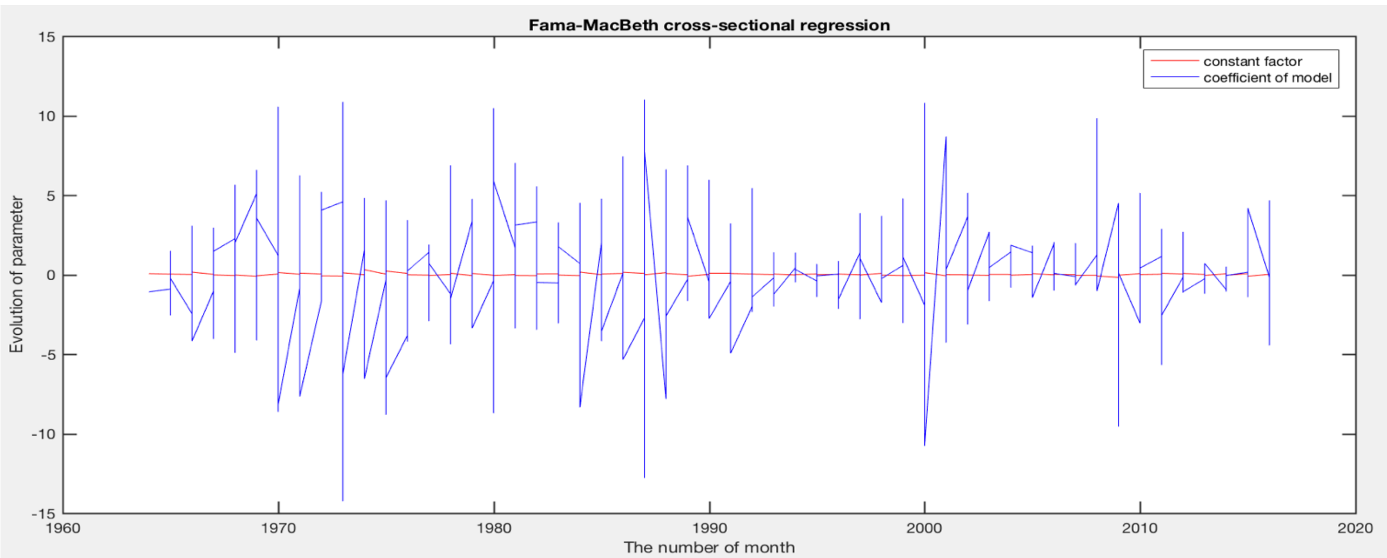
Notes: Table 5.4 reports the results of the Ordinary least squares regression and the Fama-MacBeth cross-sectional regression with the realized forward returns one month ahead of the date t for common stock listed on the NYSE and predicted returns estimated by three types of prediction models from the equation of (4.3.5.1). Model I is the prediction model we use directly for the coefficients from cross-sectional regression of realized forward returns and the seven risk factors in order to estimate the predicted returns in next month (at t+1). Model II is the prediction model we use to collect the time-varying alphas (α) and gammas (γ) from (4.3.1) and employ the average value of alphas and seven coefficients of risk factors to estimate the predicted returns in next month (at t+1). Model II is the prediction model we use as the estimator of alphas and seven coefficients of risk factors for each company to predict the returns in the next month (at t+1). Column "$\emptyset_0$"is intercept of the regressions and Column "$\emptyset_1$"is the coefficients of predicted returns from models. Column "t-stat_OLS" is the t-statistics value of parameters from OLS regression and Column "t-stat_NW" is the t-statistics value of parameters after robustness test which correct the heteroscedasticity and autocorrelation. The t-stat for "$\emptyset_1$" is used for testing whether average $\emptyset_1$ is different from 1. Adj. R-square is considered as the indicator of the quality of each regression. The symbols of *, ** and *** represent significance of the result at the 90%, 95% and 99% levels respectively.

Figure 5.2 Fama-MacBeth cross-sectional regression

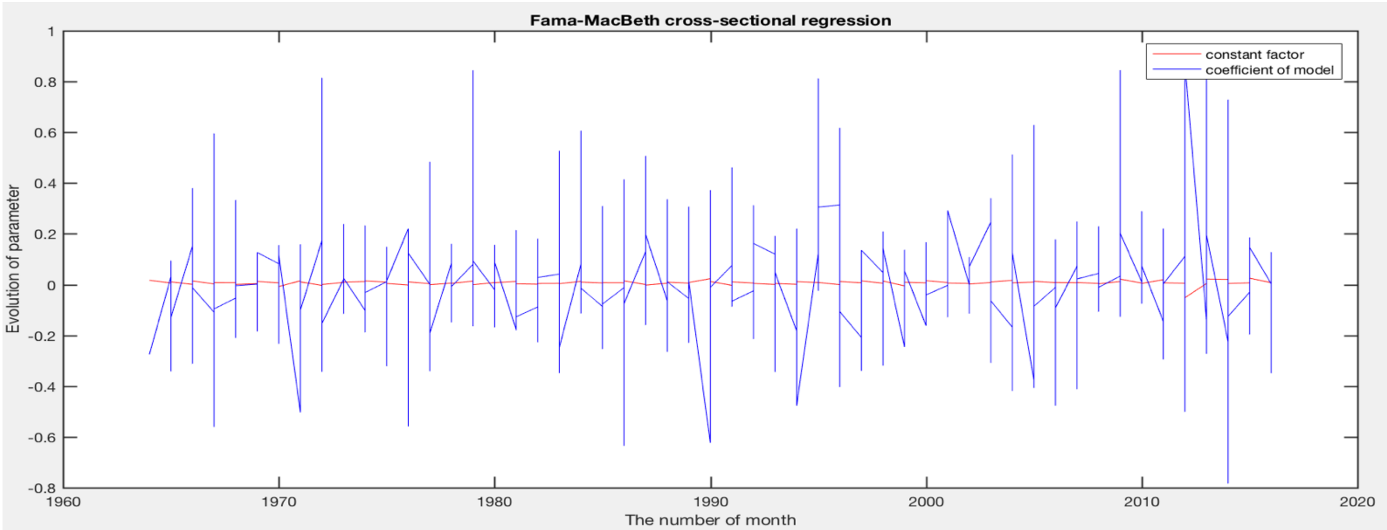
Panel A: Fama-MacBeth cross-sectional regression for Model I



Panel B: Fama-MacBeth cross-sectional regression for Model II



Panel C: Fama-MacBeth cross-sectional regression for Model III



Notes: Figure 5.2 shows the indicates, the evolution of constant factors and coefficients corresponding to predicted returns estimated by three types of prediction model, which have been collected from the equation of (4.3.5.1). The dependent variable is the realized forward returns one month ahead of the date t for common stocks listed on the NYSE. Panel A shows the evolution of constant factors and coefficients of estimated returns corresponding to Model I, in which we use directly the coefficients from cross-sectional regression of realized forward returns and seven risk factors to estimate the predict returns in the next month (at t+1). Panel B shows the evolution of constant factors and coefficients of estimated returns corresponding to Model II, in which we collect the time-varying alphas (α) and gammas (γ) from (4.3.1) and use the average value of alphas and seven coefficients of risk factors to estimate the predict returns in next month (at t+1). Panel C shows the evolution of constant factor and coefficient of estimated returns corresponding to Model II, in which we collect the time-varying alphas (α) and gammas (γ) from (4.3.1) and use the estimator of alphas and seven coefficients of risk factors for each company to predict the predict returns in next month (at t+1). The sample period for observation is from January 1965 to December 2016; there are thus 624 months for analyzing.

5.5 Further analysis

To study further the relationship between forward return and risk factors with different observation periods, we test the efficiency of each risk factor in order to choose the best combination of variables by conducting univariate and multivariate analysis. As the results show in Section 5.4, using the mean value of the chosen parameter to estimate forward return gives the best quality of regression. Thus, based on Model II's average parameter, we further perform three types of analysis to select the model by which the predicted estimates could be pertinent to use in the market.

Firstly, we conduct OLS regression to reveal the pure effect of individual risk factor on forward realized returns. Secondly, we find the model closest to the tendency of the real market by seeking the best combination of risk factors. Based on the results from the previous section, we combine the risk factors with similar characteristics such as the length of observation windows and significant statistical results. Following the combination, we regress forward realized returns and potential risk factors combinations to select the optimal result with the highest R-square value.

Finally, we verify the effect of the allocation of industries by classifying firms into 10 industries and using the prediction model to obtain checks for each industry. These checks are of paramount importance: industry factor is a significant component in the US market as we can often find that performance of one industry overrides others.

5.5.1 Results from univariate analysis

After applying the cross regression (4.4.1.1) in Panel A of Table 5.5, we gather the results of individual risk factors within different observation-periods in correspondent to one month, three-month, one-year, two-year, three-year, four-year, and five-year marks and compared the differences among them. The dependent variable is realized return of one month following the date t. Our cross-sectional regression analysis includes 624 months from January 1965 to December 2016. As a result, we obtain the coefficients of alpha and gamma on each month.

Cross-sectional regression illustrates the tendency of return's evolution through time. We use Models (1), (2), (3), (4), (5), (6), (7), in corresponding with regressions of individual risk factors with betas $\beta^{21d}, \beta^{63d}, \beta^{252d}, \beta^{24m}, \beta^{36m}, \beta^{48m}, \beta^{60m}$. Figure 5.3 shows the tendency of abnormal returns (α) and coefficients of individual risk factors over time. The y-axis of Models (1), (2), and (3) ranges from [-0.3, +0.3], higher than that of Model (4) and Model (5), which ranges from [-0.2, +0.25]. Model (5) and (6) have stable patterns of movement, of which their y-axis is the smallest ranging from [-0.15, +0.2], due to the long-run observation windows. Models (1), (2), and (3)'s regression of forward realized returns against risk factors within the observation period less than one year are more likely to fluctuate owing to short-term variants.

The results shown in Panel A are the average value of 624 parameter estimates we obtain by regression with each risk factor. Through the short-term observation window, the abnormal return, which is noted as α in Panel A, is kept stable mostly for β^{21d} and up to β^{252d}. When the length of observation window is longer, the value of α increases proportionally. Viewing the dataset as a whole, the coefficients of risk factors (γ) with the short-term observation window are positive. Our interpretation of this phenomenon is that within the one-year observation period, riskier firms are predicted to garner higher returns. On the contrary, within the long-term observation windows from β^{24m} to β^{60m}, corresponding to monthly returns from two to five years, risk factors (β) are negatively correlated to the forward stock return, indicating less riskier firms are predicted to get higher returns within the following month. With regard to the quality of regression, R-square are higher with long-term observation windows in $\beta^{252d}, \beta^{24m}, \beta^{36m}, \beta^{48m}, \beta^{60m}$, proving that the risk factors calculated with long-term monthly returns are more statistically significant.

