Ai nonni e ai nipoti

Diversification in the cross-section of returns

Michele Schincaglia*

1. Introduction

The CAPM (Treynor, 1962; Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black, 1972) is a model of market equilibrium where pricing is determined by the asset correlation with systemic market risk. An informationally efficient market (Fama, 1970) should impose CAPM excess risk premium at zero. The APT model (Ross, 1976) indicates the existence of an excess premium whenever other risk factors are not priced by CAPM. The aim of this work is to investigate whether the degree of product differentiation is a significant proxy for risk to be included within the pricing kernel. In order to appropriately set the framework of this work a survey of the literature will be needed. First will be described the theory of asset pricing since Markowitz mean-variance quadratic programming problem to the most recent multifactor models, presenting the empirical evidences and the significant results. Then, fishing from industrial organization literature, the focus will move towards the theory of industry dynamics, from the classical debate between Bain's S-C-P paradigm versus Stigler's efficiency hypothesis; to the significance of market power and degree of competition on company performance and stock returns. Once the foundations are in place, a brief introduction to network literature and textual analysis will allow to setup the methodology of this study. Limitation of the results and further directions for investigating the relationship between differentiation and returns will conclude the paper.

2. Literature Survey

2.1 Quantitative Asset Pricing and the Market Model

Early attempt of quantifying the behavior of equity prices can be found in Regnault (1867), and in the precognizant dissertation thesis of Bachelier (1900). The modern framework to analyze the relation between asset prices and returns arose from the classical analyses of Cowles and Jones (1937) where they tested on forecasting ability and performance of professional investment advisors. The rejection of the efficacy of qualitative security analysis, the foundations for a

* I want to express the sincerest gratitude to Professoressa Billio for her illuminating advice and her unconditional empathy: higher quality education flourishes thanks to these figures. I want to thank also Dr. Yanguas for her intuition, her precious support, and her refreshing positivity; Dr. Pellegrino for the opportunity of work for him and learn from him how to trust my skills, even when nobody else did: I'll always owe him a great deal, and hope many students will be trained under his mentorship. I want to thank Dr. Testoni for his support, his knowledge, his patience and generosity: without him there wouldn't be a thesis. It has been a privilege to work together for the same boss. Finally, I want to thank my wife, for her strength, her professional drive and her relentless commitment: I'm everyday taught and inspired by her passion and her wisdom. I earnestly wish them to receive all the good they offered to me, and hope the best for their ambitions…*Ad maiora!*

quantitative theory for asset pricing theory were laid down. On this ground, Markowitz and Muth posited the rules of modern portfolio theory and asset pricing models: while the first engineered a "probabilistic reformulation of security analysis" (Markowitz, 1952) that allows a statistical approach to investment problems; the latter guaranteed the functioning of the mechanism by assuring a rational economic environment in which "expectations are formed […] on the structure of the relevant system describing the economy" and agents behaved in a consistently measurable pattern (Muth, 1961). These seminal theories allowed to formalize new equilibrium pricing models: the milestone among them was the Capital Asset Pricing Model, or Market Model, and its importance is demonstrated by the simultaneous formalization by many authors of the same intuition (Treynor, 1962; Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black 1972). In the *market model*, asset returns where to be decomposed by the price of time (the pure interest rate) and the price of risk, that is the expected return caused by the asset co-movement with the systemic volatility of all securities – i.e., the market. Hence the returns, and the price of an asset is related exclusively to its relation with systemic risk; any other idiosyncratic shock would not affect firms' returns, because of nonlinearity of portfolio variance and diversification effects.
It is clear why in the *market* model asset record high rather than low prices: a security that do not move when the market moves, it is related to the price of time only – i.e., the *risk-free* interest rate – and therefore should be discounted at a lower interest rate and have a higher price today. This line of reasoning provides the most important intuition for asset pricing theories: securities markets behave like insurance market, where the return is the premium paid to the investor who decides to expose himself to higher (systemic) risk than the others. A riskier event requires a higher premium. Cochrane (2021) confirmed this intuition showing how the main determinant for asset prices movements is caused by discount rates variation. In this sense it confirms that financial markets act as more as a "giant insurance market" for risk rather than a cash-flow harvesting tool. The centrality of risk premium – discount rates can help explain the amount of research in this branch – or "trunk" (Cochrane, 2005) – of finance.

2.2 The Market Model versus Consumption CAPM

Soon after the formalization of the *Market Model*, many empirical investigation of the phenomena acknowledged that theory wasn't supported by data: Friend and Blume (1970) found that the two-parameter model produced biased estimates of performance; Hakansson recognized the limitations of the assumptions modeling agents preferences and the inconsistencies arising from the myopic horizon approach; Jensen, Black and Scholes (1970) showed the flatness of the beta-return relationship, and recognized the presence of a third significant factor premia in returns regressions. These rejections of CAPM empirical soundness didn't stop the exploration for a reliable pricing model; on the contrary, they sparked the academic interest in both postulating new models and in refining the ones in place and to further explore the set of variables able to quantify assets' behavior.

