ISCTE Business School
Instituto Universitário de Lisboa

Industry Classification Benchmark and Industries' Predictive Ability

Yingwen Yu

Dissertationsubmitted as partial requirement for the conferral of
Master in Finance

Supervisor: Prof. Sofia Correia Brito Ramos

21st April, 2014

-Spine-

Industry Classification Benchmark and Industries' Predictive Ability

Yingwen Yu

ISCTE Business School
Instituto Universitário de Lisboa

Index

- **Acknowledgements**

I am extremely grateful to Professor Sofia Correia Brito Ramos for her supervision, help and useful common to my thesis. From the data selecting to the thesis revision, she gives me many advisements which are very useful to improve my thesis.

- **Abbreviations**

CPI: consumer price index;

CRSP: the center for research in security prices;

FF: Fama French industry classification;

PMI: purchasing manager index;

GICS: global industry classification standard;

NAICS: north American industry classification system;

NASDAQ: national association of securities dealers automated quotations;

NYSE: New York stock exchange;

PER: price-earnings ratio;

SIC: standard industrial classification.

- **Glossary**

Compustat: is a database of financial, statistical and market information on active and inactive global companies throughout the world. The service began in 1962.

FTSE: called as FTSE 100, is a share index of 100 companies listed on the London Stock Exchange with the highest market capitalization.

Abstract

In this dissertation, we investigate Industry Classification Benchmark (ICB) to test if the industries are able to predict the market portfolio or not, and if so how many months ahead do they forecast? According to the final results, several conclusions have been made. Firstly, industries could forecast the aggregated market with information variables, such as Default Spread (BAA rate and AAA rate), Dividend Yield, Market Volatility and PER (price earnings ratio). Secondly, only five of thirty-two industries are able to predict the market portfolio in one month ahead, they are support services, food & drug retails, general retails, mobile telecommunications and nonlife insurance, respectively. Furthermore, those industries in the same supersectors of Industry Classification Benchmark possess the homological predictive power to the aggregated market portfolio.

Key Words: Industry Classification Benchmark, Predictive Abilities, Industrial Lags, Industrial Correlation.

Resumo

Neste tese eu investigo a classificação de industrias:IndustryClassificationBenchmark (ICB) para testar se as industrias conseguem prever as variações da carteira de mercado, e se sim com quantos meses de avanço.

Os resultados mostram em primeiro lugar que as industria podem prever para além das variáveis normalmente usadas tais como: oDefaultSpread (a diferença entre o spread das taxas BAA e AAA), a DividendYield, a volatilidade do mercado e opriceearnings ratio (PER). Em segundo lugar só 5 das 32 industrias são capazes de prever a carteira de mercado com um mês de avanço, elas são: supportservices, food&drugretails, general retails, mobile telecommunications enonlifeinsurance. Para além disso, estas industrias no mesmo supersector daIndustryClassificationBenchmarktêm o mesmo poder preditivo homólogo para o mercado.

1. Introduction

Two aims are going to be achieved in this dissertation, one is comparing the difference among the most important and commonly utilizing industry classifications, such as Standard Industrial Classification (SIC), the Global Industry Classification Standard (GICS), the North American Industry Classification System (NAICS), the Fama-French industry (FF) and Industry Classification Benchmark (ICB). The other is to investigate if there is any predictive ability possessed by industries based on the Industry Classification Benchmark.

To begin with conventional classifications, in the financial market, SIC, GICS, NAICS and Fama-French industry are the more traditional classifications and accepted widely by various professors and researchers. As Industry Classification Benchmark (ICB) is the new benchmark, it has already been accepted to use by many financial associations, like NASDAQ and NYSE comparing to the conventional classifications we mentioned above. However, only few researches have been done for ICB for testing its accuracy and scientificity, therefore, we are going to compare the different benchmark firstly, and then find out their merit and demerit in the first part of my dissertation.

Industry Classification segregates market into different groups founded on the similar products, homological property or other features. However, dissimilar classifications react differently to some specific indicators, such as variance on financial ratios or homogeneity of firms. In order to fix this problem, financial researchers are interested in comparing the diversified taxonomies and testing their efficacy to financial ratios. As Guenther and Rosman (1994) find out the difference between SIC codes across Compustat and CRSP (both of them belongs to SIC codes), CRSP SIC codes have lower intra-industry correlation in stock returns and higher intra-industry variance in financial ratios than Compustat SIC codes (Bhojraj et al., 2003).

As for the importance of industries taxonomies, of course, industrial classifications make a significant contribution on financial research analysis. Its importance is indubitable. For instance, in a period, from 1992 to 1995, there were at least 81 articles using industry classification in their tests and being published in the financial journals (Kahle and Walkling, 1996). In these publishing articles, 31% of them use Compustat SIC codes, while 26% of them prefer to use CRSP SIC codes. Beside, almost a quarter of samples use other methods except Compustat and CRSP. Financial researchers use industry classifications for four differently subsequent purposes. Thirty-eight articles(48% of sample) are for identifying control sample, twenty-eight(35% of sample) articles describe industrial structure via using industry classification, twenty-six(32%) articles use industry taxonomy to restrict sample, and last seven article(7%) categorize acquisitions and divestitures as conglomerate or no conglomerate.

Moreover, in the second part of the dissertation, some properties of industries are going to be investigated. Whether the industries predict the stock market or not is our core hypothesis. The subsequent two methods are used to demonstrate this propensity, one is testing predictive ability by using information variables in order to contain and reflect more information about the industry, market and macro-economy circumstance; meanwhile, the other focuses on how many months do industries lead the market portfolio ahead, namely that the industries may have strong or weak power to forecast the market's movement.

A host of previous reports show that researchers investigate the movement of the market and predict it by using lots of various information variables which correlated with the economic movement closely. How important are the information variables to the research after all? The results in Hong and Stein (1999) conclude more evidence about the significance of information variables. "We model a market populated by two groups of boundary rational agents: 'newswatchers' and 'momentum traders.'

Each newswatcher observes some private information, but fails to extract other newswatchers' information from prices. If information diffuses gradually across the population, prices under react in the short run. The relation between information and price becomes more active than we think, such important factors are ought to be considered into the regression to control the sample and assure test's precise definitely. What's more, Estrella and Mishkin (2009) test following relative components predicting the U.S recession, interest rate, spread, stock price and monetary aggregates, to investigate how information affects market's return. They select those variables from large numbers of indicators on candidate list and demonstrate that those variables which are reliable, elaborate and quick indicators are served as the supplement of the financial model. Last but not least, to the other related report, Keim and Stambaugh (1985) denotes that three variables, Treasury Bills, 20-years government bonds and value-weighted portfolio of New York Stock Exchange (NYSE) common stocks have predictive ability of stocks' return and bonds. In Hou (2007), researcher finds out that the diffusion of information is able to generate the lead-lag effect and influence the stock returns. Beside, this effect results in the returns on large cap leading the returns on small cap. Moreover, other paper showed more evident supporting this argument. Firm-specific information, especially negative information, diffuses only slightly across the investing public (Hong, Lim and Stein, 2000).So the information variables representing messages and information from market ought to have incomparable influence on aggregated movement and macro-economy.

In our research, we select the following six information variables in the regression, Treasury Bills, CPI (consumer price index), Default Spread (difference between BAA rate and AAA rate), Dividend Yield, volatility and PER to be the objects for our test from a long candidate list, in view of the fact that they account for the macro economy activities for the market portfolio and they are working as an market information indicators in normal industrial data analysis. For the results of our research, finding out that all the industries generate better linear regressions when adding more information variables we select, but only Default Spread, Dividend Yield, volatility and PER possess the forecasting ability in all the industries, CPI is only available in three industries, while Treasury Bills do not show any ability in predicting.

In addition, after demonstrating the exits of predictive ability for industries, testing how long do industries lead the return of stock market will generate an important conclusion with practical significant meaning because in the real market, leading market for one month is already very terrific. To be surprised, some industries with large capitalization like OILGPUS (oil & gas producers), CHMCLUS (chemicals) and INDMTUS (industrial metal & mining) they hardly present any predictive ability to the aggregated market at all. However, the industries SUPSVUS (support services), FDRGRUS (food & drug retails), GNRETUS (general retails), TELMBUS (mobile telecommunications) and NLINSUS (nonlife insurance) are capable to forecast the market's movement in one month ahead.

Above all, forecasting the market does have its own inestimable meanings. Firstly, it can predict the financial crisis and recession, as well as maximally cut down the losses. As we know that, the subprime mortgage crisis has caused a globally economic recession since 2007, nowadays, majority countries are still suffering the damages it caused or just start to recover from it slightly and slowly. However, if industries could predict this terrible movement, much more loss and damages could be avoided ideally. Secondly, if this predictive ability works with the precise process, it denotes that the Industrial Classification Benchmark outperforms several conventional ones, because it divides industries matching with the objective law of the market and complying with the practical market movement.

What's more, due to the classifications divide the market based on industries' properties, those industries correlating with each other exceedingly closed may have similar predictive propensity. Especially in the same supersectors of ICB, those industries are demonstrated to have analogous predictive ability of the market's movement. All in all, integrating the results of fundamental and further tests, the predictive ability the industries possess has been confirmed in our investigation.

2. Literature Review

The industry classification schemes, SIC, NAICS, GICS, FF industry classification and ICB are the most acceptable and useful among congeners. However, they hardly agree with each other congruously in many aspects, such as in applying or in researching. In this section, we will review the previous researches based on the different classifications to present their characteristics and scope of application, to compare their difference, as well as their merit and demerit.

2.1.Definition of industrial classifications

In order to introduce these classifications comprehensively, we provide some background information of five classifications (SIC, NAICS, GICS, FF and ICS), including their historical development, aim of usage, range of application and current availability.

2.1.1. SIC Codes(Standard Industrial Classification)

The SIC codes, being established in 1937 by the Interdepartmental Committee, is the oldest system among other four taxonomies. It classified the aggregated market by four-digit code. More specifically, the first three digits indicate the industry group and the first two digits represent the major group. The goal of establishing SIC codes is to develop it for statistical data and to promote it to be adopted by Federal Government (Pearce, 1957). Although it was used to research and develop for governmental usage, this classification was widely accepted by marketers and financial economists in the following years. As an example, this widely utilizing system reflects the economy's movement periodically and industrial organization. (Bhojraj, Lee and Oler, 2003). Due to its advanced property and creativity at that time, it was always a predominant algorithm in the market until NAICS being a substitute for it in 1990s, but U.S

Securities and Exchange Commission (SEC) still use SIC codes so far.

2.1.2. NAICS Codes (North American Industry Classification System)

In 1999, NAICS Code, which was implemented by Statistical Bureaus of United State, Canada and Mexico, newly includes 358 industries and rearranges SIC categories. The criterion of NAICS was productive. However, it has more extensive ranges covering more industries and more details in industry level than SIC codes. NAICS uses six-digits to classify the monetary market. To do so, the first two digits represent the largest business sectors, the third digit indicates the subsector, the fourth, fifth and sixth digits represent the industry group, NAICS industries and national industries respectively. Its propensities are listed following, firstly, NAICS is a production-based framework; Secondly, it identifies the new industries and regroup industries classification scheme from SIC aiming to adjust the real-time movement of economic market (Saunders, 1999). More professors and researchers may prefer to use NAICS definitions rather than SIC, because it is more advanced and its definitions lead to more cohesive industries (Jayanthi, 2003). Hence, Clarke (1989), as well as Amit and Livnat (1990) prove that NAICS outperforms SIC in terms of homologous grouping.

2.1.3. GICS Codes (Global Industry Classification Standard)

In 2008, GICS was created by a combination Standard & Poor's (S&P) and Morgan Stanley Capital International (MSCI). Ten sectors, twenty-four industry groups, sixty-eight industries and one hundred and fifty- four sub-industries in total constitute to GICS structure. Depending on the financial statements and annual reports of firms, GICS is much easier to compare the performance of companies. As soon as the information is newly updated, market researchers can define the context immediately.

The aim of this classification is that “to enhance the investment research and asset management process for financial professionals worldwide”(S&P and MSCI, 2002).