Based on the second prediction model, we calculate the mean value of the parameters (α and γ) for each firm. The dependent variable in Panel B of Table 5.5 is the one-month forward realized return which is used to regress with the return estimated from prediction model (5.5.1.3)

$$\hat{R}_{t,t+1}^{i,M_2} = \overline{\alpha}_i + \overline{\gamma}_i^j \beta_{i,t}^j$$

where i represents company i, ranging from 1 to 3289; t is the month in time series from 1 to 624; M_2 is the second prediction model for which we calculate the mean value of parameters of each company; j is the observation windows, correspondent to 21 days, 63 days, 252 days, 24 months, 36 months, 48 months and 60 months.

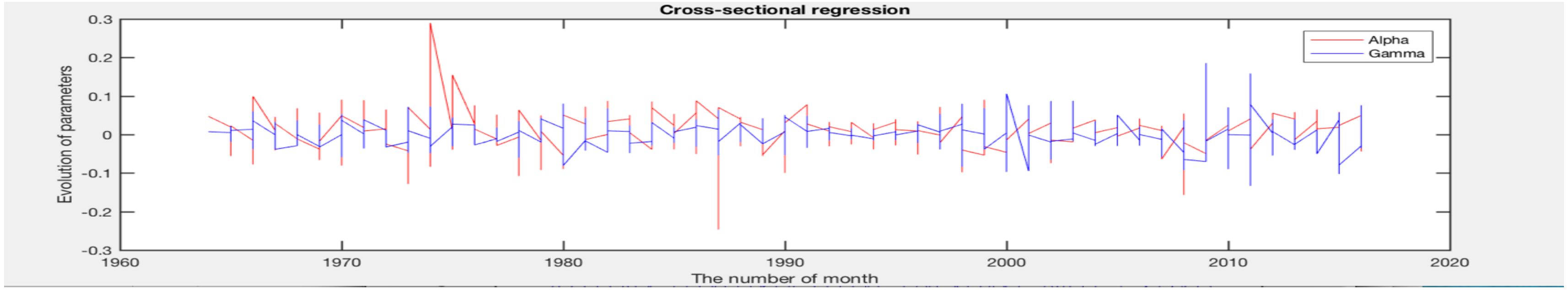
Panel B presents the result of the Fama-MacBeth cross-sectional regression (4.3.5.1) of the estimated returns from the prediction models against individual risk factors. From this result, we can see that the intercepts of the regressions, the constant difference between realized forward return and the forward return estimated by the prediction model, are kept stable for individual risk factors with different observation windows. The coefficients of risk factors (ϕ_1) show the multipliers of risk factors with short-term observation windows are higher than that with long-term windows. Especially for β^{21d} to β^{252d}, the coefficients of risk factors are between 0.2 and 0.3, which have t-statistic values higher than other betas (from β^{24m} to β^{60m}). R-square values from these regressions indicate that the quality of regression increase as the length of observation windows of each individual risk factors increases. The prediction model show that individual betas calculated using five-years monthly returns are more pertinent with the realized forward return.

Figure 5.3: Cross-sectional regression for univariate analysis

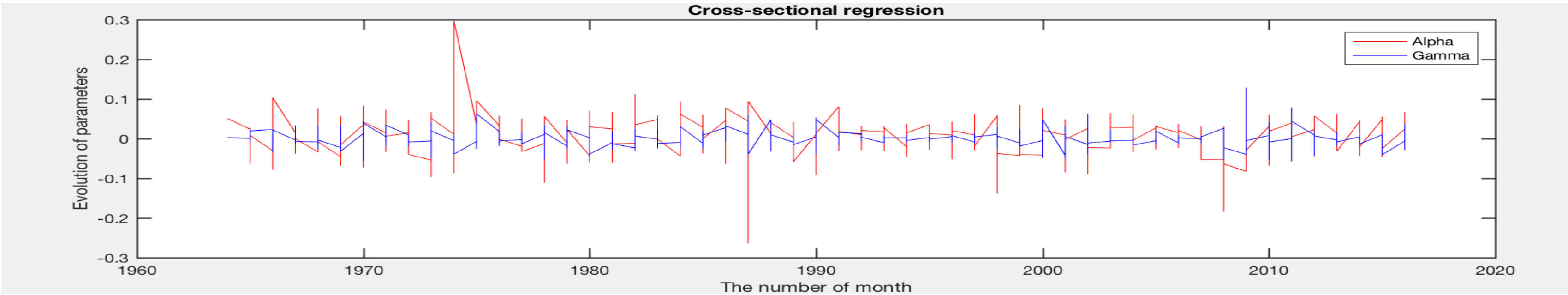
Figure 5.3 shows the evolution of constant factor and individual coefficient corresponding to seven risk factors with different observation windows along time-varying, which is collected from the cross-sectional regression of ((4.4.1.1)). The dependent variable is the realized forward returns one month ahead of the date t for common stock listed on the NYSE. At the end of each month, we compute the risk factors for each stock listed on the NYSE, which represents the short-run risk factors using daily returns during the past 21 days (21d), 36 days (36d) and 252 days (252d), and the long-run risk factors using monthly returns during the past two years (24m), three years (36m), four years (48m) and five years (60m). Panel A reports the alpha and gamma from the regression with the individual risk factor using the past 21 days of daily returns. Panel B reports the alpha and gamma from the regression with the individual risk factor using the past 63 days of daily returns. Panel C reports the alpha and gamma from the regression with the individual risk factor using the past one-year daily returns. Panel D reports the alpha and gamma from the regression with the individual risk factor which using the past two-year monthly returns. Panel E reports the alpha and gamma from the regression with the individual risk factor using the past three-year monthly returns. Panel F reports the alpha and gamma from the regression with the individual risk factor using the past four-year monthly returns. Panel G reports the alpha and gamma from the regression with the individual risk factor using the past five-year monthly returns. The sample period for observation is from January 1965 to December 2016; there are thus 624 months for analyzing.) We also record seven betas on the last date of each month.

Panel A: Cross-sectional regression with β^{21d}

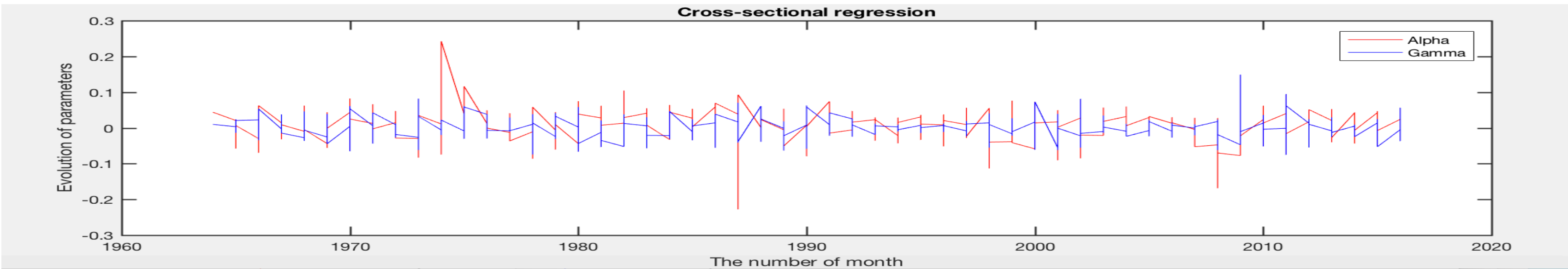
Panel B: Cross-sectional regression with β^{63d}



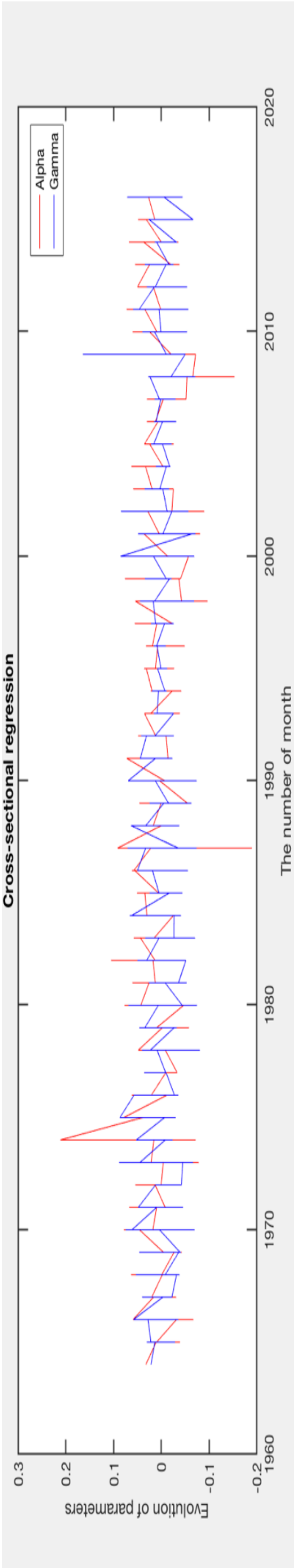
Panel C: Cross-sectional regression with β^{252d}



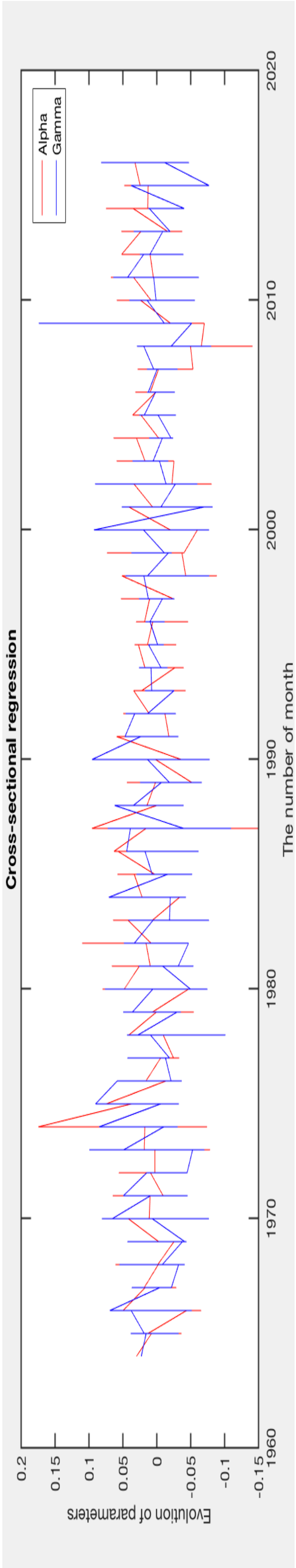
Panel D: Cross-sectional regression with β^{24m}