In the same years CAPM was formulated, it was recognized an important limit regarding agent assumptions: whilst wealth dynamics are of primary relevance for individual decisions, the most important variable influencing agent's incentives and behavior is its ability to consume. In other words, before being an investor, every agent is a consumer, and hence his set of actions is primarily driven by the utility generated from levels of consumption. On this reasoning was formalized another breakthrough in the field of asset pricing: the Consumption CAPM, or C-CAPM. Stiglitz (1970) presented a pricing theory consistent with consumption valuations, while Fama (1970) modeled agents' utility as a function of lifetime consumption. On these premises Rubinstein (1976) and Lucas (1978) formalized a market equilibrium framework, where it is the "*consumer* behavior (that) determines the equilibrium price function". In a continuous-time setting, Breeden (1979) supported this approach by showing that it is the asset sensitivities to aggregate consumption that appropriately account for risk; and that comovement with aggregate wealth "it is not an adequate measure of an asset's risk". With the increase in computing power that followed in the next decade, Breeden, Gibbons and Litzenberger (1989) was able to empirically validate the C-CAPM, confirming that equilibrium expected excess return (over risk-free rate) are proportional to each asset sensitivity to consumption growth.

After a decade of prolific research, Fama (1991) recognized the risk that multifactor models could become "licenses" for researchers. This epistemological issue should impose a careful consideration of which variables include into a consistent and meaningful set: the evidence that would ultimately assure factor validity was its ability to fit within the intertemporal consumption-based equilibrium approach. In simpler terms, those assets which produce payoffs that covaries positively with future marginal utility (i.e., assets paying well in "bad times", when marginal utility of consumption is high) should command higher prices and lower expected returns. Again, it is the insurance framework that allows rationalize in mathematical terms the determinants of risk premia (Cochrane, 2009; Connor and Korajczyk, 2010).

Nonetheless the theoretical relevance of the model, empirical result offered little satisfaction to the supporters of the approach: Campbell and Cochrane (1999) tried to solve these failures of C-CAPM tests and to save its elegant intuition by introducing habits as a conditioning variable of consumer utility; Lettau and Ludvingson (2001) collected some empirical evidence in favor of the model by assuming that consumers *smooth out* transitory expected variation on their wealth and consequently by including wealth and labor income along with consumption in the explanatory variables vector. For these reasons, Cochrane (2009) decided to impose the C-CAPM as the general asset pricing model from which derive any other theory: the price of any asset (or contingent claim) is determined by the sum of its future cash flow discounted by a pricing kernel. The kernel is determined as the stochastic intertemporal rate of substitution between present and future marginal utility of consumption. Overall, the main conquest for pricing theory regards the potential for multidimensionality of the risk premium, and that returns sequitur linearly the assets ability to insure against wealth and consumption fluctuations.

2.3 The multifactor (r)evolution

In the same years in which was recognized the central role of consumption in asset pricing theory, and relating to the same intuition of lifetime consumption *smoothing* rooted in the classical work of Friedman (1957), Merton (1973) created the foundations of the vast field of multifactor models. Expanding logically from the work of Jensen, Black and Scholes (1970) which rejected with empirical data the linearity relation between asset prices and their market beta, Merton introduced the possibility of a nonsingular set of states of the world in which risks are dependent on the realization of a specific state. While maintaining the CAPM assumptions on investors dislike for wealth uncertainty, when Merton introduced intertemporal decision making related to the need of hedging the many future realizations of the consumption-investment opportunity set (Fama, 1996), he formalized the necessity of multifactor models in order to explain asset prices consistently. Similarly, Rosenberg (1974) observed that security returns comoved with factors other than the market premium, and that those components influencing asset returns could have been related to companies' financials and industrials characteristics. He then postulated a first taxonomy of pricing models by dividing the types of factors into two groups: the first based on no a priori knowledge on what drives stock returns, and their identification would have been pursued through purely statistical decomposition of returns variances; the second group was instead grounded on economic theory, where factors explaining returns were prespecified proxies related to "meaningful economic events". Within the first group, the work of Ross (1976) constituted an important building block of theory: in the Arbitrage Pricing Model, asset prices were determined by the law of one price, and the returns where to be explained by an unknown set of factors. The relevance of this model was due mainly by its simplicity: the relaxation of some of the stringent assumptions of the *Market Model* – i.e., utility function has to be monotonic and concave, while the market portfolio does not have to be mean-variance efficient anymore (Roll and Ross, 1980) – allowed easier formulation for new pricing models. Another important aspect of the APT was its factor indefiniteness: since it didn't indicate any potential candidates for the risk premium relevant set, it paved the way for a creative and prolific exploration of the appropriate set of factors.