Different from GICS, both SIC and NAICS relying on production and technology oriented of industry are not too intuitionist to observe the movement and performance of market or industry itself. As a leading taxonomy of stock indexes and benchmark, GICS guides us in the market - oriented perspective and leads to a different inference by using the financial research data. For example, using SIC system to analyze discretionary accruals (DA) may lead to an incorrect result. But taking the place of the SIC system by GICS can eliminate the bias and demonstrate that GICS is a better measure in earnings management (Hrazdil and Scott, 2013).

2.1.4. FF Industry Classification (Fama French Industry Classification)

FF industry classification reclassified SIC codes into 48 industries groups (Fama and French, 1997), and is developed by financial academics. As we all know, Fama and French are famous for their three factors theory in investment. FF industry classification’s aim is to rearrange SIC codes considering the common risk characteristics (Bhojraj et al., 2003), however, both two classifications react differently to some industries’ movement. For example, more recently Fama and French classification begins with firms’ 4-digit SIC codes and classifies them into 48 industry groups. The FF industry classification was used to control samples in asset pricing mostly comparing to other taxonomies. (Chan, Lakonishok and Seaminathan, 2007).

2.1.5. ICB (Industry Classification Benchmark)

ICB was launched by Dow Jones and FTSE in 2005. It is a relatively new benchmark comparing to the alternative classifications. The ICB classifies the aggregated market into 10 industries, dividing them into 19 supersectors, and then subdivided them into 41 sectors, to end up with containing 114 subsectors (A Guide to the Industry Classification Benchmark, 2012). The exact details are showed in the below table 1. ICB's clearly hierarchical structure leads to the more scientific and precise grouping of industries.

Moreover, ICB, which is a newly-built measure to group the market, provides us much incomparable superiorities. First of all, ICB system covers over 70,000 companies and 75,000 securities worldwide and is supported by the ICB database, so ICB system can offer the users plenty information outperforming other classification schemes. Secondly, the lower inter-sector correlation as ICB has proved the precise grouping principle and decreased the effect of bias in scheme applying. Thirdly, global industrial landscape can be easily observed by ICB system.

Nowadays, ICB is widely used by NASDAQ, NYSE and other financial departments around the world for stock selection, data analysis and performance measurement. It also will be utilized to drive a search engine.

Table 1 Industry Classification Benchmark Structure

This table is going to provide more details about ICB in industries classifying.

Industry	Supersector	Sector	Subsector
0001 Oil & Gas	0500 Oil & Gas	0530 Oil & Gas Producers	0533 Exploration & Production
			0537 Integrated Oil & Gas
		0570 Oil Equipment, Services & Distribution	0573 Oil Equipment & Services
			0577 Pipeline
		0580 Alternative Energy	0583 Renewable Energy Equipment
			0587 Alternative Fuels
1000 Basic Materials	1300 Chemicals	1350 Chemicals	1353 Commodity Chemicals
			1357 Specialty Chemicals
	1700 Basic Resources	1730 Forestry & Paper	1733 Forestry
			1737 Paper
		1750 Industrial Metals & Mining	1753 Aluminum
			1755 Nonferrous Metals
			1757 Iron & Steel
		1770 Mining	1771 Coal
			1773 diamonds & Gemstones
			1775 General Mining
			1777 Gold Mining
			1779 Platinum & Precious Metals
2000 Industrials	2300 Construction & Materials	2350 Construction & Materials	2353 Building Materials & Fixtures
			2357 Heavy Construction
	2700 Industrial Goods & Services	2710 Aerospace &Defenses	2713 Aerospace

Industry	Supersector	Sector	Subsector
			2717 Defense
		2720 General Industrials	2723 Containers & packaging
			2727 Diversified Industrials
		2730 Electronic & Electrical Equipment	2733 Electrical Components & Equipment
		2750 Industrial Engineering	2753 Commercial Vehicles& Trucks
			2757 Industrial Machinery
		2770 Industrial Transportation	2771 Delivery Services
			2773 Marine Transportation
			2775 Railroads
			2777 Transportation Services
			2779 Trucking
		2790 Support Services	2791 Business Support Services
			2793 Business Training & Employment Agencies
			2795 Financial Administration
			2797 Industrial Suppliers
			2799 Waste & Disposal Services
3000 Consumer Goods	3300 Automobiles & Parts	3350 Automobiles & Parts	3353 Automobiles
			3355 Auto Parts
			3357 Tires
	3500 Food & Beverage	3530 Beverages	3533 Brewers
			3535 Distillers & Vintners
			3537 Soft Drinks

Industry	Supersector	Sector	Subsector
		3570 Food Producers	3573 Farming & Fishing
			3577 Food Products
	3700 Personal & Household Goods	3720 Household Goods & Home Construction	3722 Durable Household Products
			3724 Nondurable Household Products
			3726 Furnishings
			3728 Home Construction
		3740 Leisure Goods	3743 Consumer Electronics
			3745 Recreational Products
			3747 Toys
		3760 Personal Goods	3763 Clothing & Accessories
			3765 Footwear
			3767 Personal Products
		3780 Tobacco	3765 Tobacco
4000 Health Care	4500 Health Care	4530 Health Care Equipment & Services	4533 Health Care Providers
			4535 Medical Equipment
			4537 Medical Supplies
		4570 Pharmaceutical & Biotechnology	4573 Biotechnology
			4577 Pharmaceuticals
5000 Consumer Services	5300 Retail	5330 Food & Drug Retailers	5333 Drug Retailers
			5337 Food Retailers & Wholesalers
		5370 General Retailers	5371 Apparel Retailers
			5373 Broadline Retailers
			5375 Home Improvement Retailers
			5377 Specialized Consumer Services

Industry	Supersector	Sector	Subsector
			5379 Specialty Retailers
	5500 Media	5550 Media	5553 Broadcasting & Entertainment
			5555 Media Agencies
			5557 Publishing
	5700 Travel & Leisure	5750 Travel & Leisure	5751 Airlines
			5752 Gambling
			5753 Hotels
			5755 Recreational Services
			5757 Restaurants & Bars
			5759 Travel & Tourism
6000 Telecommunications	6500 Telecommunications	6530 Fixed Line telecommunications	6535 Fixed Line Telecommunications
		6570 Mobile Telecommunications	6575 Mobile Telecommunications
7000 Utilities	7500 Utilities	7530 Electricity	7535 Conventional Electricity
			7537 Alternative Electricity
		7570 Gas, Water & Multi-Utilities	7573 Gas Distribution
			7575 Multi-utilities
			7577 Water
8000 Financials	8300 Banks	8350 Banks	8355 Banks
	8500 Insurance	8530 Nonlife Insurance	8532 Full Line Insurance
			8534 Insurance Brokers
			8536 Property & Casualty Insurance
			8538 Reinsurance
		8570 Life Insurance	8575 Life Insurance

Industry	Supersector	Sector	Subsector
	8600 Real Estate	8630 Real Estate Investment & Services	8633 Real Estate Holding Development
			8637 Real Estate Services
		8670 Real Estate Investment Trusts	8671 Industrial & Office REITs
			8672 Retail REITs
			8673 Residential REITs
			8674 Diversified REITs
			8675 Specialty REITs
			8676 Mortgage REITs
			8677 Hotel & Lodging REITs
	8700 Financial Services	8770 Financial Services	8771 Asset Mangers
			8773 Consumer Finance
			8775 Specialty Finance
			8777 Investment Services
			8779 Mortgage Finance
		8980 Equity Investment Instruments	8985 Equity Investment Instruments
		8990 Nonequity Investment Instruments	8995 Nonequity Investment Instruments
9000 Technology	9500 Technology	9530 Software & Computer Services	9533 Computer Services
			9535 Internet
			9537 Software
		9570 Technology Hardware & Equipment	9572 Computer Hardware
			9574 Electronic Office Equipment
			9576 Semiconductors
			9578 Telecommunications Equipment

2.2. Comparison of classification benchmark

Disparity of those alternative classifications for many aspects, such as the scope of application, will suggest researchers to apply them based on their own specific features. To begin with SIC, Guenther and Rosman (1994) indicate that SIC focuses on the industrial production process and product output. However, SIC codes are not capable to compare companies with similar operating function. In comparison, Krishman and Press (2003) demonstrate that NAICS expects to define industries based on production press. More differently, the disparity is that NAICS have an ability to discover and define new industries, such as services industries. As for FF industry classification, it has been utilized in many financial fields, such as asset pricing (Brennan et al. 2004), corporate finance (Graham and Kumar, 2006), financial anomalies (Chan et al. 2004), accounting (Chan, et al, 2004) and economics (Bebchuk and Grunstein, 2005). To end up with GICS, this classification results in more reliable industry groupings in financial analysis and research (Hrazdil and Scott, 2013).

Bhojraj (2003) has done a test to prove some evident about classification's difference in industries partition. It selects some firms into a sample and divides them by two-digit SIC codes. NASIC shows 80% of firms matching with grouping. This is a relatively high percentage of correspondence, while FF has a slightly higher percentage of correspondence, 84%. However, GICS just agrees with SIC codes grouping into 54%. Although NAICS presents a high correspondence with SIC, they still have great differences according to the result of test. Firstly, NAISC has an ability to define new industries as previous mention. Comparing to the SIC, NAICS has software publishers (such as magazine publishers and newspaper companies) as a new category. In 1987, the newly published version of SIC brought software publishers into computer-related services categories (data-processing services, computer-programming services, software reproduction), but not defining it as a new industries in taxonomy. In which case, SIC lays particular emphasis on industries' characteristics, while NAICS focuses on the production process. Table 2 will show more details about their principle of grouping industries. Secondly, NAICS is developed from SIC, it is scientific and keeping up with the market movement, like updating industrial data along with the time changing. Naturally, it definitely pays more attention to emerging industries and un-to-date services. For instant, diet and weight-reducing centers, paging, telemarketing bureaus are not included in SIC codes but NAICS. Thirdly, NAICS provides more financial information than SIC does.

Table 2 Comparison of the structures of SIC and NAICS (Krishnan and Press, 2003):

This table shows the difference in two classifications. For example, SIC defines agriculture, forestry and fishing in the same industry, while NAICS includes hunting in it as well. Due to NAICS is well developed from SIC codes, it possesses finer partition with showing more details to investors, namely that it contains more subsection comparing to SIC.

SIC divisions	**NAICS sectors**
A. Agriculture, forestry and fishing	11. Agriculture, forestry, fishing and hunting
B. Mining	21. Mining
C. Construction	22. Utilities
D. Manufacturing	23. Construction
E. Transportation	31-33. Manufacturing
F. Wholesale trade	42. Wholesale trade
G. Retail trade	44-45. Retail trade
H. Finance, insurance and real estate	48-49. Transportation and warehousing
I. Services	51. Information
J. Public administration	52. finance and insurance

K. No classifiable establishments	53. Real estate and rental and leasing
	54. Professional, scientific and technical services
	55. Management of companies and enterprises
	56. Administrative and support; waste management; and remediation services
	61. Educational services
	62. Health care and social assistance
	71. Arts, entertainment and recreation
	72. Accommodation and food services
	81. Other services (except public administration)
	92. Public administration
	99. Unclassified establishments

Furthermore, different classification schemes react differently to financial ratio. Basically, financial ratios are the indicators to provide some information to investors and researchers without any bias. Besides, they faithfully reflect the states about firms' circumstance, such as profitability, management state and return of equity. Following Bhojraj et al (2003), it redefines which classification benchmarks are preferable to interpret the financial ratios depending on homogeneous categories. The result shows that GICS has better performance in valuation multiples, as it can explain more proportion of the variation in firm-level than alternative classifications. Those financial ratios, such as the variance in firm-level price-to-book, enterprise value-to-sales, and price-to-earnings ratio are explained better by GICS also, as the adjusted R^2generated by GICS is 10% to 30% higher than the other classification schemes do. As we have mention above, GICS works well in financial analysis and researches, this hypothesis is demonstrated here. Moreover, different from alternative taxonomies, GICS seems have the other path to group the market, likewise, it definitely owns its strong advantage in explaining co movements of stock return, interpreting the cross-sectional variations in valuation multiples, forecasting the growth rate, developing expenditures and other various key financial aspects. Specifically, in forecasting the growth rate of firm, prior researchers use GICS in cost of capital investigation (Gebhardt, Lee, and swaminathan, 2001; Claus and Thomas, 2001), growth rate in equity estimation (Frankel and Lee, 1998; Lee, Myers and Swaminathan, 1999), market efficiency valuation (Laporta, 1996), and identification of homologous firms (Bhojraj and Lee, 2002). Moreover, for another example, the average forecasted growth for each industry can explain a greater proportion of the firm-level variations. What's more, GICS does not always generate better results in all the financial ratios. For example, for debt-book equity, SIC and NAICS have better performance in evaluating it than others, perhaps due to SIC and NAICS emphasize on production.