Panel E: Cross-sectional regression with β^{36m}



Panel F: Cross-sectional regression with β^{48m}



Panel G: Cross-sectional regression with β^{60m}

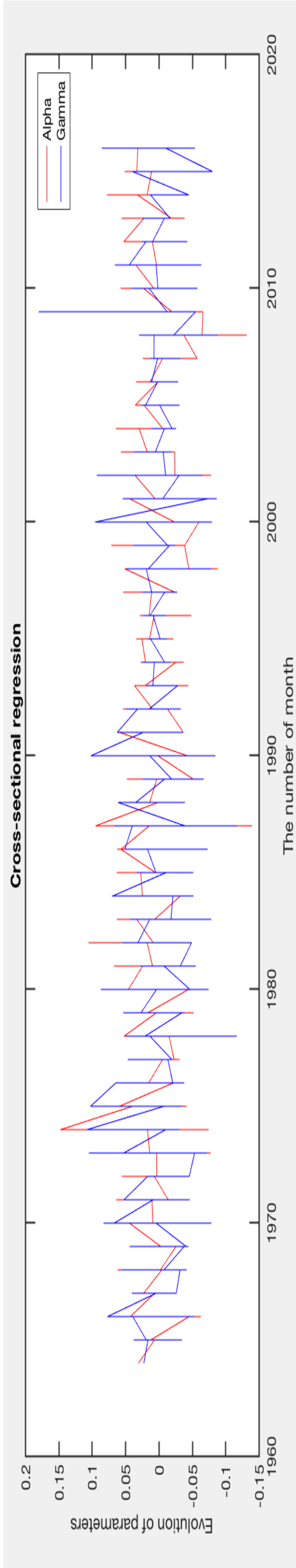


Table 5.5: Univariate analysis

Panel A: Cross-sectional regression with individual risk factor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α	0.0077**	0.0078*	0.0078**	0.0091**	0.0096**	0.0098**	0.0101**
t-stat_OLS	(2.118)	(1.7869)	(2.0116)	(2.0922)	(2.0479)	(2.0396)	(2.0570)
γ	0.000426	0.000497	0.000584	-0.0018	-0.0021	-0.0022	-0.0025
t-stat_OLS	(-0.2179)	(-0.3121)	(-0.4377)	(-0.7261)	(-0.7542)	(-0.7722)	(-0.8298)
Adj. R-square	0.1290	0.0923	0.1340	0.1450	0.1427	0.1435	0.1437

Panel B: Fama-MacBeth cross-sectional regression with prediction model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\emptyset_0$	0.0063**	0.0056**	0.006**	0.0063**	0.0063**	0.0062**	0.0061**
t-stat_OLS	(2.4657)	(2.2348)	(2.3548)	(2.1328)	(2.1259)	(2.0994)	(2.0661)
t-stat_NW	(2.3211)	(2.1105)	(2.2105)	(2.0237)	(2.0205)	(1.9946)	(1.9655)
$\emptyset_1$	0.2077	0.2918	0.2325	0.1669	0.4792	0.1756*	0.1627*
t-stat_OLS	(-0.4074)	(-0.7345)	(0.7198)	(0.8125)	(1.3748)	(1.7945)	(1.8039)
t-stat_NW	(-0.3991)	(-0.6379)	(0.7014)	(0.8091)	(1.3720)	(1.6868)	(1.6757)
Adj. R-square	0.1270	0.1370	0.1220	0.1810	0.1920	0.1960	0.2090

Table 5.5: Univariate analysis

Notes: Table 5.5 shows the result of univariate analysis. Panel A reports the results of cross-sectional OLS regression with the individual risk factor within different observation-periods, which correspond to one month, three months, one year, two years, three years, four years and five years. The dependent variable is the realized forward returns one month ahead of the date t for common stock listed on the NYSE. Column (1) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 21-day daily returns. Column (2) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 63-day daily returns. Column (3) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 1-year daily returns. Column (4) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 2-year monthly returns. Column (5) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 3-year monthly returns. Column (6) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 4-year monthly returns. Column (7) represents the regression of realized forward returns on the next month with the individual systematic risk factor calculated on the past 5-year monthly returns. Panel B reports the results of Ordinary least squares regression and Fama-MacBeth cross-sectional regression with the realized forward returns one month ahead of the date t for common stock listed on the NYSE and predicted returns estimated by the prediction of the Model II from which we collect the time-varying alphas (α) and gammas (γ) of (4.3.1) and use the average value of alphas and the seven coefficients of risk factors to estimate the predicted returns in the next month (at t+1). Row "$\emptyset_0$"is the intercept of the regressions and Row "$\emptyset_1$"is the coefficient of the predicted returns from models. The t-stat for "$\emptyset_1$" is used for testing whether average $\emptyset_1$ is different from 1. Column (1) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 21-day daily returns. Column (2) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 63-day daily returns. Column (3) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 1-year daily returns. Column (4) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 2-year monthly returns. Column (5) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 3-year monthly returns. Column (6) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 4-year monthly returns. Column (7) represents the regression of realized forward returns on the next month with estimated returns predicted by the prediction model with the individual systematic risk factor calculated on the past 5-year monthly returns. Adj. R-square is considered as the indicator

5.5.2 Results from multivariate analysis

According to our analysis using the univariate test in Table 5.5.1.1, we conclude that the length of observation period of risk factors has a crucial effect on the univariate analysis result. By better fitting the prediction model with the tendency of market realized return, the multivariate analysis can help combine constructed variables with similar characteristics and test the significance from a statistical perspective and from the perspective of regression quality. In this section, six models with different combinations of constructed risk factors are created for analytical purposes. Table 5.6 shows the definition of the combination models.

Table 5.6: the combination models for multivariate analysis

	Model A	Model B	Model C	Model D	Model E	Model F
α	√	√	√	√	√	
β^{21d}	√				√	√
β^{63d}	√				√	√
β^{252d}	√				√	√
β^{24m}		√	√			√
β^{36m}		√	√	√		√
β^{48m}		√	√	√	√	√
β^{60m}		√		√		√

Notes: Table 5.6 reports the definition of six combination models for multivariate analysis and indicates the risk factors used to be analyzed in the models.

Panel A of Table 5.7 shows the statistics of six models noted as Model A, B, C, D, E, F, by applying the cross regression (5.5.2.2). The regression is performed based on the dependent variable which is the realized return one month following the date t against the combination of constructed risk factors with different observation-periods.

$$R_{t,t+1} = \alpha_t + \gamma^M \beta_t^M + \varepsilon_t \tag{5.5.2.2}$$

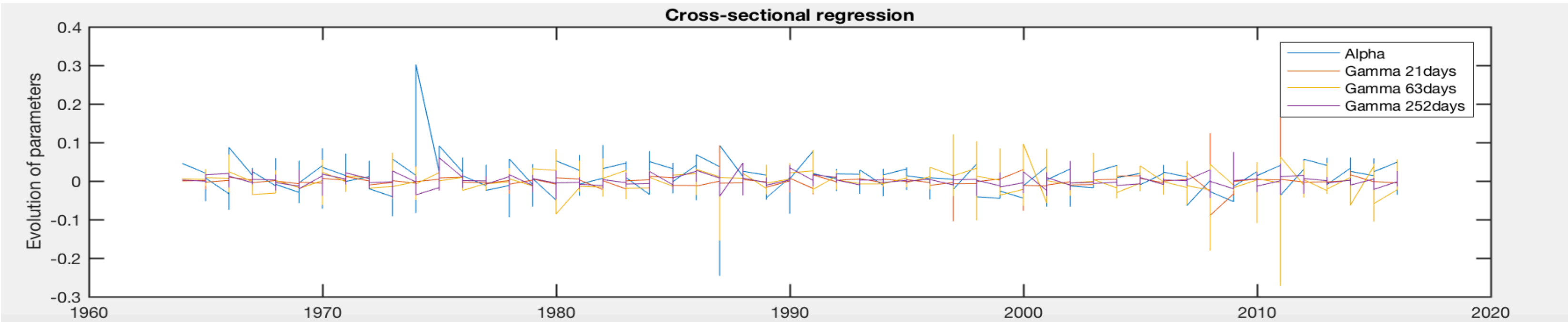
In which M representing six models of combination, in correspondence to Model A, B, C, D, E, F; β_t^M, is the matrix of explicative variables, representing the combination of risk factors with different observation windows in each model. Our analysis includes 624 months from January 1965 to December 2016. Further, by using cross-sectional regression, we get the coefficients of alpha and gamma on each month.

Figure 5.4 illustrates the tendency of abnormal returns (α) and the change of coefficients of combined risk factors with different observation periods from regression 5.5.2.2 for each model over time. The combinations for long-term observation period betas, like Model B and Model C with the y-axis ranging from [-0.5, +0.5], are the most fluctuant. The figure shows the existence of serial correlations among the coefficients, which are also proven by the analysis of robustness tests. Using the second prediction model, we obtain the average value of parameters for each company. Panel B of Table 5.5.2.1 shows result of the Fama-MacBeth cross-sectional regression (4.3.5.1) of the estimated returns from the prediction models against each combination of risk factors.

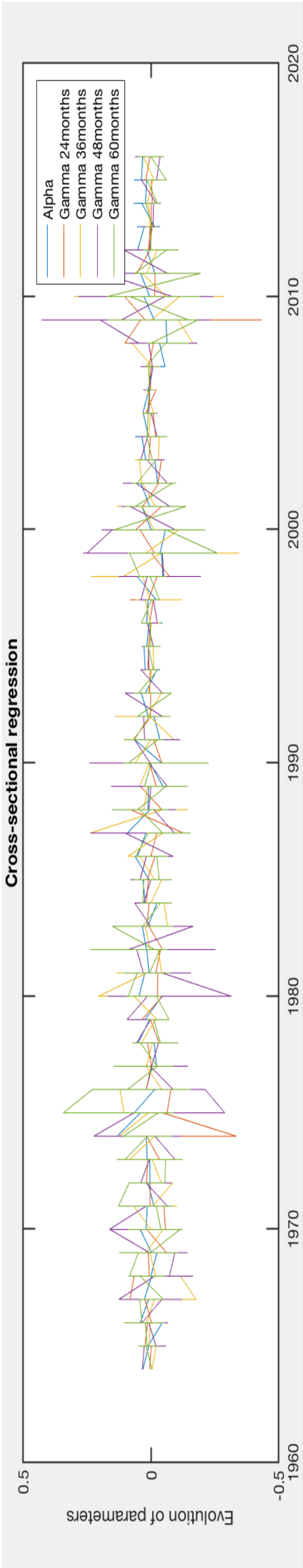
Figure 5.4: Cross-sectional regression for multivariate analysis

Figure 5.4 indicates the evolution of constant factor and coefficient corresponding to risk factors along time-varying, which are collected from the cross-sectional regression of ((4.4.1.1)) for each model. The dependent variable is the realized forward returns one month ahead of the date t for common stock listed on the NYSE. Panel A reports the alpha and gamma from the regression of Model A with risk factors using the past 21-day returns, 63-day returns and one-year returns. Panel B reports the alpha and gamma from the regression of Model B with risk factors using the past two-year returns, three-year returns, four-year returns and five-year returns. Panel C reports the alpha and gamma from the regression of Model C with risk factors using the past two-year returns, three-year returns and four-year returns. Panel D reports the alpha and gamma from the regression of Model D with risk factors using the past three-year returns, four-year returns and five-year returns. Panel E reports the alpha and gamma from the regression of Model E with risk factors using the past 21-day returns, 63-day returns, one-year returns and four-year returns. Panel F reports the alpha and gamma from the regression of Model F with risk factors using the past 21-day returns, 63-day returns one-year returns, two-year returns, three-year returns, four-year returns and five-year returns, but without considering the constant factor. The sample period for observation is from January 1965 to December 2016, there are thus 624 months for analyzing.

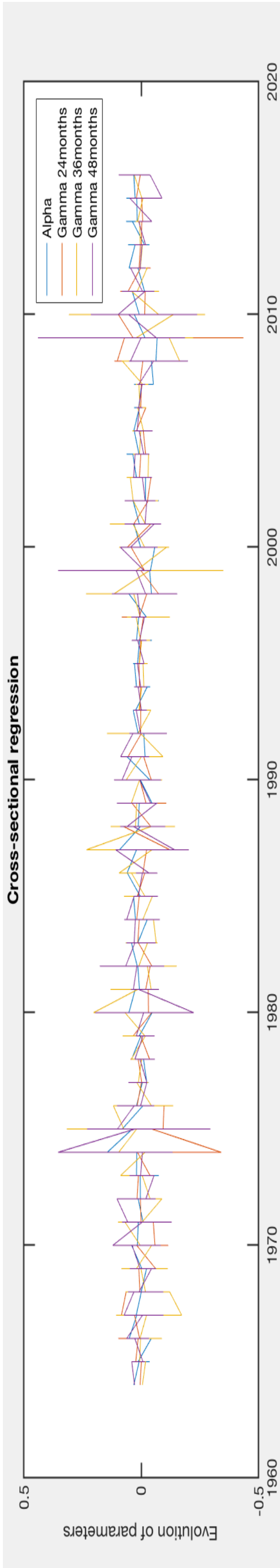
Panel A: Cross-sectional regression for model A



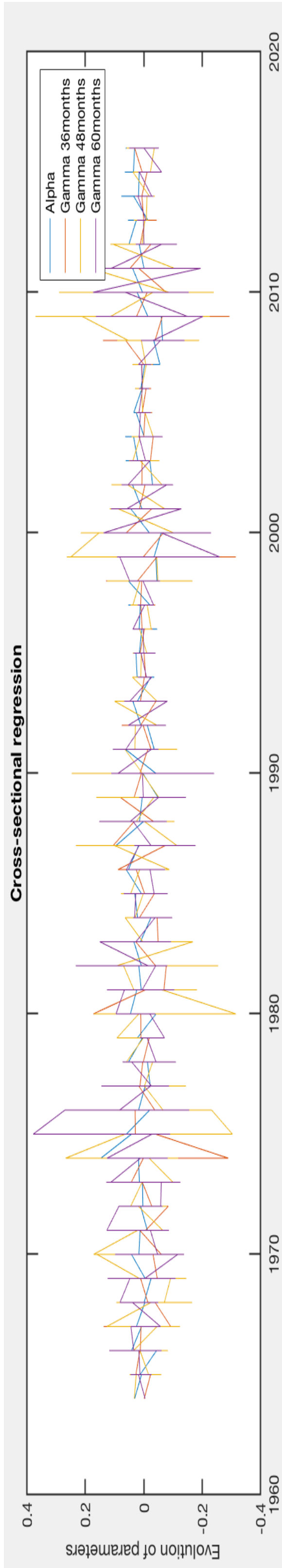
Panel B: Cross-sectional regression for model B



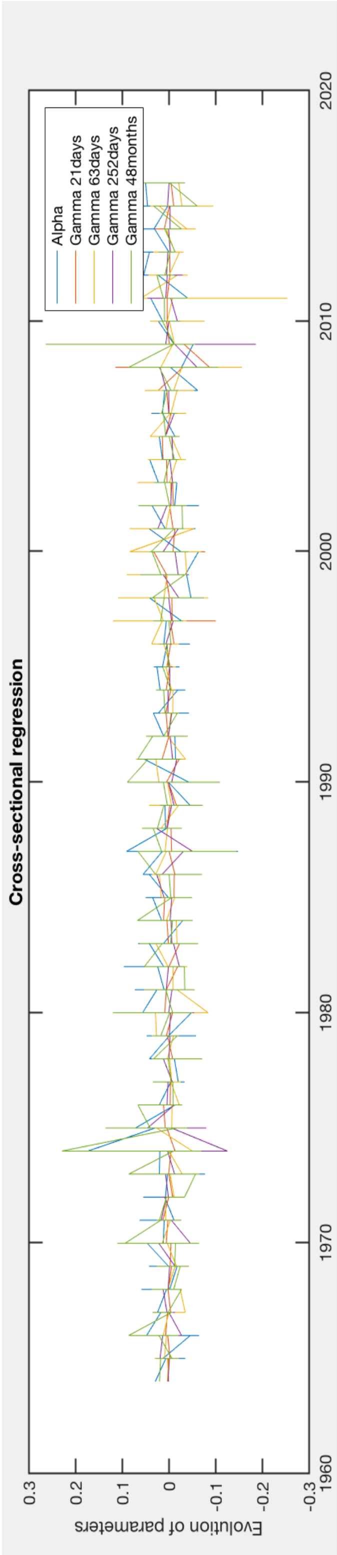
Panel C: Cross-sectional regression for model C



Panel D: Cross-sectional regression for model D



Panel E: Cross-sectional regression for model E



Panel F: Cross-sectional regression for model F

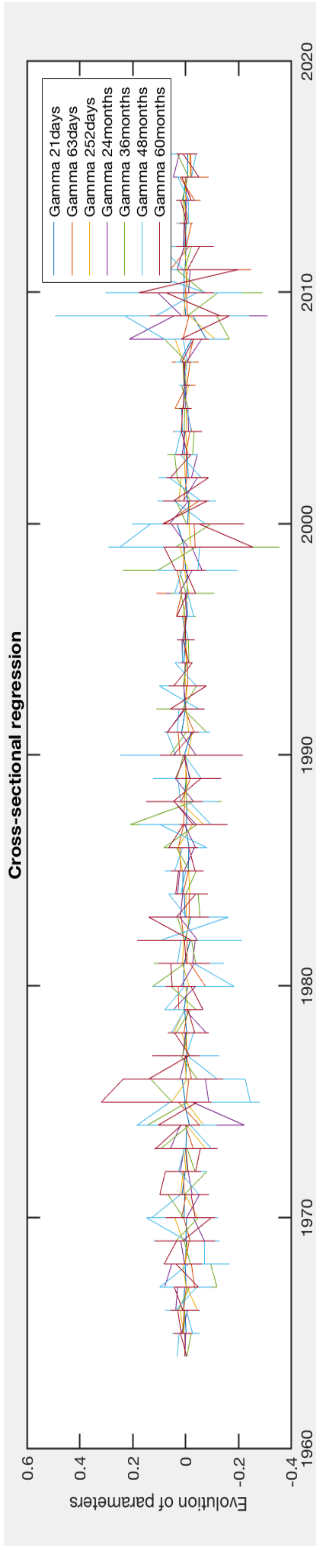


Table 5.7: Multivariate analysis

Panel A: Cross-sectional regression with individual risk factor

	(A)	(B)	(C)	(D)	(E)	(F)
α	0.0083*	0.0099*	0.0097**	0.0099*	0.0096*	
	(1.763)	(1.933)	(1.963)	(1.951)	(1.802)	
γ^{21d}	-0.001618				-0.000343	-0.0023*
	(-0.095)				(-0.088)	(1.672)
γ^{63d}	-0.00024				0.00695*	0.0045
	(-0.178)				(0.169)	(0.871)
γ^{252d}	-0.00079				-0.00946	-0.00454*
	(-1.384)				(-1.177)	(-1.791)
γ^{24m}		-0.0035	-0.0034			-0.0038
		(-1.241)	(-1.239)			(-1.372)
γ^{36m}		0.0022	0.0022	-0.0019		0.0026
		(0.613)	(0.797)	(-0.699)		(-1.105)
γ^{48m}		0.0052	-0.001*	0.0057	-0.0013	-0.001*
		(0.088)	(-1.774)	(0.096)	(-0.489)	(-1.991)
γ^{60m}		-0.0063		-0.0062		-0.0065
		(-0.259)		(-0.263)		(-0.288)
Adj. R-square	0.156	0.167	0.152	0.1505	0.155	0.173

Panel B: Fama-MacBeth cross-sectional regression with prediction model

	(A)	(B)	(C)	(D)	(E)	(F)
$\emptyset_0$	0.005*	0.003	0.004	0.005	0.006***	0.009***
t-stat_OLS	(1.842)	(1.168)	(1.296)	(1.550)	(2.849)	(5.173)
t-stat_NW	(1.747)	(1.101)	(1.236)	(1.467)	(2.674)	(4.862)
$\emptyset_1$	0.174	0.553	0.479	0.401	0.167	0.497
t-stat_OLS	(0.182)	(-1.059)	(-0.065)	(0.743)	(0.089)	(-0.861)
t-stat_NW	(0.106)	(-0.913)	(-0.008)	(0.689)	(0.073)	(-0.793)
Adj. R-square	0.165	0.158	0.219	0.215	0.215	0.316