In brief, all the previous researches have been done in order to investigate and group firms into finer partition, likewise, to secure their special application based on their own properties. Due to classification scheme's disparity in launching and limitation of applying, academic researchers are ought to select an appropriate classification benchmark to control samples and be applied for tests.

3. Data

3.1.Data definition and database

To measure the relationship between industries and the aggregated market, we select monthly data from Datastream, in order to assure data resource's reliability and veracity. Data's selection is far from easy. The examining period for data starts from January 1973 and end up with February 2013, adding up 40 years. Large sample can decrease or even eliminate the bias misleading our judgment. In our test, all the industries data is ought to be defined by ICB. Specifically, data representing the market is selected from two resources, one is S&P 500 and the other is Dow Jones Industrials. In terms of data's numbers and covering ranges, S&P 500 covers more industries than Dow Jones Industrials, such as retails, manufactory and resources, while Dow Jones industrials only focuses on industry, such as chemistry and engineering. Above arguments support us to select the more appropriate one, S&P 500, as market data.

However, one important problem is that ICB partitions the aggregated market into 41 industries, only 32 industries data is gain from above mentioned database. It means that there are still 9 subsequent observations missing, Alternative Energy, Household Goods & Home Construction, Leisure Goods, Personal Goods, Tobacco, Health Care Equipment & Services, Real Estate Investment Trusts, Financial Services, and Non-equity Investment Instruments. In this case, we have to drop those industries from investigating portfolio and contribute the model based on only 32 industrial data. What's more, not all the data in our sample begins with the same date, some of them only have been recorded in the past decade. For example, the data of Equity Investment Instruments and Software do not start from January 1973 as other data does, but from August, 1997 and from December 1981 respectively, because of the lack of records about them in the past. Nevertheless, this is understandable because both of them are technical industries which are only developed from the past two decades. What we are considering is that whether the different estimating period can mislead our final result or not. The answer is definitely no, according to Hong et al (2006), the lack of time series data for investing the relation between aggregated market and industries do not have much effect on the final results.

From the table 3 below, some industries relate with others closely and they are self-explained. For example, for those two industries, OILGPUS (oil & gas production) and OILESUS (oil equipment, services &distribution) belong to oil industry together, however, each branch focuses on the disparity areas, one pays more attention in production while the other lays more emphasize on the ancillary foundation, such as service and manufacturing equipment. For another example, INDENUS (industrial engineering) and INDTRUS (industrial transportation) have something in common that they are working as two proxies of industry contracture. Those above industries are similar but still have disparity.

3.2.Predictive Variables

The market is not independent. However, it is influenced by various variables. Following the Hou (2007), it denotes that the information diffusion is capable to lead to lead-lag effect of market. Therefore, not only the data from industries, but also the information variables which may correlate or affect the final result should be considered as well, in order to control the bias in the sample.

We select the following representative variables from a long candidate list. To do so, Dividend Yield (DY) represents a dividend per share and reflects the earning on investment for firms. Price – earnings Ratio (PER) is defined as market share price and is the most important indicator to evaluate firms' current value. Market Volatility indicates the fluctuation of market movement. Based on this character, many investors view it as a sensitive indicator to evaluate market's risk. As foe Consumer Price Index (CPI) is working as a proxy to represent the change of market inflation. Treasury Bills is viewed as the risk-free rate. Default Spread is the difference between BAA-rated and AAA-rated bonds. We collect the data of those variables starting from January 1973 to February 2013, the same period as industrial data has. However, Dividend Yield and PER miss observations since October 2012, while data of Market Volatility just were recorded from January 1990. The specific details are shown in the below table 3.What's more, for the data analysis process, we generate monthly return of each industry, as well as the market data and information variables.

Table 3

It is the summary of the data, including thirsty-two industries listing in Panel A and eight predictive variables are posting in Panel B. The testing period is both from January 1973 to February 2013. The summary describes data from six aspects, such as mean, maximum and minimum figures, standard deviation, observation and the definition of industries, in order to present a whole picture of the data we use in this test.

Industry	Mean	Max	Min	Standard deviation	Observation	Definition
Panel A						
AERSPUS	0.008	0.212	-0.308	0.062881	481	Aerospace and Defense
AUPRTUS	0.003	0.318	-0.404	0.065815	481	Automobiles, Auto Parts and Tires
BANKUS	0.004	0.204	-0.408	0.065162	481	Banks
BEVESUS	0.007	0.226	-0.317	0.055544	481	Brewers, Distillers & Vintners, Soft Drinks
CHMCLUS	0.006	0.194	-0.316	0.060332	481	Commodity Chemicals and Specialty Chemicals
CNSTMUS	0.006	0.286	-0.352	0.070732	481	Building Materials & Fixtures and Heavy Construction
ELECTUS	0.002	0.195	-0.161	0.044212	481	Conventional Electricity and Alternative Electricity
ELTNCUS	0.008	0.236	-0.29	0.065464	481	Electrical Components & Equipment and Electronic Equipment
EQINVUS	2E-04	0.337	-0.77	0.12235	186	Equity Investment Instruments
FDRGRUS	0.008	0.272	-0.249	0.053371	481	Drug Retailers and Food Retailers & Wholesalers

Industry Classification Benchmark

Industry	Mean	Max	Min	Standard deviation	Observation	Definition
FOODSUS	0.008	0.2	-0.196	0.045508	481	Farming & Fishing and Food Products
FSTPAUS	0.004	0.587	-0.472	0.084736	481	Forestry and Paper
GNINDUS	0.004	0.187	-0.311	0.059205	481	Containers & Packaging and Diversified Industrials
GNRETUS	0.007	0.22	-0.359	0.062905	481	Apparel Retailers, Broadline Retailers, Home Improvement Retailers, Specialized Consumer Services and Specialty Retailers
GWMUTUS	0.004	0.213	-0.196	0.051961	481	Gas Distribution, Multi-utilities and Water
INDENUS	0.007	0.23	-0.373	0.067066	481	Commercial Vehicles & Trucks and Industrial Machinery
INDMTUS	0.005	0.323	-0.437	0.088648	481	Aluminum, Nonferrous Metals and Iron & Steel
INDTRUS	0.007	0.223	-0.358	0.062176	481	Delivery Services, Marine Transportation, Railroads, Transportation Services and Trucking
LFINSUS	0.007	0.33	-0.566	0.074343	481	Life Insurance
MEDIAUS	0.005	0.168	-0.281	0.057356	481	Broadcasting & Entertainment, Media Agencies and Publishing
MNINGUS	0.003	0.368	-0.817	0.104118	481	Coal, Diamonds & Gemstones, General Mining, Gold Mining and Platinum & Precious Metals
NLINSUS	0.007	0.291	-0.181	0.055199	481	Full Line Insurance, Insurance Brokers, Property & Casualty Insurance,

Industry	Mean	Max	Min	Standard deviation	Observation	Definition
						Reinsurance and Life Insurance
OILESUS	0.006	0.304	-0.329	0.076321	481	Oil Equipment & Services and Pipelines
OILGPUS	0.006	0.194	-0.194	0.055117	481	Exploration & Production and Integrated Oil & Gas
PHARMUS	0.007	0.282	-0.191	0.053348	481	Biotechnology and Pharmaceuticals
RLESTUS	0.006	0.336	-0.424	0.078101	481	Real Estate Holding & Development and Real Estate Services
SOFTWUS	0.018	0.346	-0.28	0.093479	374	Computer Services, Internet and Software
SUPSVUS	0.006	0.155	-0.288	0.05933	481	Business Support Services, Business Training & Employment Agencies, Financial Administration, Industrial Suppliers and Waste & Disposal Services
TECHDUS	0.005	0.203	-0.383	0.075305	481	Computer Hardware and Electronic Office Equipment, Semiconductors and Telecommunications Equipment
TELFLUS	0.003	0.258	-0.194	0.05121	481	Fixed Line Telecommunications
TELMBUS	0.005	0.41	-0.435	0.08702	481	Mobile Telecommunication
TRLESUS	0.008	0.307	-0.431	0.074208	481	Airlines, Gambling, hotels, Recreational Services, Restaurants & Bars and Travel & Tourism

Industry	Mean	Max	Min	Standard deviation	Observation	Definition
S&P 500	0.005	0.151	-0.245	0.045606	481	Benchmark
Panel B						
Predictive Variables						
DY	3.083	6.38	1.08	1.351047	476	Dividend Yield
PER	19.76	142.2	6.69	15.91826	476	Price-earnings Ratio
Volatility	20.4	59.89	10.42	7.751869	278	Market Volatility
CPI	139.5	232.8	43	54.15904	481	Consumer Price Index
Treasury Bills	5.264	15.52	0.01	3.318396	481	Risk-free rate
DS	1.117	3.37	0.54	0.474606	481	Default Spread, the difference between AAA and BAA

4. Methodology

In this section, we investigate the predictive ability of industries, because this propensity of each industry is going to be more and more important for economic activities and guide the investing direction, as well as the industries' developing. The subsequent two methods are going to be tested our hypothesis that industries portfolio is able to forecast the market's movement. What's more, we are expecting to gain the significant results to figure out how this mechanism processes and lead to the industrially predictive ability and if there is any disparity among different industries.

4.1 Predictive regressions by using information variable

4.1.1. To test the predictive ability by using the following equation

$$\mathbf{M_t = \alpha_1 V_1 + \alpha_2 V_2 + \ldots + \alpha_t V_t + \beta\, I_t + \varepsilon_t} \qquad (1)$$

According to the equation above, M_t indicates market portfolio's data. Here, we use S&P 500 to represent the whole market, because it is accepted widely and covers complete relatively industries comparing to other index, such as the financial industry, public transportation and manufactory. V_t represents the predictive variables respectively, in this test, we use subsequent six variables, Treasury Bills, CPI (consumer price index), Default Spread, Market Volatility, Dividend Yield and PER (price-earnings ratio); and I_t indicates the numbers of lags used in the test, in order to find out if the industries cause a lead-lag effect to the aggregated market. All above figures are used as monthly return instead the original data.

4.1.2. Information variables

How important are the information variables? An increasing numbers of information variables are created by researchers aiming at delivering the more comprehensive information to investors, such as generally economic circumstance, earnings of firms or any other private messages. For example, PMI (purchasing managers index) is widely used into evaluating socio-economic development status. During the global crisis period, many reports and researches mention PMI a lot and view it as a critical indicator to reflect circumstance about economic recession. More specifically, when those countries or economic entities suffer the worse time, the figure of this index will be lower than 50. The lower the figures are, the worse the recession is, vise verse. Moreover, PMI is conductive to stock market's fluctuation immediately, especially stocks in emerging markets. Obviously, reacting sensitively to this index indirectly explains that the market attaches the great importance to this index.

For one more example, CPI denotes the level of inflation or deflation in a specific period of the society. Some relative researchers find out that "the inflation pressure has a negative but not significant effect on the performance of the company" (Shao, 2011). Therefore, inflation or deflation has an impact on companies directly, even though it is not a significant influence as that research says, however, negative influence to the final result is ought to be eliminated or cut down its impact at least. All in all, we select those information variables in order to consider the impact of variables and then to assure the precise and reliable result. However, we cannot predicate that those six variables are capable to capture the predictive power to the aggregated market, so a conclusion should be based on the final result.