Notes: Table 5.7 shows the result of multivariate analysis. Panel A reports the results of cross-sectional OLS regression with some of risk factors with different observation-periods, which correspond to one month, three months, one year, two years, three years, four years and five years. The dependent variable is the realized forward returns one month ahead of the date t for common stock listed on the NYSE. Column (A) represents the regression of realized forward returns on the next month with risk factors using the past 21-day, 63-day and 252-day daily returns for Model A, Column (B) represents the regression of realized forward returns on the next month with risk factors using the past the past two-year, three-year, four-year and five-year monthly returns for Model B. Column (C) represents the regression of realized forward returns on the next month with risk factors using the past the past two-year, three-year, four-year monthly returns for Model C. Column (D) represents the regression of realized forward returns on the next month with risk factors using the past three-year, four-year and five-year monthly returns for Model D. Column (E) represents the regression of realized forward returns on the next month with risk factors using the past 21-day, 63-day and 252-day daily returns and four-year monthly returns for Model E. Column (F) represents the regression of realized forward returns on the next month with risk factors using the past 21-day, 63-day, one-year, two-year, three-year, four-year and five-year returns for Model F. Panel B reports the results of Ordinary least squares regression and Fama-MacBeth cross-sectional regression with the realized forward returns one month ahead of the date t for common stock listed on the NYSE and predicted returns estimated by the prediction of Model II from which we collect the time-varying alphas (α) and gammas (γ) of (4.3.1) and use the average value of alphas and coefficients of risk factors of each combination model to estimate the predict returns in next month (at t+1). Row "$\emptyset_0$"is intercept of the regression and the Row "$\emptyset_1$"is the coefficient of predicted returns from models. Row "t-stat_OLS" is the t-statistics value of parameters from OLS regression and Row "t-stat_NW" is the t-statistics value of parameters after robustness test which corrects the heteroscedasticity and autocorrelation. The t-stat for "$\emptyset_1$" is used for testing whether average $\emptyset_1$ is different from 1. Adj. R-square is considered as the indicator of the quality of each regression. The symbols of *, ** and *** represent significance of the result at the 90%, 95% and 99% levels respectively, based on the t-statistics value from Newey-West test.

5.5.2.1 Model for short-run betas

We are interested in the effect of short-term observation-periods on the one-month ahead returns due to the nature of risk factors. The constructed risk factors with observation periods less than one year (from β^{21d} to β^{252d}) are considered as independent variables in Model A. The result of Model A shown in Panel A indicates that all the risk factors with daily returns less than one year are negatively correlated to the one-month forward realized return. This indicates that in the short-term period, riskier firms compared to the market index are expected to get higher returns in the coming month. The result from Panel B of Table 5.7 shows results of the robustness test. Results are corrected by Newey-West to delete the autocorrelation and heteroscedasticity; furthermore, they indicate that the coefficient of the prediction model from Model A is statistically significant. We conclude that predicted returns estimated by Model A are significantly related to the realized return one month ahead of t.

5.5.2.2 Model for long-run betas

Model B is established to analyze the effect of the risk factors with long-term observation periods on the forward returns. Thus, we selected betas calculated with monthly returns over one year, β^{24m}, β^{36m}, β^{48m}, β^{60m}., to regress against the forward realized monthly returns. In the univariate analysis, the signs risk factor coefficients within the long-term observation period is negative. However, in the multivariate analysis in Panel A, the result of Model B indicates risk factors using three-year and 4-year monthly returns that have positive correlations with the β^{36m}, β^{48m}. This is interpreted to be due to obvious collinearity between the risk factors with long-term observation periods. The absolute value of β^{24m} and β^{60m} are higher than the value in the univariate analysis, and the R-square is also higher, indicating the quality of regression has improved. The result of Model B in Panel B shows the prediction model from Model B statistically related to the return in next month. However, the R-square value is 0.158, which does not represent a good quality regression.

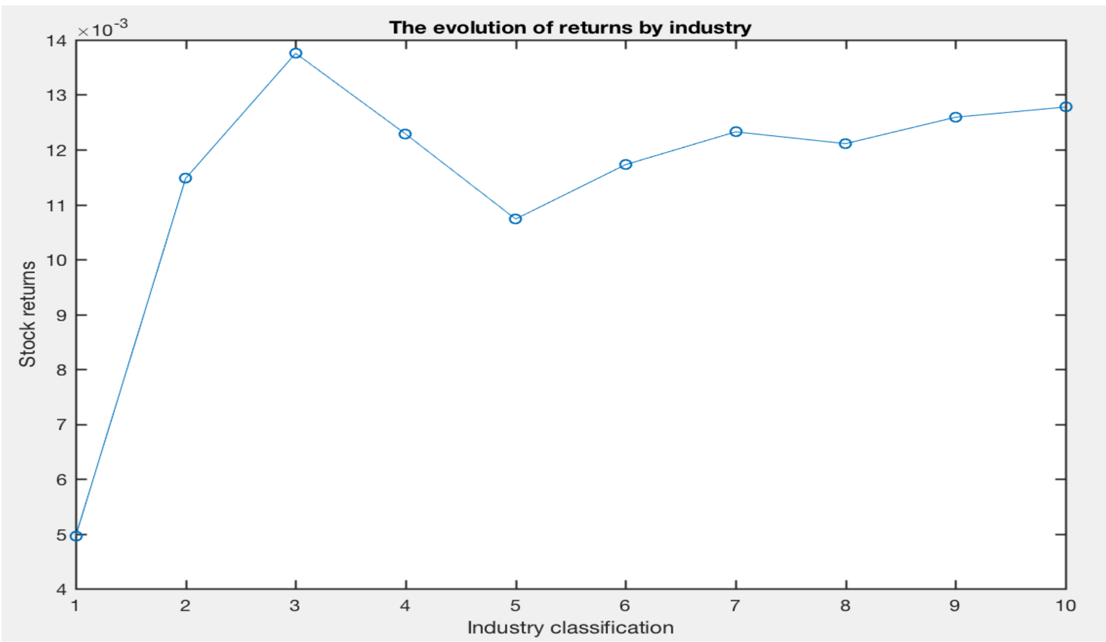
Panel B of univariate test indicates that regression of the returns predicted from risk factors with long-term observation windows has better quality than that of others, and the regression with individual β^{60m} has the highest R^2. Thus, we combine β^{24m}, β^{36m} and β^{48m}as explicative variables in Model C andβ^{36m}, and β^{48m} and β^{60m} are categorized as Model D. Due to the mutual effect among the risk factors with long-term observation period, Panel A of Table 5.7 shows that β^{36m} is positively related to returns in the next month in Model C but negatively related to the forward returns in Model D. The quality of regression is 15.2% and 15.5%, respectively. As shown in Panel B, the coefficients of risk factors for Model C and D are both significant, but the t-values of intercepts are not significant enough to demonstrate risk premium in the market.

Model E includes short-window and long-window risk factors and is established byβ^{21d}, β^{63d}, β^{252d}, β^{36m}, β^{48m}. Due to the assumption that the expectation of abnormal returns is null, we construct Model F including all seven risk factors without the constant (α). Our results in Panel B show the t-statistic value of coefficients for Model F is very significant, which is 4.5121 after correcting autocorrelation and heteroscedasticity. The quality of regression of Model F is the highest than that of others. Thus, we conclude the best combination of risk factors to predict the returns in next month should include all seven risk factors without constant parameter (α). Overall, in this final model, two observations can be made: first, the abnormal return should be neglected; second, the length of observation period of risk factors show a crucial effect on the prediction.

5.5.3 Industry factors

To study the effect of the industry factors to return prediction, we calculate the average returns for each month according to the industry classification depending on the SIC categories. Figure 5.5 shows the average returns by each industry. The participation of Agriculture, Forestry, and Fishing (SIC<10) and Public administration (SIC>90) are negligible due to lack of firms available in the sample. As a result, we exclude these industries to reduce the complexity of charts and improve the accuracy of prediction. From the Figure, construction has the highest mean value of returns and transportation, communications, electric, gas, and sanitation have the lowest returns during the period observed.

Figure 5.5: Stock returns by industry classification



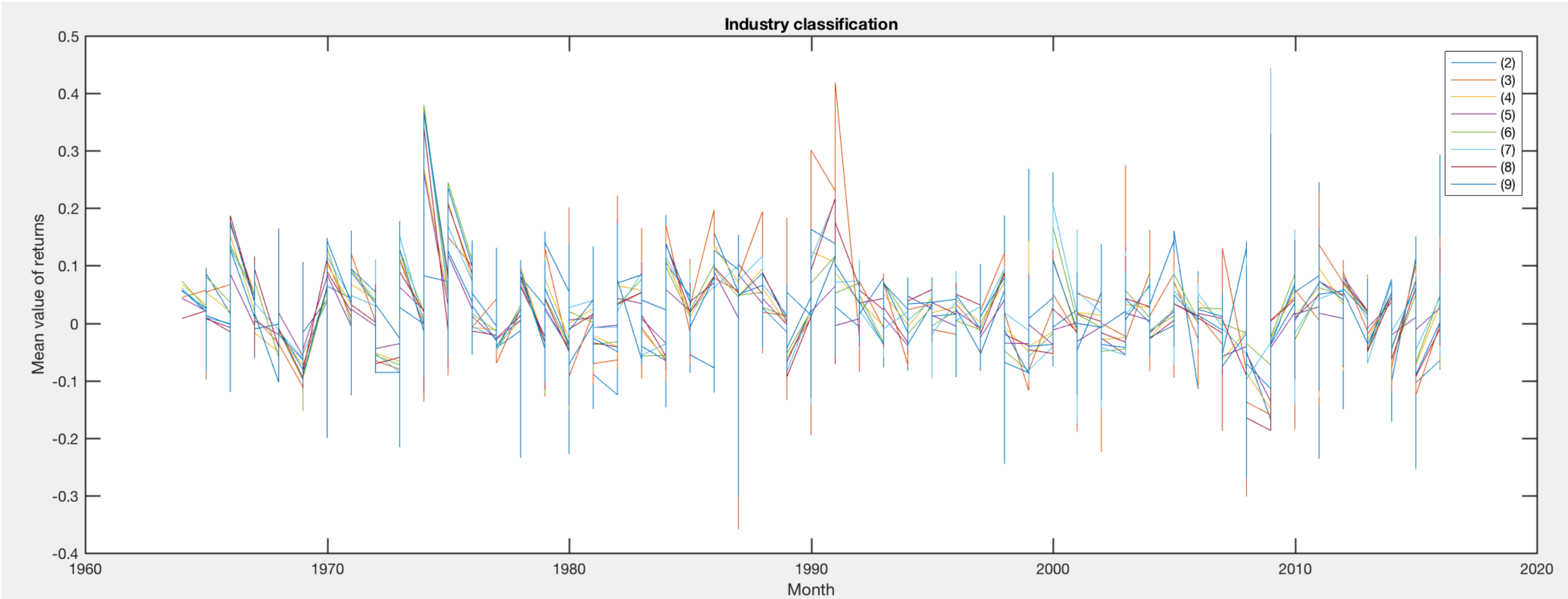
Notes: Figure 5.5 the average returns by each industry. The y-axis represents the mean value of stock returns for each industry and the x-axis represents the industry classification based on the Table 4.4.3.1