About the data selection, six following variables are taken into our test, Treasury Bills, CPI (consumer price index), Default Spread (BAA rate and AAA rate), Dividend Yield, Market Volatility and PER (price-earnings ratio).Given that the available observations of them in investigating are only 273, this testing period starts from 31st January, 1990 to 28th September, 2012. Missing observations are caused by the lack of recording data for Market Volatility before January, 1990, while observations of other two variables Dividend Yield and price-earnings ratio are missing since September, 2012. Nevertheless, this data deficiency problem may be conductive to generate a larger standard error in the final result and affect the accuracy of our test, to do so, all the variables are ought to possess in the same testing time interval with the same numbers of observation to eliminate the bias.

4.1.3. Result Analysis

The results are presented in the subsequent table 4.According to the above mentioned equation, the elements we estimate are six information variables, industry data and market data. The market data is pulled in the left side of equation as a dependent variable, while the rest variables and industrial data are pulled in the right side as independent variables. To do so, the mechanism is that three regressions are established for each industry. To begin with the first regression, it contains industry data and Treasury Bills, namely this regression only has one information variable and the industrial data with one month lag. Continually, three more variables (CPI, Default Spread and Dividend Yield) are pulled into the second regression. Therefore, there are four information variables in total included into this regression equation based on the first one. Last but not least, the final regression considers all those six information variables in the end. The premise of doing this is that pulling the different numbers of industrial information variables into different regressions will not cause any interaction among all the regressions, because those three regressions run separately.

We expect to gain the final reason with subsequent properties. Firstly, an increasing R^2 accords with our test. Each regression will generate one R^2 respectively in the end. It means that we will capture three R^2 in total with different numbers of information variables for each industry. The higher R^2 is generated, the better linear regression is gained. In addition, for the statistical aspect, the standard error of regression is used to assure the precision of predicting the relationship between industry and market. Secondly, non-zero coefficients are in our expectation, because it is able to construct the final result and prove that they have predictive power. In other words, the information diffusion hypothesis will be demonstrated based on non-zero coefficients.

For specific results analyzing processes, industry OILGPUS (oil & gas producers) is presented as an example, then these analyzing steps and theories can still be applied to the rest industries. To start with OILGPUS (oil & gas producers), the value of R^2 in the first regression is the lowest one, 0.0033, and it is increasing in the second and third regressions, they are 0.0144 and 0.2355 respectively. Rising R^2 is an expectant consequence and accords with our initial hypothesis. Besides, we run the last regression and find out that coefficient CPI is in the 10% significant interval, while coefficients of Default Spread, Dividend Yield, volatility and price-earnings ratios are in the 5% significant interval. It denotes that those mentioned coefficients are not zero and they have contribution to forecast the market portfolio. Go so far as to the rest 31 industries, the similar results, increasing R^2, has been generated. Comparing the R^2 of the third regression of each industry, some are higher than 0.2, such as OILGPUS (oil & gas production), OILESUS (oil equipment, service & distribution) and CHMCLUS (chemicals), while some are lower than 0.03, like MNINGUS (mining) and AERSPUS (aerospace & defense). These results reflect the stronger and weaker linear regressions generated by the different industries. What’s more, in terms of significant of coefficients, to be surprise, Treasury Bills are not statistically significant in all the 32 industries, while CPI is only significant in three of thirty-two industries, OILGPUS (oil & gas producers), OILESUS (oil equipment, services & distribution) and CHMCLUS (chemicals). Generally, all above evident hardly support an argument that Treasury Bills have a predictive ability, nevertheless, only does CPI show this power in few industries.

All in all, a conclusion should be made that each industry has its predictive ability to the market by using those information variables, Default Spread (BAA rate and AAA rate), Dividend Yield, volatility and PER.

Table 4 Predictive ability of information variables

Based on the ICB, this table contains all the results of testing the predictive power of industries with information variables. Treasury Bills, CPI (consumer price index), Default Spread (BAA rate and AAA rate), Dividend Yield, Market Volatility and PER (price-earnings ratio) are six information variables for our investigation. There are three linear regressions being generated for each industry via using different numbers of information variables. The probability and R^2 are the results being focused on and analyzed. Also, the heteroskedasticity may happen in time series data, white test is going to be used to control it.

	OILGPUS			OILESUS			CHMCLUS			FSTPAUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0015	0.0067	0.1118	0.0015	0.0066	0.1090	0.0013	0.0058	0.1128	0.0014	0.0056	-0.0334
probability	0.7897	0.8442	0.0008	0.7850	0.8476	0.0013	0.8184	0.8636	0.0008	0.8005	0.8707	0.4069
Industry(-1)	-0.0069	-0.0097	-0.1201	0.0407	0.0383	-0.0458	0.0280	0.0233	-0.1058	0.0128	0.0104	0.0228
	0.9107	0.8733	0.0239	0.3984	0.4258	0.2831	0.6775	0.7301	0.0508	0.7322	0.7856	0.5747
Treasury Bills(-1)	0.0012	-0.0003	0.0004	0.0011	-0.0004	0.0005	0.0012	-0.0003	0.0003	0.0012	-0.0003	0.0002
	0.3878	0.8298	0.7608	0.4215	0.8136	0.7612	0.3923	0.8506	0.8244	0.3912	0.8565	0.9061
CPI(-1)		0.0000	-0.0002		0.0000	-0.0002		0.0000	-0.0002		0.0000	0.0001
		0.9266	0.0971		0.8797	0.0998		0.9206	0.0841		0.9318	0.5845
Default spread(-1)		-0.0105	0.0290		-0.0092	0.0292		-0.0099	0.0295		-0.0101	-0.0246
		0.3599	0.0259		0.4171	0.0277		0.3842	0.0287		0.3733	0.0920
Dividend yield(-1)		0.0058	-0.0160		0.0059	-0.0155		0.0059	-0.0157		0.0060	0.0116
		0.2978	0.0028		0.2918	0.0044		0.2919	0.0039		0.2853	0.0776
Volatility(-1)			-0.0035			-0.0035			-0.0036			0.0009
			0.0000			0.0000			0.0000			0.1777
PER(-1)			0.0004			0.0004			0.0004			0.0002
			0.0001			0.0003			0.0000			0.2303
R square	0.0033	0.0144	0.2355	0.0085	0.0188	0.2225	0.0046	0.0152	0.2338	0.0040	0.0147	0.0311
Observation	273	273	273	273	273	273	273	273	273	273	273	273

	INDMTUS			MNINGUS			CNSTMUS			AERSPUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0015	0.0060	-0.0353	0.0017	0.0085	-0.0270	0.0014	0.0054	-0.0404	0.0015	0.0071	-0.0318
probability	0.7881	0.8584	0.3566	0.7700	0.8005	0.4836	0.8091	0.8731	0.3051	0.7891	0.8364	0.4208
Industry(-1)	0.0076	0.0071	0.0263	-0.0171	-0.0196	-0.0124	0.0151	0.0131	0.0491	-0.0084	-0.0124	0.0173
	0.8684	0.8730	0.5435	0.6728	0.6174	0.7450	0.7781	0.7989	0.3301	0.8891	0.8375	0.7726
Treasury Bills(-1)	0.0011	-0.0003	0.0001	0.0011	-0.0005	0.0001	0.0012	-0.0003	0.0003	0.0012	-0.0003	0.0002
	0.3904	0.8340	0.9324	0.4257	0.7605	0.9449	0.3920	0.8525	0.8613	0.3841	0.8269	0.9230
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001
		0.9276	0.5470		0.8941	0.6434		0.9353	0.5057		0.9176	0.5974
Default spread(-1)		-0.0103	-0.0256		-0.0107	-0.0281		-0.0102	-0.0266		-0.0106	-0.0242
		0.3692	0.0776		0.3509	0.1232		0.3692	0.0717		0.3609	0.0996
Dividend yield(-1)		0.0060	0.0120		0.0058	0.0107		0.0060	0.0126		0.0058	0.0113
		0.2838	0.0608		0.2993	0.1007		0.2824	0.0530		0.2958	0.0801
Volatility(-1)			0.0009			0.0007			0.0010			0.0008
			0.1143			0.2638			0.1024			0.1877
PER(-1)			0.0002			0.0002			0.0002			0.0002
			0.1856			0.1290			0.1761			0.1680
R square	0.0035	0.0145	0.0317	0.0048	0.0162	0.0299	0.0038	0.0147	0.0336	0.0034	0.0145	0.0296
Observation	273	273	273	273	273	273	273	273	273	273	273	273

Table 4 Predictive ability of information variables (Continued)

	GNINDUS			ELTNCUS			INDENUS			INDTRUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0016	0.0038	-0.0395	0.0014	0.0034	-0.0418	0.0011	0.0015	-0.0468	0.0013	0.0054	-0.0365
probability	0.7641	0.9103	0.3090	0.8002	0.9200	0.2833	0.8533	0.9646	0.2155	0.8194	0.8737	0.3500
Industry(-1)	0.0797	0.0733	0.1017	0.0430	0.0414	0.0707	0.0535	0.0519	0.0867	0.0339	0.0291	0.0595
	0.2076	0.2392	0.0909	0.4146	0.4233	0.1605	0.3550	0.3593	0.1113	0.5837	0.6284	0.2948
Treasury Bills(-1)	0.0011	-0.0002	0.0002	0.0011	-0.0003	0.0001	0.0012	-0.0002	0.0003	0.0012	-0.0003	0.0002
	0.4197	0.9009	0.9052	0.4191	0.8492	0.9411	0.3906	0.9169	0.8569	0.3930	0.8557	0.8983
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001
		0.9255	0.5361		0.9591	0.4878		0.9860	0.4307		0.9264	0.5567
Default spread(-1)		-0.0083	-0.0253		-0.0096	-0.0268		-0.0094	-0.0279		-0.0098	-0.0254
		0.4448	0.0696		0.3913	0.0620		0.3914	0.0476		0.3802	0.0759
Dividend yield(-1)		0.0058	0.0122		0.0062	0.0129		0.0064	0.0135		0.0059	0.0120
		0.2957	0.0578		0.2623	0.0448		0.2524	0.0328		0.2878	0.0622
Volatility(-1)			0.0011			0.0011			0.0012			0.0010
			0.0628			0.0728			0.0397			0.1139
PER(-1)			0.0001			0.0002			0.0001			0.0002
			0.3169			0.2319			0.2594			0.1860
R square	0.0146	0.0235	0.0449	0.0077	0.0183	0.0392	0.0102	0.0206	0.0437	0.0051	0.0156	0.0340
Observation	273	273	273	273	273	273	273	273	273	273	273	273

	SUPSVUS			AUPRTUS			BEVESUS			FOODSUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0012	0.0054	-0.0398	0.0013	0.0044	-0.0401	0.0015	0.0067	-0.0305	0.0015	0.0065	-0.0310
probability	0.8338	0.8744	0.3069	0.8197	0.8983	0.3146	0.7926	0.8447	0.4477	0.7939	0.8504	0.4359
Industry(-1)	0.0562	0.0460	0.0893	0.0342	0.0296	0.0596	0.0166	0.0109	0.0309	0.0195	0.0137	0.0356
	0.3278	0.4167	0.0906	0.4969	0.5426	0.2262	0.7815	0.8533	0.6013	0.7833	0.8478	0.6139
Treasury Bills(-1)	0.0012	-0.0002	0.0004	0.0012	-0.0002	0.0003	0.0011	-0.0004	0.0001	0.0011	-0.0003	0.0001
	0.3799	0.9000	0.8273	0.3794	0.9030	0.8400	0.4090	0.8230	0.9527	0.4034	0.8269	0.9359
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001
		0.9250	0.5209		0.9485	0.5088		0.9169	0.6168		0.9196	0.6114
Default spread(-1)		-0.0094	-0.0265		-0.0098	-0.0265		-0.0103	-0.0242		-0.0102	-0.0239
		0.4067	0.0675		0.3790	0.0623		0.3684	0.1049		0.3742	0.1043
Dividend yield(-1)		0.0056	0.0119		0.0060	0.0123		0.0058	0.0111		0.0059	0.0112
		0.3174	0.0634		0.2864	0.0562		0.2950	0.0934		0.2943	0.0896
Volatility(-1)			0.0011			0.0010			0.0008			0.0008
			0.0799			0.1038			0.2080			0.1961
PER(-1)			0.0002			0.0002			0.0002			0.0002
			0.1967			0.2252			0.1752			0.1669
R square	0.0078	0.0172	0.0385	0.0062	0.0164	0.0362	0.0036	0.0144	0.0303	0.0036	0.0144	0.0302
Observation	273	273	273	273	273	273	273	273	273	273	273	273

	PHARMUS			FDRGRUS			GNRETUS			MEDIAUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0016	0.0078	-0.0295	0.0015	0.0071	-0.0312	0.0010	0.0014	-0.0418	0.0012	0.0039	-0.0387
probability	0.7665	0.8197	0.4554	0.7885	0.8361	0.4263	0.8565	0.9680	0.2905	0.8271	0.9076	0.3167
Industry(-1)	0.0572	0.0554	0.0664	0.0343	0.0312	0.0543	0.0652	0.0637	0.0866	0.0433	0.0392	0.0687
	0.3392	0.3627	0.2509	0.5980	0.6319	0.3732	0.2196	0.2243	0.0939	0.5028	0.5295	0.2524
Treasury Bills(-1)	0.0010	-0.0005	-0.0001	0.0011	-0.0004	0.0000	0.0011	-0.0002	0.0003	0.0012	-0.0002	0.0003
	0.4682	0.7409	0.9496	0.4142	0.7961	0.9951	0.4000	0.9029	0.8481	0.3902	0.8880	0.8707
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001
		0.8837	0.6411		0.8955	0.6278		0.9836	0.4643		0.9468	0.5262
Default spread(-1)		-0.0099	-0.0241		-0.0099	-0.0245		-0.0101	-0.0263		-0.0096	-0.0254
		0.3872	0.1019		0.3855	0.0908		0.3645	0.0682		0.3937	0.0791
Dividend yield(-1)		0.0058	0.0112		0.0058	0.0114		0.0063	0.0125		0.0061	0.0123
		0.2991	0.0859		0.2967	0.0765		0.2571	0.0526		0.2706	0.0544
Volatility(-1)			0.0008			0.0009			0.0010			0.0010
			0.1731			0.1459			0.1228			0.1080
PER(-1)			0.0002			0.0002			0.0002			0.0002
			0.1842			0.1725			0.1787			0.2255
R square	0.0072	0.0179	0.0343	0.0047	0.0154	0.0324	0.0101	0.0208	0.0403	0.0062	0.0166	0.0353
Observation	273	273	273	273	273	273	273	273	273	273	273	273

	TRLESUS			TELFLUS			TELMBUS			ELECTUS		
	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0013	0.0053	-0.0367	0.0015	0.0072	-0.0324	0.0012	-0.0005	-0.0461	0.0016	0.0078	-0.0305
probability	0.8159	0.8762	0.3513	0.7901	0.8335	0.4088	0.8330	0.9873	0.2387	0.7670	0.8196	0.4370
Industry(-1)	0.0192	0.0139	0.0387	0.0507	0.0453	0.0642	0.0471	0.0451	0.0586	0.0787	0.0756	0.0929
	0.7005	0.7741	0.4256	0.3710	0.4216	0.2316	0.1957	0.2130	0.1043	0.3713	0.3860	0.2729
Treasury Bills(-1)	0.0012	-0.0003	0.0003	0.0011	-0.0003	0.0002	0.0012	0.0000	0.0006	0.0011	-0.0004	0.0000
	0.3896	0.8624	0.8543	0.3974	0.8377	0.9120	0.3756	0.9992	0.7141	0.4355	0.7829	0.9947
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001
		0.9386	0.5405		0.8912	0.6094		0.9742	0.4315		0.8507	0.6681
Default spread(-1)		-0.0101	-0.0250		-0.0096	-0.0245		-0.0089	-0.0253		-0.0091	-0.0237
		0.3789	0.0898		0.3912	0.0912		0.4240	0.0785		0.4237	0.1057
Dividend yield(-1)		0.0059	0.0118		0.0056	0.0113		0.0064	0.0128		0.0059	0.0114
		0.2857	0.0655		0.3173	0.0818		0.2465	0.0457		0.2945	0.0770
Volatility(-1)			0.0009			0.0009			0.0010			0.0009
			0.1558			0.1453			0.1105			0.1336
PER(-1)			0.0002			0.0002			0.0002			0.0002
			0.1959			0.1494			0.1735			0.1846
R square	0.0041	0.0147	0.0320	0.0076	0.0177	0.0357	0.0153	0.0250	0.0463	0.0093	0.0197	0.0371
Observation	273	273	273	273	273	273	273	273	273	273	273	273

Table 4 Predictive ability of information variables (Continued)

	GWMUTUS			BANKSUS			NLINSUS			LFINSUS			RLESTUS			EQINVUS			SOFTWUS			TECHDUS		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
CONST	0.0016	0.0078	-0.0294	0.0017	0.0025	-0.0415	0.0017	-0.0009	-0.0490	0.0016	0.0038	-0.0415	0.0014	0.0060	-0.0367	0.0027	0.0703	0.0231	0.0015	0.0050	-0.0341	0.0013	0.0041	-0.0374
probability	0.7734	0.8217	0.4608	0.7626	0.9427	0.3297	0.7562	0.9787	0.2106	0.7780	0.9118	0.3149	0.8096	0.8626	0.3775	0.6491	0.1126	0.6928	0.7929	0.8840	0.3895	0.8154	0.9040	0.3387
Industry(-1)	0.0457	0.0388	0.0513	0.0366	0.0325	0.0564	0.1221	0.1187	0.1588	0.0253	0.0239	0.0512	0.0134	0.0052	0.0340	0.0194	0.0297	0.0346	0.0324	0.0275	0.0380	0.0441	0.0403	0.0550
	0.4694	0.5406	0.3894	0.4630	0.5124	0.3008	0.0700	0.0783	0.0194	0.5828	0.5986	0.3108	0.8009	0.9226	0.5400	0.5831	0.4250	0.3921	0.3752	0.4560	0.3038	0.2625	0.3039	0.1533
Treasury Bills(-1)	0.0011	-0.0004	0.0001	0.0011	-0.0002	0.0002	0.0009	-0.0002	0.0001	0.0011	-0.0003	0.0001	0.0012	-0.0003	0.0003	-0.0002	-0.0012	0.0000	0.0010	-0.0003	0.0001	0.0011	-0.0002	0.0002
	0.4258	0.8068	0.9742	0.4244	0.8817	0.8885	0.4958	0.8833	0.9479	0.4141	0.8495	0.9493	0.3885	0.8462	0.8322	0.9063	0.5736	0.9980	0.4376	0.8392	0.9306	0.4083	0.8785	0.8958
CPI(-1)		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		0.0000	0.0001		-0.0007	-0.0006		0.0000	0.0001		0.0000	0.0001
		0.8638	0.6729		0.9918	0.4887		0.9714	0.4020		0.9623	0.4938		0.9273	0.5605		0.0194	0.0918		0.9572	0.5552		0.9582	0.5251
Default spread(-1)		-0.0092	-0.0232		-0.0096	-0.0258		-0.0084	-0.0277		-0.0099	-0.0269		-0.0102	-0.0244		-0.0282	-0.0375		-0.0100	-0.0247		-0.0096	-0.0256
		0.4265	0.1158		0.3846	0.0747		0.4326	0.0470		0.3782	0.0715		0.3808	0.0989		0.0421	0.0207		0.3800	0.0892		0.3952	0.0765
Dividend yield(-1)		0.0057	0.0111		0.0062	0.0125		0.0064	0.0136		0.0062	0.0129		0.0059	0.0118		0.0584	0.0605		0.0057	0.0112		0.0058	0.0118
		0.3087	0.0911		0.2695	0.0655		0.2493	0.0356		0.2696	0.0571		0.2920	0.0798		0.0062	0.0040		0.3096	0.0844		0.2985	0.0675
Volatility(-1)			0.0008			0.0010			0.0012			0.0011			0.0009			0.0007			0.0009			0.0010
			0.1700			0.1622			0.0477			0.1187			0.1775			0.4205			0.1608			0.1142
PER(-1)			0.0017			0.0002			0.0001			0.0001			0.0002			0.0002			0.0002			0.0002
			0.1847			0.2590			0.3171			0.2943			0.1973			0.1456			0.1814			0.2070
R square	0.0064	0.0164	0.0328	0.0069	0.0171	0.0363	0.0236	0.0330	0.0585	0.0055	0.0162	0.0361	0.0036	0.0143	0.0310	0.0026	0.0467	0.0618	0.0071	0.0170	0.0341	0.0105	0.0203	0.0395
Observation	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273	181	181	181	273	273	273	273	273	273

4.2.The predictive regression between industry and market portfolio

In the previous section, we have investigated if returns of industrial portfolio with information variables can predict the market or not. As for this section, we continue to test how long industries lead the market portfolio ahead. In this case, we test as many months lags as possible to prove how that predictive ability is. The more lags an industry has, the better achievement we gain.

4.2.1. Equation and its description

The dependent variable would be the return of market, while the explanatory variables are the return of industries. The general equation is following:

$$\mathbf{M_t = C + \alpha_1 IND\ (-1) + \alpha_2 IND\ (-2) + \ldots + \alpha_t\, IND(-t) + \varepsilon_t} \qquad (2)$$

Note: M represents the market returns by selecting the S&P 500 as the market return data. IND represents industry, like OILGPUS (oil & gas producers), CHMCLUS (chemicals), FSTPAUS (forestry & paper) and so on. The numbers in bracket indicate the numbers of lags we considers in the test. In this trial, we are going to test 3 months lags, 6 months lags and 12months lags of industries data respectively to check probability of each coefficient is statistically significant or not. To be noticed, as we have verified the predictive ability of information variables in the previous section. Here, it is no need to consider them into this regression again. Otherwise, the equation would be much more complicate. Table 5 contains all the results of this trial. As many statistical research does, to choose 90% and 95% to be our sample’s confidence interval. If the probability of coefficient is out of that confidence interval, the result is statistically significant.

4.2.2. Problem in this method

Due to the data we investigate is time series, the investigating period of data sample is from 1973 to 2013 covering 40 years in total, heteroskedasticity is probable to exist and affect the precise of final result. Moreover, it may cause bias about inconsistent variance and covariance which would lead to an inaccurate t test. Aiming to prevent the test result from being affected by the heteroskedasticity, we should use white test in the data analysis process to control or even eliminate its influence.

4.2.3. Result analysis

From table 5, we run three regressions for each industry in this test, and they include 3 months lags, 6 months lags and 12 months lags respectively. Our expectant result is getting non-zero coefficient, namely the probabilities of them are statistically significant. To begin with the first regression, industrial returns are pulled into the regression. Nearly all the corresponding probabilities with 1 lag are not statistically significant, except SUPSVUS (support services), FDRGRUS (food & drug retailers), GNRETUS (general retailers), TELMBUS (mobile telecommunications), and NLINSUS (nonlife insurance). More specifically, both GNRETUS and TELMBUS are out of 90% confidence interval, while SUPSVUS, FDRGRUS and NLINSUS are eliminated from 95% confidence interval. Continually, the findings in 2 months lags are disappointed, because none industries provide the significant results any more, namely all of them is not possible to predict the market in two months ahead. Nevertheless only five above industries have predictive ability to lead the market in one month, however, in the real market, forecasting ability of one month is already achievable. One month leading to the market portfolio will still have an incomparably significant meaning.

Moreover, the following two regressions groups with six and twelve months respectively are run for investigating the long-horizon predictive ability. The consequence is the same as the previous regression that all the industries are virtually impossible to forecast the market in a long-period. This result is in our anticipation.

In conclusion, only five of thirsty-two industries are capable to predict the aggregated market in one month. However, they are not all the large cap in the market as our original hypothesis, such as nonlife insurance and mobile telecommunications. Following Sadorsky (1999), the authors estimated how the oil price affects the market price, and she found out that the oil price movement could forecast the actual market movement. But in our test, oil industry does not show any predictive ability to the market. Beside, based on the paper Valkanov (2003), as well asTorous, Valkanov and Yan (2004), they came up with a problem that long-horizon regressions hardly generate any better consequences. Hence, in our test result, long-horizon data do not show any forecasting power.