Figure 5.6 shows the evolution of stock returns through time varying and according to industry classification. We calculate the mean value of stock returns on the last day of each month for common stocks operating in each industry. From the results of Figure 5.6, we can see that these eight industries, in general, have a similar pattern and vary with the tendency of the market. For example, in 1987, most stocks in all industries dropped sharply to their lowest point. When comparing industries, the mining, retail and services sectors are the most fluctuant industries with high variation. In contrast, manufacturing, transportation, communications, electric, gas and sanitation along with the wholesale trade have relatively flat plots, which means their evolutions tend to be stable.

Based on the result of multivariate analysis, we first calculate the predicted returns for each industry by using the mean value of seven coefficients without the constant parameter and regress the estimated forward return with the realized ones. Table 5.8 reports the results of Fama-MacBeth cross-sectional regression (4.3.5.1) including eight parts of industry classification.

From Table 5.8, we can see that the t-statistic value of coefficients of prediction returns are obviously significant for all industries excluding the Retail trade. The result indicates that our prediction model is very suitable for the firms operating in the field of construction, transportation, communications, electric, gas and sanitation, with the quality of regression for these two last industries at 67.2% and 63.9%, respectively. Referring to the result from Figure 5.6, these figures could be explained by the stability of firms in the construction, transportation, communications, electric, gas and sanitation industries, especially with regard to their operations and when compared with other industries. Mining, retail and services are the most fluctuant industries which always tend to vary with market change, and thus the appropriation of the prediction model is lower than other industries, which are 32.4%, 26.2% and 21.4%. Thus, we could confirm here that our prediction model is more appropriate for the industries with stable tendencies when compared to the market, such as those in the field of construction, transportation, communications, electric, gas and sanitation.

Figure 5.6 The evolution of stocks by industry



Notes: Figure 5.6 reports the evolution of stock returns through time varying, according to the industry classification. We calculate the mean value of stock returns onthe last day of each month for common stocks operating in each industry. The y-axis represents the mean value of stock returns for each industry and the x-axis represents the month. The sample period for observation is from January 1965 to December 2016, there are thus 624 months for analyzing. "(2)" represents the evolution of returns for the industry of mining. "(3)" represents the evolution of returns for the industry of construction. "(4)" represents the evolution of returns for the industry of manufacturing. "(5)" represents the evolution of returns for the industry of transportation, communications, electric, gas, sanitation. "(6)" represents evolution of returns for the industry of the wholesale trade. "(7)" represents evolution of returns for the industry of the retail trade. "(8)" represents evolution of returns for the industry of finance, insurance, and real estate. "(9)" represents evolution of returns for the industry of service.

Table 5.8: Fama-MacBeth cross-sectional regression for industry classification

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\emptyset_0$	0.008***	0.012***	0.010***	0.009***	0.011***	0.008***	0.009***	0.011***
t-stat_OLS	(2.603)	(3.443)	(5.014)	(6.2469)	(5.195)	(3.566)	(4.605)	(4.596)
t-stat_NW	(2.516)	(3.273)	(4.711)	(6.035)	(4.951)	(3.303)	(4.325)	(4.248)
$\emptyset_1$	0.469	0.509	0.555	0.642	0.703	0.0051***	0.355	0.296**
t-stat_OLS	(-0.043)	(1.012)	(-0.435)	(1.355)	(0.837)	(3.142)	(-0.710)	(2.104)
t-stat_NW	(-0.039)	(0.923)	(-0.397)	(1.248)	(0.771)	(3.029)	(-0.654)	(1.946)
Adj. R-square	0.324	0.672	0.203	0.639	0.457	0.262	0.373	0.214

Notes: Table 5.8 reports the results of Ordinary least squares regression and Fama-MacBeth cross-sectional regression with the realized forward returns one month ahead of the date t for common stock listed on the NYSE and predicted returns estimated by the prediction of Model II from which we collect the time-varying alphas (α) and gammas (γ) of (4.3.1) and use the average value of alphas and coefficients of risk factors for each industry to estimate the predict returns in the next month (at t+1). Column (2) represents the regression of realized forward returns with predicted returns in the industry of mining. Column (3) represents the regression of realized forward returns with predicted returns in the industry of construction. Column (4) represents the regression of realized forward returns with predicted returns in the industry of manufacturing. Column (5) represents the regression of realized forward returns with predicted returns in the industry of transportation, communications, electric, gas, and sanitation. Column (6) represents the regression of realized forward returns with predicted returns in the industry of the wholesale trade. Column (7) represents the regression of realized forward returns with predicted returns in the industry of the retail trade. Column (8) represents the regression of realized forward returns with predicted returns in the industry of finance, insurance, and real estate. Column (9) represents the regression of realized forward returns with predicted returns in the industry of service. Row "$\emptyset_0$"is intercept of the regressions and Row "$\emptyset_1$"is the coefficient of predicted returns from models. Row "t-stat_OLS" is the t-statistics value of parameters from OLS regression and Row "t-stat_NW" is the t-statistics value of parameters after robustness test which correct the heteroscedasticity and autocorrelation. The t-stat for "$\emptyset_1$" is used for testing whether average $\emptyset_1$ is different from 1.Adj. R-square is considered as the indicator of the quality of each regression. The symbols of *, ** and *** represent significance of the result at the 90%, 95% and 99% levels respectively, based on the t-statistics value from Newey-West test.

VI. Conclusion

In our study, based on the CAPM model, we have developed a new nonparametric methodology for returns prediction, from the conditional risk factor model. After reviewing the literature related to the problematic caused by using the traditional CAPM in empirical tests and the proposition of conditional beta coefficient with time-varying, we have conducted extensive research and found the answer to the following questions: What is the difference among the risk factors using the past returns with different observation windows? What is the effect of these risk factors with different observation windows on the realized forward returns one-month ahead of the date? What is the optimal method to estimate the forward returns? Which risk factors could be combined to form an optimal prediction model, according to the length of the observation windows of risk factors? And furthermore, are there industry factors which have crucial effect on the return estimation?

The subject of this study focuses on stock returns of mature companies listed on the New York Stock Exchange (NYSE): the NYSE market index reflects the overall market situation and could be considered as the standard market index in our study. The total number of companies to be analyzed is 3829 with the sample period for observation running from January 1965 to December 2016; there are thus 624 months for analyzing.

We have constructed seven risk-factors variables using respectively the past 21-day daily returns, 63-day daily returns, 252-day daily returns, 2-year monthly returns, 3-year monthly returns, 4-year monthly returns and 5-year monthly returns. There exist differences among the risk factors with various observation windows. In general, betas demonstrate a tendency toward normal distribution; however, beta coefficients using short-term returns in the past have higher kurtosis, which indicates that such risk factors have more abnormal betas than predicted. Further analysis on the serial correlation of risk factors leads us to the conclusion that the risk factors with long-run observation windows, such as $\beta^{252d}, \beta^{24m}$, β^{36m}, β^{48m}, β^{60m}, calculated with returns more than one year in duration, are more dependent on the past performance of conditional factor loadings, and that the adjusted R-square of the regression for $\beta^{252d}, \beta^{24m}$, β^{36m}, β^{48m} and β^{60m} are higher than that of others, which are 89.4%, 89.7%, 89.0% and 87.7%, respectively.

When observing the results of the cross-sectional regression of realized forward returns one-month ahead of the date t with seven beta coefficients with short-term and long-term past returns, we can trace the pattern of coefficients with time varying, which captures the overall market information in the given period. The Fama-MacBeth cross-sectional regression tests three types of prediction models and points to the fittest one with realized returns in the market -- Model II. Here, we collect the time-varying alphas (α) and gammas (γ) and use the average value of alphas and seven coefficients of risk factors to estimate the predicted returns in next month (at t+1).

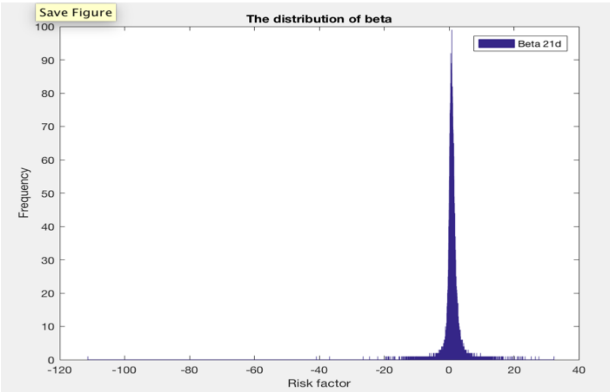
Further study shows the practicability of the prediction model. The figure of the evolution of constant factor and individual coefficient, correspond to the seven risk factors with different observation windows over time, and indicate serial correlation among the risk factors. The result of univariate analysis shows that the quality of regression increases with the length of observation windows of each individual risk factor. Moreover, the prediction model with individual betas calculated with five-year monthly returns are more appropriate with the realized forward returns, because long observation periods tend to be more stable and can capture the common situation of the capital market. The multivariate analysis continues the study of the combination of risk factors on the prediction model. After correcting the heteroscedasticity and autocorrelation among the risk factors by Newey-West, we conclude that the optimal prediction model formed by the combination of risk factors entails the use of seven risk factors with different observation periods and without the constant factor. As abnormal returns in capital markets are random and expectations should equal zero, we therefore do not consider abnormal returns when estimating forward returns.