Table 5 Predictive ability of industries with lags

This table presents the results of predictive ability in long horizon. To generate the return of each industry and then to run three linear regressions to test whether three, six and twelve monthly lags with predictive power or not is the original hypothesis. In order to generate comparable and precise results, testing the trial in the same period, launching white test to avoid heteroskedasticity in time series data and controlling linear regressions by R^2, are going to be our core principle.

5.1 Three monthly lags

Code	OILGPUS	OILESUS	CHMCLUS	FSTPAUS	INDMTUS	MNINGUS	CNSTMUS	AERSPUS	GNINDUS	ELTNCUS	INDENUS	INDTRUS	SUPSVUS	AUPRTUS	BEVESUS	FOODSUS	PHARMUS	FDRGRUS	GNRETUS	MEDIAUS	TRLESUS
Lag 1(coef)	-0.0313	0.0121	0.0287	0.0114	-0.0054	-0.0113	0.0346	0.0101	0.0512	0.0309	0.0355	0.0205	0.0972	0.0328	0.0417	0.0311	0.0026	0.0972	0.0706	0.0710	0.0419
Probability	0.504	0.733	0.528	0.712	0.878	0.702	0.363	0.796	0.284	0.468	0.415	0.633	**0.014**	0.405	0.438	0.554	0.957	**0.029**	**0.079**	0.119	0.254
Lag 2	-0.0594	-0.0181	-0.0108	-0.0302	0.0149	-0.0168	-0.0323	-0.0345	-0.0260	-0.0256	-0.0150	-0.0247	-0.0450	-0.0644	-0.0019	0.0160	0.0525	-0.0328	-0.0335	-0.0341	-0.0066
Probability	0.152	0.540	0.785	0.250	0.606	0.477	0.281	0.371	0.521	0.449	0.639	0.504	0.252	0.042	0.965	0.751	0.213	0.445	0.369	0.387	0.823
Lag 3	0.0394	0.0268	0.0026	0.0229	0.0198	-0.0016	0.0204	-0.0089	0.0248	0.0304	0.0064	0.0018	-0.0201	0.0453	-0.0284	0.0101	0.0169	0.0014	-0.0102	0.0712	0.0182
Probability	0.371	0.416	0.945	0.463	0.498	0.953	0.533	0.807	0.553	0.419	0.867	0.958	0.628	0.215	0.560	0.859	0.734	0.974	0.781	0.113	0.546
Constant	0.0059	0.0054	0.0054	0.0055	0.0054	0.0056	0.0054	0.0058	0.0054	0.0053	0.0054	0.0056	0.0053	0.0055	0.0055	0.0051	0.0050	0.0050	0.0053	0.0051	0.0051
Probability	0.013	0.015	0.019	0.010	0.015	0.008	0.017	0.012	0.015	0.024	0.019	0.015	0.017	0.010	0.020	0.035	0.032	0.032	0.019	0.027	0.024
R square	0.009	0.003	0.002	0.005	0.002	0.002	0.005	0.003	0.006	0.005	0.003	0.002	0.018	0.013	0.004	0.001	0.004	0.013	0.011	0.016	0.005
observation	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478	478

Code	TELFLUS	TELMBUS	ELECTUS	GWMUTUS	BANKSUS	NLINSUS	LFINSUS	RLESTUS	EQINVUS	SOFTWUS	TECHDUS
Lag 1(coef)	0.0421	0.0529	0.0935	0.0235	0.0651	0.1335	0.0553	0.0400	0.0214	0.0157	0.0347
Probability	0.369	**0.087**	0.125	0.631	0.112	**0.006**	0.118	0.246	0.543	0.567	0.318
Lag 2	0.0104	-0.0181	-0.0058	0.0100	-0.0502	-0.0661	-0.0616	-0.0261	-0.0217	-0.0118	-0.0091
Probability	0.822	0.455	0.915	0.836	0.173	0.134	0.037	0.378	0.584	0.682	0.761
Lag 3	0.0665	0.0644	0.0861	0.0470	0.0231	0.0377	0.0153	0.0234	0.0450	0.0404	0.0474
Probability	0.181	0.024	0.120	0.332	0.581	0.418	0.682	0.391	0.295	0.146	0.112
Constant	0.0051	0.0051	0.0051	0.0052	0.0054	0.0048	0.0055	0.0053	0.0025	0.0062	0.0052
Probability	0.019	0.018	0.019	0.020	0.014	0.034	0.015	0.016	0.473	0.014	0.018
R square	0.008	0.023	0.014	0.004	0.013	0.029	0.015	0.007	0.019	0.009	0.009
observation	478	478	478	478	478	478	478	478	183	371	478

Table 5 Predictive ability of industries with lags (continually)

5.2 Six monthly lags

Code	OILGPUS	OILESUS	CHMCLUS	FSTPAUS	INDMTUS	MNINGUS	CNSTMUS	AERSPUS	GNINDUS	ELTNCUS	INDENUS	INDTRUS	SUPSVUS	AUPRTUS	BEVESUS	FOODSUS	PHARMUS	FDRGRUS	GNRETUS	MEDIAUS	TRLESUS
Lag 1(coef)	-0.0334	0.0109	0.0268	0.0070	-0.0058	-0.0117	0.0331	0.0116	0.0483	0.0313	0.0310	0.0169	0.0961	0.0308	0.0444	0.0417	0.0022	0.0984	0.0738	0.0720	0.0450
Probability	0.480	0.760	0.551	0.816	0.866	0.687	0.391	0.772	0.310	0.462	0.470	0.699	0.016	0.432	0.414	0.444	0.964	0.029	0.067	0.111	0.226
Lag 2	-0.0572	-0.0195	-0.0116	-0.0294	0.0143	-0.0185	-0.0351	-0.0333	-0.0223	-0.0260	-0.0137	-0.0205	-0.0432	-0.0667	-0.0039	0.0125	0.0526	-0.0311	-0.0363	-0.0398	-0.0100
Probability	0.171	0.512	0.770	0.256	0.621	0.438	0.254	0.389	0.580	0.443	0.665	0.590	0.279	0.043	0.930	0.809	0.217	0.475	0.336	0.312	0.740
Lag 3	0.0407	0.0303	0.0053	0.0269	0.0207	-0.0009	0.0243	-0.0101	0.0265	0.0345	0.0101	0.0023	-0.0211	0.0522	-0.0291	0.0076	0.0168	-0.0003	-0.0097	0.0819	0.0222
Probability	0.362	0.360	0.888	0.386	0.483	0.974	0.459	0.782	0.522	0.361	0.791	0.948	0.623	0.151	0.547	0.892	0.731	0.995	0.789	0.070	0.468
Lag 4	-0.0072	-0.0209	-0.0177	-0.0086	-0.0021	-0.0185	-0.0349	-0.0091	-0.0048	0.0072	-0.0039	-0.0089	-0.0047	0.0055	-0.0175	-0.0473	-0.0126	0.0365	-0.0068	-0.0268	-0.0227
Probability	0.861	0.514	0.625	0.774	0.943	0.443	0.441	0.813	0.908	0.840	0.913	0.817	0.905	0.886	0.711	0.413	0.769	0.385	0.842	0.528	0.502
Lag5	0.0419	0.0001	0.0194	0.0221	0.0077	0.0097	0.0255	0.0213	0.0271	0.0335	0.0089	0.0053	0.0348	0.0372	0.0359	0.0616	0.0225	0.0337	0.0507	0.0225	0.0357
Probability	0.280	0.997	0.589	0.408	0.785	0.659	0.411	0.548	0.469	0.344	0.782	0.880	0.381	0.238	0.348	0.187	0.552	0.394	0.097	0.549	0.213
lag6	-0.0272	-0.0238	-0.0636	-0.0406	-0.0279	-0.0165	-0.0645	-0.0705	-0.0572	-0.0580	-0.0776	-0.0665	-0.0149	-0.0710	-0.0604	-0.0798	-0.0212	-0.0590	-0.0859	-0.0488	-0.0470
Probability	0.475	0.403	0.063	0.076	0.260	0.400	0.031	0.051	0.126	0.078	0.012	0.059	0.710	0.019	0.158	0.110	0.595	0.120	0.014	0.227	0.101
constant	0.0058	0.0057	0.0058	0.0056	0.0055	0.0058	0.0058	0.0063	0.0055	0.0054	0.0058	0.0060	0.0052	0.0056	0.0058	0.0056	0.0051	0.0049	0.0056	0.0052	0.0053
Probability	0.019	0.013	0.015	0.011	0.015	0.007	0.013	0.012	0.016	0.032	0.015	0.014	0.027	0.011	0.022	0.039	0.038	0.054	0.017	0.020	0.026
R square	0.013	0.006	0.010	0.011	0.006	0.006	0.018	0.012	0.013	0.014	0.016	0.010	0.020	0.026	0.011	0.013	0.006	0.022	0.028	0.021	0.014
observation	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475	475

Code	TELFLUS	TELMBUS	ELECTUS	GWMUTUS	BANKSUS	NLINSUS	LFINSUS	RLESTUS	EQINVUS	SOFTWUS	TECHDUS
Lag 1(coef)	0.0400	0.0556	0.1059	0.0302	0.0702	0.1334	0.0523	0.0359	0.0202	0.0168	0.0315
Probability	0.394	0.076	0.079	0.538	0.090	0.007	0.143	0.300	0.556	0.545	0.358
Lag 2	0.0081	-0.0256	-0.0096	0.0100	-0.0562	-0.0654	-0.0547	-0.0276	-0.0139	-0.0109	-0.0130
Probability	0.857	0.310	0.857	0.834	0.126	0.155	0.074	0.343	0.705	0.706	0.669
Lag 3	0.0711	0.0705	0.1010	0.0469	0.0287	0.0427	0.0191	0.0266	0.0488	0.0407	0.0505
Probability	0.152	0.020	0.073	0.322	0.503	0.370	0.602	0.330	0.226	0.141	0.092
Lag 4	0.0432	-0.0116	-0.0001	-0.0104	-0.0245	-0.0215	-0.0106	0.0104	-0.0032	0.0493	0.0248
Probability	0.357	0.690	0.998	0.817	0.571	0.648	0.774	0.747	0.924	0.078	0.392
Lag5	0.0796	0.0287	0.1180	0.0958	0.0396	0.0562	0.0272	-0.0003	-0.0547	0.0250	0.0364
Probability	0.069	0.296	0.031	0.030	0.215	0.179	0.398	0.992	0.031	0.310	0.208
lag6	-0.0127	-0.0069	-0.1069	-0.0679	-0.0580	-0.0454	-0.0644	-0.0576	-0.0164	-0.0183	-0.0280
Probability	0.782	0.800	0.051	0.126	0.074	0.267	0.022	0.026	0.527	0.492	0.362
constant	0.0047	0.0050	0.0050	0.0051	0.0055	0.0049	0.0058	0.0056	0.0020	0.0053	0.0050
Probability	0.037	0.023	0.025	0.029	0.015	0.054	0.016	0.015	0.574	0.059	0.026
R square	0.018	0.026	0.036	0.021	0.022	0.035	0.027	0.017	0.042	0.024	0.017
observation	475	475	475	475	475	475	475	475	180	368	475

Table 5 Predictive ability of industries with lags (continually)

5.3 Twelve monthly lags

Code	OILGPUS	OILESUS	CHMCLUS	FSTPAUS	INDMTUS	MNINGUS	CNSTMUS	AERSPUS	GNINDUS	ELTNCUS	INDENUS	INDTRUS	SUPSVUS	AUPRTUS	BEVESUS	FOODSUS	PHARMUS	FDRGRUS	GNRETUS	MEDIAUS	TRLESUS
Lag 1(coef)	-0.0330	0.0131	0.0301	0.0046	-0.0071	-0.0091	0.0364	0.0236	0.0523	0.0334	0.0287	0.0199	0.0863	0.0237	0.0541	0.0551	0.0043	0.1019	0.0715	0.0634	0.0473
Probability	0.488	0.718	0.518	0.882	0.845	0.746	0.345	0.574	0.276	0.440	0.525	0.654	0.034	0.558	0.326	0.311	0.929	0.033	0.087	0.174	0.212
lag6	-0.0393	-0.0184	-0.0631	-0.0391	-0.0264	-0.0169	-0.0621	-0.0714	-0.0565	-0.0551	-0.0757	-0.0791	-0.0215	-0.0626	-0.0518	-0.0809	-0.0191	-0.0591	-0.0741	-0.0354	-0.0385
Probability	0.304	0.525	0.063	0.102	0.262	0.370	0.045	0.061	0.137	0.105	0.013	0.022	0.599	0.044	0.226	0.107	0.642	0.128	0.036	0.394	0.187
Lag 7	-0.0373	-0.0353	-0.0266	-0.0186	-0.0244	-0.0218	-0.0056	0.0137	-0.0083	-0.0050	-0.0155	-0.0244	0.0335	-0.0395	0.0424	0.0173	0.0264	0.0087	0.0129	0.0166	0.0193
Probability	0.369	0.195	0.377	0.448	0.278	0.264	0.839	0.656	0.808	0.873	0.569	0.431	0.304	0.192	0.216	0.676	0.468	0.799	0.669	0.637	0.466
Lag 8	-0.0088	0.0071	0.0049	0.0069	-0.0445	-0.0385	-0.0128	0.0079	0.0475	-0.0093	-0.0077	-0.0482	-0.0114	0.0278	0.0268	0.0411	0.0526	-0.0026	0.0103	0.0219	0.0256
Probability	0.828	0.795	0.886	0.783	0.056	0.059	0.657	0.820	0.190	0.747	0.790	0.152	0.744	0.382	0.413	0.357	0.195	0.944	0.764	0.544	0.403
Lag 9	-0.0730	-0.0679	-0.0524	-0.0348	-0.0305	-0.0430	-0.0594	-0.0161	-0.0249	-0.0601	-0.0827	-0.0521	-0.0267	-0.0229	-0.0720	-0.1004	-0.0091	0.0074	-0.0107	-0.0149	-0.0116
Probability	0.132	0.046	0.184	0.242	0.230	0.023	0.062	0.640	0.514	0.112	0.016	0.123	0.525	0.477	0.106	0.076	0.838	0.861	0.752	0.730	0.738
Lag 10	-0.0248	-0.0081	0.0268	0.0215	0.0088	0.0009	0.0333	0.0481	0.0191	0.0283	-0.0080	0.0260	0.2740	0.0151	0.0647	0.0420	0.0196	0.0106	0.0656	0.0451	0.0650
Probability	0.531	0.767	0.462	0.420	0.704	0.968	0.259	0.218	0.585	0.377	0.786	0.461	0.274	0.670	0.079	0.430	0.631	0.799	0.070	0.256	0.050
Lag11	-0.0412	-0.0311	-0.0214	-0.0177	-0.0052	-0.0400	-0.0262	-0.0376	0.0188	-0.0115	-0.0099	-0.0566	0.0067	-0.0356	-0.0171	-0.0217	0.0155	-0.0065	-0.0199	0.0307	-0.0141
Probability	0.247	0.220	0.574	0.506	0.820	0.025	0.350	0.282	0.610	0.709	0.739	0.068	0.860	0.278	0.613	0.654	0.667	0.878	0.562	0.428	0.695
lag12	0.0140	0.0121	-0.0023	-0.0094	-0.0065	-0.0292	0.0421	0.0372	0.0207	0.0421	0.0100	0.0478	0.0232	0.0197	0.1085	0.1082	0.0796	0.0433	0.0342	0.0313	0.0492
Probability	0.708	0.668	0.948	0.701	0.775	0.087	0.156	0.305	0.568	0.214	0.735	0.147	0.546	0.540	0.004	0.045	0.039	0.309	0.303	0.384	0.071
constant	0.0074	0.0066	0.0065	0.0062	0.0066	0.0066	0.0062	0.0058	0.0056	0.0061	0.0069	0.0071	0.0049	0.0061	0.0047	0.0054	0.0046	0.0049	0.0052	0.0051	0.0047
Probability	0.001	0.003	0.004	0.004	0.003	0.002	0.006	0.018	0.013	0.009	0.002	0.002	0.034	0.006	0.054	0.024	0.050	0.050	0.023	0.024	0.053
R square	0.016	0.022	0.0174	0.013	0.017	0.032	0.0290	0.019	0.017	0.021	0.033	0.033	0.020	0.019	0.043	0.035	0.015	0.021	0.031	0.018	0.030
observation	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469

Code	TELFLUS	TELMBUS	ELECTUS	GWMUTUS	BANKSUS	NLINSUS	LFINSUS	RLESTUS	EQINVUS	SOFTWUS	TECHDUS
Lag 1(coef)	0.0377	0.0470	0.1118	0.0372	0.0624	0.1263	0.0561	0.0372	0.0195	0.0112	0.0359
Probability	0.422	0.121	0.067	0.474	0.130	0.011	0.134	0.317	0.579	0.698	0.311
lag6	-0.0105	0.0168	-0.0775	-0.0681	-0.0470	-0.0476	-0.0712	-0.0603	0.0001	-0.0236	-0.0241
Probability	0.824	0.563	0.160	0.124	0.150	0.252	0.011	0.027	0.997	0.353	0.435
Lag 7	0.0316	0.0085	0.0409	0.0506	0.0064	0.0216	0.0086	0.0393	0.0243	0.0206	0.0104
Probability	0.468	0.771	0.411	0.203	0.849	0.561	0.749	0.142	0.367	0.470	0.730
Lag 8	0.0304	0.0135	0.0199	-0.0053	0.0192	0.0131	0.0165	-0.0189	-0.0209	-0.0343	-0.0130
Probability	0.493	0.610	0.687	0.905	0.599	0.765	0.532	0.517	0.400	0.184	0.636
Lag 9	-0.0396	-0.0540	-0.1342	-0.1083	-0.0358	-0.0811	-0.0628	-0.0093	-0.0313	0.0054	0.0046
Probability	0.406	0.047	0.024	0.021	0.306	0.042	0.034	0.781	0.180	0.878	0.888
Lag 10	0.0213	0.0555	0.0288	0.0046	-0.0043	0.0049	0.0221	0.0513	-0.0242	0.0300	0.0170
Probability	0.610	0.050	0.574	0.918	0.901	0.887	0.426	0.108	0.333	0.247	0.570
Lag11	-0.0111	0.0114	-0.0842	-0.0001	0.0213	-0.0183	-0.0005	-0.0267	0.0475	0.0189	0.0161
Probability	0.795	0.667	0.081	0.998	0.495	0.598	0.987	0.407	0.041	0.463	0.552
lag12	0.0090	0.0335	0.1193	-0.0026	0.0573	0.0525	0.0132	0.0408	-0.0040	-0.0332	0.0261
Probability	0.838	0.179	0.028	0.949	0.068	0.178	0.640	0.114	0.834	0.264	0.389
constant	0.0056	0.0052	0.0058	0.0063	0.0055	0.0054	0.0060	0.0055	0.0027	0.0067	0.0055
Probability	0.014	0.018	0.009	0.005	0.013	0.024	0.011	0.016	0.450	0.016	0.015
R square	0.007	0.031	0.050	0.025	0.0247	0.040	0.031	0.032	0.035	0.021	0.010
observation	469	469	469	469	469	469	469	469	174	361	469

4.3.Conclusion of Methodology

The original intention of measuring the predictive ability of industries is forecasting the whole market in order to prevent some dreadfully economic crisis, such as subprime mortgage crisis and any other financial storm. Our intention is supported by several previous researches. Following to the Wolfers and Zitzewitz (2004), it indicates that the market context can predict some future events of economy. What's more, the market's predictive ability outperforms the most accurate classification benchmarks.

According to the results of two above methods, information variables show their predictive power coming with industries data, especially the Default Spread (BAA rate and AAA rate), Dividend Yield, Market Volatility and PER (price earnings ratio) are able to lead the aggregated market. However, only several industries possess predictive power in one month ahead. For the long horizon forecasting test, we get null in the final results.

5. Further Research

5.1. Correlation among the industries

Although ICB is famous for its lower inter-sector correlation, in the real market, there is no absolutely clear boundary among industries. For example, OILGPUS (Oil & Gas Production) and OILESUS (Oil Equipment, Services & Distribution) are classified into different industries in the classification, however, they are working as two sub-industries of oil and gas industry, one focuses on the production while the other accounts for the equipment and service. Most of the industries from the same branch tend to correlate with others, because of the globalization resulting in industrial cooperation and compensation.

In this section, we investigate if there is any correlation among industries which belong to the same trunk, if so, do they have power to forecast the market portfolio?

5.1.1. Methodology

The most critical step in methodology is how to find out relative industries and divide them into a group. After that, estimating the correlation among industries would be the best way to evaluate their relationship.

Table 6 An example quoted from ICB

We take a specific group from ICB as an example to explain the mechanism of classifying. There are four gradations into a group as the following table shows.

Industry	Supersector	Sector	Subsector
0001 Oil & Gas	0500 Oil & Gas	0530 Oil & Gas Production	0533 Exploration & Production
			0537 Integrated Oil & Gas
		0570 Oil Equipment, Services & Distribution	0573 Oil Equipment & Services
			0577 Pipelines
		0580 Alternative Energy	0583 Renewable Engergy Equipment
			0587 Alternative Fuels

More specifically, taking Oil & Gas industry as an example, the first column, Industry, is a broad heading, and then it is classified into different Supersectors, Oil & Gas, which are presented in the second column. Continually, Supersectors are divided into several Sectors in the third column, such as Oil & Gas Production and Oil Equipment, Services & Distribution or other alternative energy. Those sectors are the objects in our trial, we call them as industries in the preceding text.

Continually, to test the inter-sector correlation among industries, some supersectors only have one sector, like Chemicals, Construction & Materials, Automobiles & Parts, Health Care, Banks, Real Estate and Financial Services, they will be removed them from the test. Hence, eight supersectors and 21 industries in total are going to be to test and the results are in table 7.

Table 7 Correlation in different sectors

From ICB, we launch the correlating trial to test how close they correlate with each other in the same supersectors and find our other commonness of it.