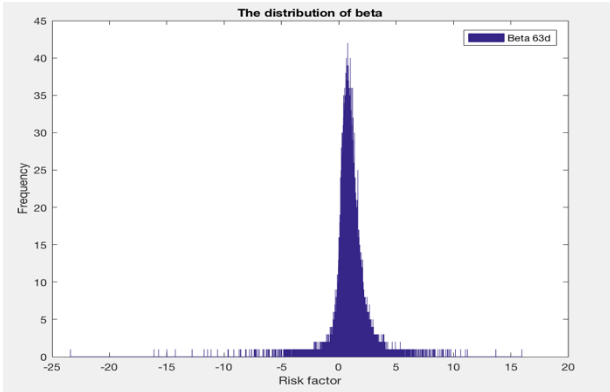
We also take the industry factor into consideration. Due to limitation of samples, we classified securities into eight groups. The Mining, retail, and services sectors are the most fluctuant industries, experiencing high variation. On the hand, manufacturing, transportation, communications, electric, gas and sanitation, along with the wholesale trade, tend to be stable, according to the flat plots, which means they evolve down a slow and steady path. . To put it another way, industry factors have an impact on the predicted returns estimated by the prediction model. Ultimately, industries such as construction, transportation, communications, electric, gas, and sanitation, in short industries with a stable tendency, when compared to the market, are more appropriate for our prediction model.

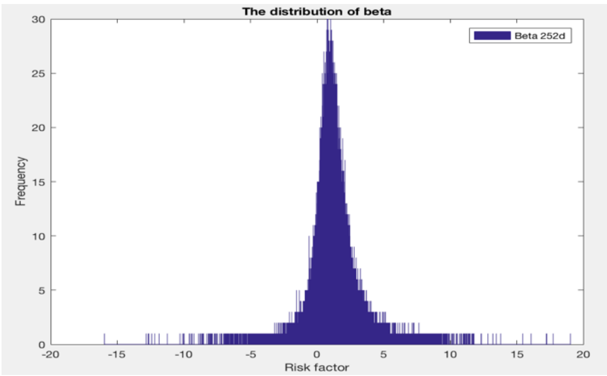
To conclude, the limitations of this study can be summarized in the following statements Firstly, this work focuses mainly on the returns of common stock listed on the NYSE and in the US capital market. In future studies, we could expand the examination to include the NASDAQ and other countries' capital markets, like S&P 500 in Canada. Secondly, this paper examines only the beta-coefficients and the stock returns at the individual securities level. We did not conduct, for example, a conclusive study of risk factors for combined portfolios owing to the possibility of the latter losing the basic characteristics of the company, which could adversely affect the results of the study. Thirdly, this paper investigates mainly the prediction model to estimate forward stock return by using the combination of risk factors with different observation period. The direction of future inquiries could focus on the predictability of beta coefficient by using the combination of risk factors as defined in this study. In the end, we can safely say that this field offers the scholar a myriad of possible subjects to explore.

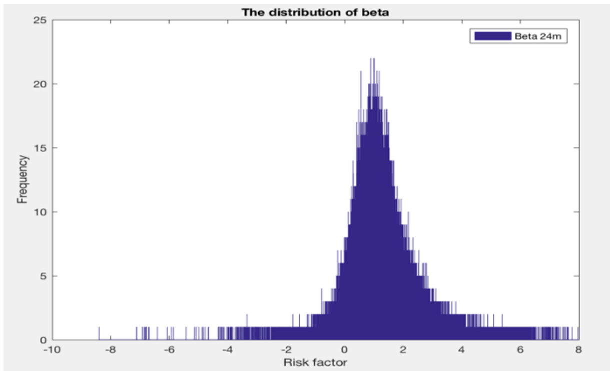
Appendix

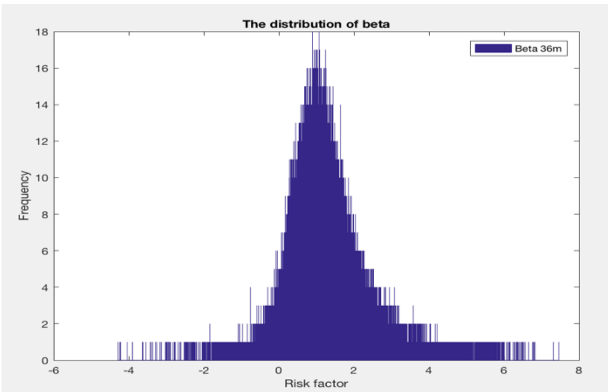
Appendix 1 The distribution of constructed risk factors

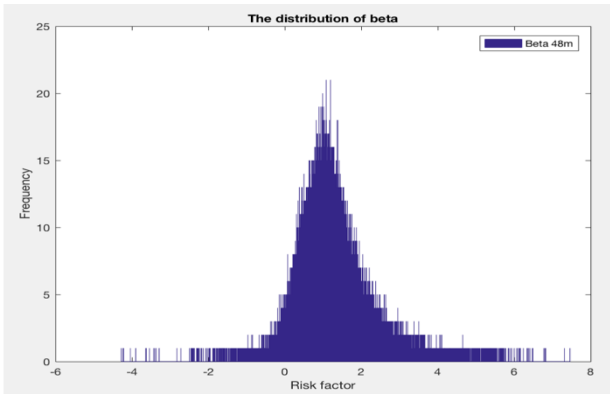


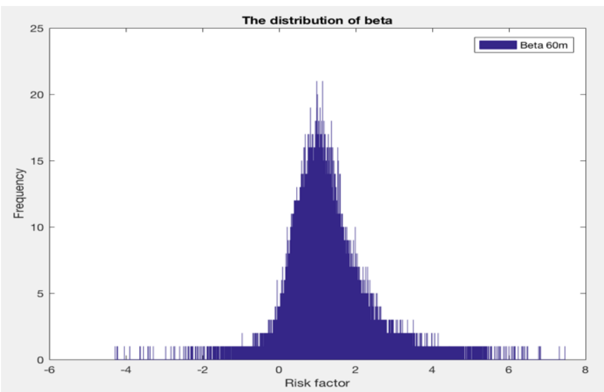












Appendix 2 Number of securities listing in NYSE

Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE
19641231	1080	19680229	1121	19710430	1209	19740628	1350	19770831	1373	19801031	1347
19650129	1081	19680329	1120	19710528	1210	19740731	1355	19770930	1376	19801128	1347
19650226	1083	19680430	1114	19710630	1207	19740830	1362	19771031	1373	19801231	1344
19650331	1091	19680531	1107	19710730	1211	19740930	1359	19771130	1380	19810130	1345
19650430	1095	19680628	1108	19710831	1214	19741031	1364	19771230	1375	19810227	1343
19650528	1097	19680730	1102	19710930	1217	19741129	1365	19780131	1368	19810331	1340
19650630	1103	19680830	1103	19711029	1219	19741231	1368	19780228	1369	19810430	1340
19650730	1111	19680930	1098	19711130	1223	19750131	1373	19780331	1368	19810529	1337
19650831	1116	19681031	1095	19711231	1224	19750228	1374	19780428	1367	19810630	1334
19650930	1120	19681129	1096	19720131	1225	19750331	1373	19780531	1365	19810731	1334
19651029	1116	19681231	1093	19720229	1230	19750430	1373	19780630	1365	19810831	1334
19651130	1119	19690131	1091	19720330	1235	19750530	1373	19780731	1368	19810930	1336
19651231	1116	19690228	1089	19720428	1236	19750630	1373	19780831	1367	19811030	1330
19660131	1118	19690328	1088	19720531	1238	19750731	1372	19780929	1372	19811130	1332
19660228	1119	19690430	1089	19720630	1246	19750829	1375	19781031	1366	19811231	1330
19660331	1124	19690529	1095	19720731	1254	19750930	1376	19781130	1365	19820129	1332
19660429	1122	19690630	1094	19720831	1258	19751031	1377	19781229	1363	19820226	1330
19660531	1124	19690731	1099	19720929	1268	19751128	1376	19790131	1359	19820331	1324
19660630	1123	19690829	1109	19721031	1276	19751231	1372	19790228	1356	19820430	1320
19660729	1126	19690930	1107	19721130	1281	19760130	1374	19790330	1356	19820528	1318
19660831	1133	19691031	1121	19721229	1285	19760227	1372	19790430	1358	19820630	1316
19660930	1137	19691128	1127	19730131	1289	19760331	1370	19790531	1358	19820730	1318
19661031	1138	19691231	1126	19730228	1297	19760430	1368	19790629	1355	19820831	1314
19661130	1137	19700130	1133	19730330	1300	19760528	1367	19790731	1355	19820930	1311
19661230	1138	19700227	1141	19730430	1305	19760630	1366	19790831	1356	19821029	1315
19670131	1140	19700331	1143	19730531	1308	19760730	1365	19790928	1353	19821130	1317
19670228	1144	19700430	1152	19730629	1310	19760831	1346	19791031	1356	19821231	1313
19670331	1145	19700529	1157	19730731	1310	19760930	1365	19791130	1358	19830131	1311
19670428	1144	19700630	1164	19730831	1316	19761029	1347	19791231	1355	19830228	1312
19670531	1139	19700731	1170	19730928	1320	19761130	1369	19800131	1357	19830331	1313
19670630	1137	19700831	1175	19731031	1328	19761231	1367	19800229	1355	19830429	1311
19670731	1137	19700930	1180	19731130	1328	19770131	1365	19800331	1351	19830531	1313
19670831	1138	19701030	1186	19731231	1334	19770228	1366	19800430	1346	19830630	1312
19670929	1140	19701130	1187	19740131	1340	19770331	1367	19800530	1349	19830729	1318
19671031	1138	19701231	1190	19740228	1347	19770429	1367	19800630	1350	19830831	1316
19671130	1138	19710129	1191	19740329	1351	19770531	1365	19800731	1355	19830930	1314
19671229	1138	19710226	1198	19740430	1348	19770630	1366	19800829	1353	19831031	1309
19680131	1125	19710331	1201	19740531	1348	19770729	1363	19800930	1350	19831130	1310

Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE
19831230	1306	19870227	1220	19900430	1172	19930630	1327	19960830	1479	19991029	1517
19840131	1303	19870331	1215	19900531	1172	19930730	1336	19960930	1479	19991130	1503
19840229	1297	19870430	1213	19900629	1176	19930831	1347	19961031	1479	19991231	1493
19840330	1298	19870529	1213	19900731	1178	19930930	1353	19961129	1483	20000131	1498
19840430	1291	19870630	1205	19900831	1180	19931029	1356	19961231	1484	20000229	1497
19840531	1294	19870731	1203	19900928	1186	19931130	1357	19970131	1494	20000331	1485
19840629	1295	19870831	1205	19901031	1188	19931231	1360	19970228	1495	20000428	1483
19840731	1290	19870930	1205	19901130	1190	19940131	1364	19970331	1495	20000531	1482
19840831	1292	19871030	1209	19901231	1193	19940228	1370	19970430	1498	20000630	1477
19840928	1287	19871130	1204	19910131	1200	19940331	1374	19970530	1499	20000731	1476
19841031	1284	19871231	1204	19910228	1207	19940429	1380	19970630	1493	20000831	1474
19841130	1280	19880129	1215	19910328	1206	19940531	1386	19970731	1501	20000929	1467
19841231	1280	19880229	1213	19910430	1213	19940630	1393	19970829	1491	20001031	1464
19850131	1284	19880331	1214	19910531	1215	19940729	1404	19970930	1488	20001130	1451
19850228	1279	19880429	1208	19910628	1222	19940831	1408	19971031	1492	20001229	1447
19850329	1278	19880531	1199	19910731	1228	19940930	1411	19971128	1502	20010131	1443
19850430	1286	19880630	1189	19910830	1232	19941031	1415	19971231	1507	20010228	1440
19850531	1284	19880729	1190	19910930	1232	19941130	1421	19980130	1517	20010330	1438
19850628	1285	19880831	1191	19911031	1239	19941230	1419	19980227	1519	20010430	1433
19850731	1286	19880930	1191	19911129	1240	19950131	1429	19980331	1525	20010531	1429
19850830	1285	19881031	1188	19911231	1240	19950228	1434	19980430	1521	20010629	1424
19850930	1284	19881130	1190	19920131	1243	19950331	1435	19980529	1521	20010731	1423
19851031	1282	19881230	1187	19920228	1247	19950428	1436	19980630	1512	20010831	1416
19851129	1277	19890131	1185	19920331	1246	19950531	1438	19980731	1514	20010928	1411
19851231	1267	19890228	1186	19920430	1248	19950630	1439	19980831	1520	20011031	1406
19860131	1265	19890331	1188	19920529	1259	19950731	1447	19980930	1525	20011130	1401
19860228	1262	19890428	1187	19920630	1264	19950831	1450	19981030	1520	20011231	1399
19860331	1262	19890531	1187	19920731	1271	19950929	1450	19981130	1523	20020131	1389
19860430	1254	19890630	1185	19920831	1273	19951031	1449	19981231	1526	20020228	1385
19860530	1253	19890731	1181	19920930	1277	19951130	1449	19990129	1533	20020328	1381
19860630	1247	19890831	1179	19921030	1275	19951229	1449	19990226	1532	20020430	1380
19860731	1245	19890929	1176	19921130	1280	19960131	1463	19990331	1531	20020531	1378
19860829	1241	19891031	1177	19921231	1293	19960229	1463	19990430	1531	20020628	1384
19860930	1242	19891130	1177	19930129	1300	19960329	1465	19990528	1534	20020731	1386
19861031	1233	19891229	1178	19930226	1302	19960430	1470	19990630	1526	20020830	1390
19861128	1232	19900131	1176	19930331	1305	19960531	1472	19990730	1524	20020930	1395
19861231	1221	19900228	1173	19930430	1316	19960628	1473	19990831	1527	20021031	1391
19870130	1221	19900330	1172	19930528	1325	19960731	1478	19990930	1529	20021129	1386

APPENDIX

Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE	Date	NYSE
20021231	1381	20060228	1346	20090430	1246	20120629	1215	20150831	1082
20030131	1386	20060331	1341	20090529	1249	20120731	1217	20150930	1077
20030228	1386	20060428	1342	20090630	1252	20120831	1216	20151030	1071
20030331	1379	20060531	1341	20090731	1255	20120928	1217	20151130	1067
20030430	1374	20060630	1335	20090831	1255	20121031	1210	20151231	1059
20030530	1377	20060731	1336	20090930	1257	20121130	1210	20160129	1050
20030630	1382	20060831	1334	20091030	1261	20121231	1205	20160229	1045
20030731	1384	20060929	1336	20091130	1259	20130131	1205	20160331	1037
20030829	1381	20061031	1337	20091231	1258	20130228	1200	20160429	1028
20030930	1385	20061130	1335	20100129	1256	20130328	1197	20160531	1024
20031031	1384	20061229	1331	20100226	1254	20130430	1193	20160630	1021
20031128	1385	20070131	1326	20100331	1250	20130531	1189	20160729	1012
20031231	1383	20070228	1324	20100430	1249	20130628	1186	20160831	1009
20040130	1384	20070330	1320	20100528	1246	20130731	1181	20160930	1002
20040227	1386	20070430	1319	20100630	1247	20130830	1176	20161031	1000
20040331	1385	20070531	1315	20100730	1241	20130930	1171	20161130	995
20040430	1382	20070629	1308	20100831	1239	20131031	1163	20161230	990
20040528	1377	20070731	1306	20100930	1238	20131129	1159		
20040630	1374	20070831	1299	20101029	1232	20131231	1155		
20040730	1370	20070928	1298	20101130	1233	20140131	1149		
20040831	1366	20071031	1289	20101231	1232	20140228	1146		
20040930	1367	20071130	1281	20110131	1232	20140331	1145		
20041029	1365	20071231	1275	20110228	1237	20140430	1143		
20041130	1369	20080131	1283	20110331	1233	20140530	1139		
20041231	1365	20080229	1279	20110429	1231	20140630	1138		
20050131	1369	20080331	1280	20110531	1228	20140731	1135		
20050228	1371	20080430	1278	20110630	1222	20140829	1133		
20050331	1366	20080530	1277	20110729	1220	20140930	1132		
20050429	1366	20080630	1278	20110831	1222	20141031	1129		
20050531	1366	20080731	1278	20110930	1226	20141128	1127		
20050630	1363	20080829	1278	20111031	1223	20141231	1123		
20050729	1362	20080930	1276	20111130	1222	20150130	1114		
20050831	1365	20081031	1273	20111230	1215	20150227	1110		
20050930	1364	20081128	1264	20120131	1213	20150331	1103		
20051031	1355	20081231	1258	20120229	1218	20150430	1101		
20051130	1353	20090130	1255	20120330	1217	20150529	1099		
20051230	1348	20090227	1256	20120430	1215	20150630	1094		
20060131	1343	20090331	1249	20120531	1215	20150731	1087		

References

Andrew, M., 2012. Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability. *Review of Financial Studies 25*, 2789–2839.

Ang, A., Chen, J., 2007. CAPM over the Long Run: 1926–2001. *Journal of Empirical Finance* 14, 1–40.

Balvers, R.J., Huang, D., 2009. Evaluation of Linear Asset Pricing Models by Implied Portfolio Performance. *Journal of Banking and Finance* 33, 1586–1596.

Blume, M. E., Stambaugh, R. F., 1983. Biases in Computed Returns: An Application to the Size Effects. *Journal of Financial Economics 12 (3)*, 387-404

Brownlees, C., and R. Engle. 2016. SRISK: A Conditional Capital Shortfall Measure of Systemic Risk.

Bali, Turan G., and Robert F. Engle. 2010. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. *Journal of Monetary Economics*, 57: 377–390

Corsi, F., 2004. A Simple Long Memory Model of Realized Volatility

Dennis, K., 2010. Testing Conditional Factor Models.

Engle, R., 2014. “Modeling Commodity Prices with Dynamic Conditional Beta.” *Oxford University Press*, Chapter 11, pp. 269–287

Fama, E., French, K., 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, 3–56.

Fama, E., French, K., 1997. Industry Costs of Equity. *Journal of Financial Economics* 43, 153–193.

Fama, E., MacBeth, J., 1973. Risk, Return and Equilibrium: Empirical Tests. *Journal of*

Political Economy 81, 607–636

Grauer, R.R., Janmaat, J.A., 2009. On the Power of Cross-sectional and Multivariate

Tests of the CAPM. *Journal of Banking and Finance 33*, 775–787

Hou, K., Moskowitz, T., 2005. Market Frictions, Price Delay, and the Cross-section of Expected Returns. Review of Financial Studies 18 (3), 981{1020.

Jensen, M., 1968. The Performance of Mutual Funds in the Period 1954 – 1964. *Journal of Finance* 23, 389–416.

Jones, C., 2006. A Nonlinear Factor Analysis of S and P 500 Index Option Returns. *Journal of Finance* 41, 2325–2363.

Kapetanios, G., 2007. Estimating Deterministically Time-varying Variances in Regression Models. *Economics Letters* 97, 97–104.

Lintner, J., 1965. The Valuation of Risk Assets and the Selection of Risky Investments in

Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 13–27.

Lewellen, J., Nagel, S., 2006. The Conditional CAPM Does Not Explain Asset-Pricing Anomalies. *Journal of Financial Economics* 82 (2), 289-314.

Merton, R., 1980. On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics* 8, 323–361.

Patton, Andrew, and Michela Verardo. 2012. Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability.

Pakel, Cavit, Niel Shephard, Kevin Sheppard, and Robert Engle. 2014. “Fitting Vast Dimensional Time-varying Covariance Models.”

Robinson, P., 1989. Nonparametric Estimation of Time Varying Parameters. *Statistical Analysis and Forecasting Economic Structural Change*.

Robert, E., 2009. Anticipating Correlations: A New Paradigm for Risk Management. *Princeton University Press*.

Sharpe, W., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions

of Risk. *Journal of Finance* 33, 885–901.

Scholes, M. and J. Williams, 1977, “Estimating Betas from Non-Synchronous Data,” *Journal of Financial Economics*, 5, 309-328.

Stanton, R., 1997. A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance* 52, 1973–2002.

Thomas, G., Christopher, H., and Jonathan, K., 2013. Daily Data is Bad for Beta: Opacity and Frequency-Dependent Betas

Tse, Y. K., and Albert K. C. Tsui. 2002. A Multivariate Generalized Autoregressive Conditioinal Heteroscedasticity Model with Time-Varying Correlations. *Journal of Business and Economic Statistics* 20: 351–362.

Vuong, Quang. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* 57: 307–333.

van Dijk, M.A., 2011. Is Size Dead? A Review of the Size Effect in Equity Returns. *Journal of Banking and Finance* 35, 3263–3274.

Wang, K., 2003. Asset Pricing with Conditioning Information: a New Test. *Journal ofFinance* 58, 161–196.