Resources		Forestry & Paper	Industrial Metals & Mining	Mining
Forestry & Paper	**Person Correlation**	1	.816	.581
	Sig.		.000	.000
Industrial Metals & Mining	**Person Correlation**	.816	1	.818
	Sig.	.000		.000
Mining	**Person Correlation**	.581	.818	1
	Sig.	.000	.000	

Table 7 Correlation in different sectors (continue)

Oil & Gas		Oil & Gas Production	Oil Equipment, Services & Distribution
Oil & Gas Production	**Person Correlation**	1	.972
	Sig.		.000
Oil Equipment, Services & Distribution	**Person Correlation**	.972	1
	Sig.	.000	

Food& Beverage		Beverages	Food Producers
Beverages	**Person Correlation**	1	.982
	Sig.		.000
Food Producers	**Person Correlation**	.982	1
	Sig.	.000	

Retail		Food & Drug Retailers	General Retailers
Food & Drug Retailers	**Person Correlation**	1	.967
	Sig.		.000
General Retailers	**Person Correlation**	.967	1
	Sig.	.000	

Utilities		Electricity	Gas, Water & Multi-utilities
Electricity	**Person Correlation**	1	.884
	Sig.		.000
Gas, Water & Multi-utilities	**Person Correlation**	.884	1
	Sig.	.000	

Insurance		Nonlife Insurance	Life Insurance
Nonlife Insurance	**Person Correlation**	1	.965
	Sig.		.000
Life Insurance	**Person Correlation**	.965	1
	Sig.	.000	

Table 7 Correlation in different sectors (continue)

Telecommunication		Fixed Line Tele	Mobile Tele	Travel & Leisure
Fixed Line Tele	**Person Correlation**	1	.780	.726
	Sig.		.000	.000
Mobile Tele	**Person Correlation**	.780	1	.784
	Sig.	.000		.000
Travel & Leisure	**Person Correlation**	.726	.784	1
	Sig.	.000	.000	

Industrial Goods and Services		Aerospace & Defense	General Industrials	Electronic & Electrical Equipment	Industrial Engineering	Industrial Transportation	Support Services
Aerospace & Defense	**Person Correlation**	1	.756	.956	.965	.984	.831
	Sig.		.000	.000	.000	.000	.000
General Industrials	**Person Correlation**	.765	1	.782	.612	.713	.917
	Sig.	.000		.000	.000	.000	.000
Electronic & Electrical Equipment	**Person Correlation**	.956	.782	1	.931	.942	.782
	Sig.	.000	.000		.000	.000	.000
Industrial Engineering	**Person Correlation**	.965	.612	.931	1	.977	.707
	Sig.	.000	.000	.000		.000	.000
Industrial Transportation	**Person Correlation**	.984	.713	.942	.977	1	.816
	Sig.	.000	.000	.000	.000		.000
Support Services	**Person Correlation**	.831	.917	.782	.707	.816	1
	Sig.	.000	.000	.000	.000	.000	

5.1.2. Data analysis

There are correlations among industries inevitably in the same supersector. According to the above table 6,a conclude is able to be made obviously that the different sectors in each supersector have correlation with each other found on their lower probability of trial and high Pearson Correlation 1. Among the final results, the lowest correlating rate, 0.581, exists in two subsequent industries, Forestry & Paper and Mining which are followed by the other two industries, Fixed Line Tele and Mobile Tele whose correlation rate is 0.78. On the contrary, Beverages and Food Production relates with each other extremely closely, generating the highest industrial correlation rate, 0.982. For the rest of industries, all of them at least have 0.7 in Pearson Correlation indicating that closer and homologous correlation appearing among those relative industries.

To be honest, it is understandable that the sectors in the same supersector belong to the identical field. For example, sectors Forestry & Paper, Industrial Metals & Mining and Mining, they belong to the basic resources sort. Their mainly economic function is supporting construction and industries developing, moreover, all of them is not artificial resource but from the nature. Similar industrial function is conductive to the inter-sector correlation.

Apart from distant and close relationship among industries, their predictive ability investigation is critical analysis to our hypothesis. Comparing the data in table 7 to table 5, most of them have the same or closer predictive ability to the market portfolio. As we mention above, industries possessing the predictive propensity are SUPSVUS (support services), FDRGRUS (food & drug retails), GNRETUS (general retails), TELMBUS (mobile telecommunication) and NLINSUS (nonlife insurance). Two of them, food & drug retails and general retails correlate with each other closely. Why do not the other congeners of other three industries reveal any predictive ability in common? Because food & drug retails and general retails have the highest correlation among them, which is 0.967, while others only are estimated less than 0.9 in correlating rate. Therefore, we are able to conclude that industries with exceedingly closed correlation rate are capable to show the same forecasting propensity to the aggregated market.

5.2. Correlation between industries and market

5.2.1. Induction of the original hypothesis

Different from the previous research about investing the inter-sector correlation among industries, the relation between industries and market portfolio is worth to evaluate. Because each industry is an element to the market and establishes to the market portfolio, therefore they must have some positive or negative relation with this organic system.

Before analyzing the results we gain, we find out some general evident supporting our original hypothesis. For example, the technical industries relates to the market portfolio negatively. More specifically, computers or any other high-technical equipments are not necessary to people’s life in the recession period. They are defined as entertainments in normal life and have many substitutes. It means that when the market is not active or has recession, these industries are possible to suffer sales decreasing from the constrictive retails market at the same period as well, namely its sales volume changes along with the market’s movement. Continually, some data from the 2012 annual report of Cisco is going to be analyzed and support our hypothesis specifically. From the annual report, it shows that its net income went down a lot in the 2009 from 2008, 36,117 million dollars and 39,540 million dollars respectively, approximately 8.7% decreasing, owing to the severe global crisis. In the following three years, the net sales increased slightly, 40,040, 43,218 and 46,061 million dollars respectively, due to the global economy recover slowly from the recession. Above figures demonstrate our opinion that the economic downturn affects the business of technical industries, because they correlate with the market closely and positively. Such a brand leader, Cisco, suffers much impact from the market’s activities, not mention that middle and small size enterprises in the same industries may face much more stress than leaders do, such as reducing purchasing power of costumers and the intensive capital.

5.2.2. Analysis of correlation between industries and market

The results are listed in table 3. Besides, we have to mention that information variables in the regression have no impact to the final results.

To begin with some large cap industries, like OILGPUS (oil & gas producers), OILESUS (oil equipment, services & distribution), MNINGUS (mining), and CHMCLUS (chemicals), their coefficients are negative to the market portfolio. Even so, it is not suspicious because those industries explore and produce the resource as oil and coal which are non-renewable energy for factories' producing, cars' consuming and any other industries developing. Hence, negative correlation is acceptable, due to their necessary and unique developing position. In the paper of Gogineni (2007), author demonstrates that reaction of the aggregated market correlates with the oil price exceedingly negatively, due to oil industries' own dependence in oil requirement is conductive to the oil's irreplaceable and incomparable position in industries' evolution and construction.

On the contrary, some industries count on the market a lot. We are supposed to say that their revenue depends on the market movement. For instance, TELFLUS (fixed line telecommunication) and TECHDUS (technology hardware & equipment) are technical industries and relate to the market positively, because thriving market would lead to a high return in those industries, and vice versa. And a lot of substitutions are available to be replaced them, namely that they are not as unique as oil or chemistry industries do, so such industries depend on the market's movement so much. For one more example, retails industries move following the market in the same direction, because during the recession, less fortune used to be consumed into the unnecessary goods, the sale volume of details is influenced by the decreasing purchasing power heavily.

6. Conclusion

Based on the tests, we could conclude that the return of industries is able to forecast the market portfolio directly, and this predictive ability correlates with information variables which reflects the macro-economy movement and works as economic indicators in many industrial aspects widely. Only five of thirsty-two industries in our sample are possible to predict the market in one month, however, this is already an exceeding achievement. Moreover, homologous industries are provided with similar predictive propensity, due to they have a strong correlation with each other and possess similar characteristics.

References

Amit, R., Livnat, J. 1990. Groupings of conglomerates by their segments' economic attributes: Towards a more meaningful rato analysis. ***Journal of Business Finance and Accounting,*** 17: 85-100.

A Guide to Industry Classification Benchmark (Equity). Version 1.4., December 2012

Estrella,A.,Mishkin, F.S. 2009. Predicting U.S. Recessions: Financial Variables as Leading Indicators. ***The Review of Economics and Statistics***, 80(1) :45-61

Bebchuk, L., Grinstein, Y. 2005. The growth of executive pay. ***Oxford Review of Economic Policy,*** 21: 283-303.

Bhojraj,S., Lee,C., Oler,D.K. 2003. What's my line? A comparison of industry classification schemes for capital market research. ***Journal of Accounting Research,*** 41: 745-769.

Boni, L., Womack, K.2006. Analysts, industries, and price momentum. ***Journal of Financial and Quantitative Analysis,*** 41(1):85-109.

Brennan, M., Asgley, J., Wang, W. 2004. Estimation and test of a simple model of intemporal capital asset pricing. ***Journal of Financial***, 59(4): 1743-1776

Chan, K., Jegadeesh, N., Sougiannis, T. 2004. The accrual effect on future earnings.***Review of Quantitative Finance and Accounting***,22(2):97-121.

Chan, W., Frankel, R., Kothari, S.P. 2004. Testing behavioral finance theories using trends ad consistency in financial performance. ***Journal of Accounting Economics***, 38(1):3-50.

Chan, L., Lakonishok, J., Swaminathan, B. 2007.Industry classifications and return comovement. ***Financial Analysts Journal,*** 63(6):56-70.

Cisco System, Inc. 2012 Annual report.

Clarke, R. 1989. SICs as delineators of economic market. ***Journal of Business,*** 62: 17-31

Claus, J., Thomas, J. 2001. Equity Premia as Low as three percent? Evidence from analysts' earnings forecasts for domestic and international stocks.***Journal of Finance***, 56:1629-66.

Keim, D.B., Stambaugh R.F. 1986. Predicting Returns in the Stock and Bonds Markets.***Journal of Financial Economics,*** 17(2): 357-390.

Fama, E.F., French, K.R. 1997. Industry costs of equity. ***Journal of Financial Economics,*** 43: 153 -93.

Frankel, R.,Lee. C.M.C. 1998. Accounting valuation, market expectation, and cross-sectional stock returns.” ***Journal of Accounting and Economics,*** 25: 283-319.

Gebhaedt,W.R., Lee, C.M.C., Swaminathan, B. 2001. Toward an implied cost-of-capital. ***Journal of Accounting Research,*** 39 : 135-76.

Gogineni, S. 2007. The Stock Market Reaction to Oil Price Changes. Unpublished dissertation of finance, Michael F. Price College of Business, University of Oklahoma.

Graham, J., Kumar, A. 2006. Do dividend clienteles exist? Evidence on dividend preferences of retail investors. ***Journal of financial economics***, 61(3):1305-1336.

Guenther, D., Rosman, A. 1994.Differences between COMPUSTAT and CRSP SIC codes and related effects on research. ***Journal of Accounting and Economics,*** 18(1):115-128.

Hong, H., Torous, W., Valkanov, R. 2006. Do industries lead stock market?***Journal of Financial Economics***, 83 (2007): 367-396.

Hong, H.,Lim, T.,Stein, J.C. 2000. Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. ***The Journal of Finance,*** 55(1): 265-295

Hrazdil, K., Scott, R. 2013.The role of industry Classification in estimating discretionary accruals.***Review of Quantitative Finance and Accounting,*** 40 (1): 15-39.

Hrazdil, K., Zhang, R. 2012. The importance of industry classification in estimating concentration ratios. ***Economics Letters,*** 114: 224-227.

Hrazdil, K., Trottier, K., Zhang, R. 2013. A comparison of industry classification schemes: A large sample stury. ***Economics Letters***, 118: 77-80.

Wolfers, J., Zitzewitz, E. 2004. Prediction Markets.***Journal of Economic Perspectives***, 18(2): 107-126.

Kahle, K., Walkling, R. 1996. The impact of industry classifications on financial research. ***Journal of Financial and Quantitative Analysis***, 31: 309-332.

Hou, K. 2007. Industry Information Diffusion and the Lead-lag effect in Stock Returns. ***Review of Financial Studies,*** 20(4): 1113-1138

Kile, C., Philliops, M.2009. Using industry classification codes to sample high-technology firms: Analyssis and recommendations. ***Journal Accounting, Auditing &Financ,*** 24(1): 35-58.

Krishnan, J., Press, E.2003.The north American industry classification system and its implication for accounting research. ***Contemporary Accounting Research***, 20(4):685-717.

La Porta, R. 1996. Expectations and the cross-section of stock returns. ***Journal of Finance***, 51(5): 1715-1742.

Pearce, E. 1957. History of the standard industrial classification. Mimeo, executive office of the president, office of statistical standards. U.S. Bureau of the budget.

Sadorsky, P. 1999. Oil Price Shocks and Stock Market Activity. ***Enery Economics,*** 21(5): 449-469

Saunders, N.C., 1999. The North American Industry Classification System: change on the horizon. ***Occupational Outlook Quarterly***, 43(3): 34-37.

Shao, J. 2011. Inflation pressure, stock option incentive and the company performance taking the food industry for an example. ***Business Management and Electronic Information (BMEI), 2011 International Conference on,*** 1: 624-627.

Standard & Poor's (S&P) and Morgan Stanley Capital International (MSCI). Global Industry Classification Standard - A Guide to the GICS Methodology, 2002.

Torous, W., Valkanov, R., Yan, S. 2004. On predicting stock returns with nearly integrated explanatory variables. ***Journal of Business,*** 77: 937-966.

Valkanov, R. 2003. Long-horizon regressions: theoretical results and applications. ***Journal of Financial Economics,*** 68:201-232.