Starting the prolific literature of multifactor models, Ball (1978) studied the effect of earnings, and their potential to be a and omitted variables proxy in the canonical Market Model. He found that P/E ratios should be included in a model of asset pricing to improve its explanatory power. Fama (1980) followed the multifactor route investigating if stock prices were related to pervasive economic phenomena as inflation, and to other real variables that were "fundamental determinants of equity values" like capital expenditures, sales and return on capital. Banz (1981) discovered in a study over 40 years of returns the renowned (and ambiguous) relation between stock returns and firms' size: whilst the size factor effect on prices did not have any solid theoretical foundations, it empirically succeeded in explaining both returns and the P/E effect premium, as explained by Reinganum (1980) analysis on the CAPM as a misspecified model for predicting returns. On a similar tide, Basu (1983) recognized that both E/P ratios and size were proxies for more fundamental factors driving asset prices, while Chan, Chen, Hsieh (1985) confirmed empirically that the size effect produces a significant premium in a multifactor pricing model and that in an efficient market, should constitute a rational explanatory factor. Campbell and Shiller (1988) tested the significance of real earnings in predicting over long horizon future cashflows and price, emphasizing again the importance of E/P ratios as a factor to model returns.

In their influential work, Chen, Roll and Ross (1986) investigated which undiversifiable systematic economic events could influence asset prices: they considered systemic risk factors those variables that are necessary to appropriately define the state of world. They choose as explanatory variable the industrial production, the market premium, changes in the (bond) yield curve (i.e., the term spread) and unanticipated changes in inflation (because agents' rationality would already price expected inflation). The importance of their work is related to the contribute it provides to the taxonomy of factors: explicit economic factors are split into two subgroups: those related to macro variables; and those proxied by company characteristics. On the latter set Bhandari (1988) found that, despite the recognized validity of Modigliani-Miller (1958) theorem positing the irrelevancy of a company capital structure, a characteristic as firms' leverage relates to a specific risk premium moving equity prices, even after controlling for the already known market beta and size factor. In many years of studies and experiments, asset pricing tests focused largely on the US stock universe: Chan, Hamao and Lakonishok tested whether cross-sectional effects of fundamental firm characteristics operated in other countries as well: they found significant evidence that earnings, cash flow, size and book value helped to explain returns in the Japanese equity market. Hou et al (2011) studied the significance of firm characteristics in pricing the cross-section of returns on a global scale: they found that a pricing model build on cashflow-to-price and momentum factors captures a large share of common variation among equity returns.

Fama and French (1992 and 1993) published two milestone of the asset pricing theory literature: they tested the prediction of a *three* and a *five*-factor pricing model to explain the return process of both equities and bonds. One important intuition of this work is that risk premiums should be pervasive across different types of assets, and in fact they find that, other than the market factor, company size and its book value, there can be found equity returns explanatory power both from unexpected changes in the term spread (between government bonds with different maturities) as well increases of default risk on corporate bond portfolio versus government bonds.

In 1993, Jegadeesh and Titman discovered the presence of the factor that shacked the foundations of market rationality and efficiency: it was shown that a zero-cost long-short portfolio based on relative strength produced positive performance and a statistically significant factor. The same result was reached by Hendricks, Patel and Zeckhauser (1993) and Carhart (1997) on parallel studies investigating portfolio performance: they uncovered the presence of the *hot hands* phenomenon among active fund managers.

Due to the prolific literature on multifactor model, Connor (1995) decided to deepen the taxonomy of Rosenberg and Chen, Roll and Ross providing a classification of theories, where factors could be organized into three main categories: macroeconomic, fundamental and statistical variables. Giving to the first group the highest power in term of theoretical consistency and explanatory power; and the least to the latter, due to the weak power of statistical techniques in identifying variable of intuitive appeal. Cochrane (1996) supported this classification and the view that macroeconomic variables are the only candidates that could help explaining the behavior of asset returns, while fundamental variables, being in their nature comparative among firms, could at most *describe* the returns. In this context, he tested a single-variable macro pricing model where the relevant factor was proxied returns on physical investment, finding that it performed as well as traditional C-CAPM and 5-factor model by Chen, Roll and Ross. To keep track of the evolution of asset pricing literature, Cochrane (2011) formalized a further categorization of multifactor models, where asset pricing theories can be divided into two main classes. The first class of models relates to investor dynamics, and includes: macroeconomic theories are linked to consumption aggregate risk; behavioral theories and their effects of agents' irrationality; and finance theories where returns are based on return covariance with characteristics. The second class is of pricing theories are based on frictions caused by segmented and intermediated markets, and liquidity effects.

Refuting Cochrane results, Haugen and Baker (1996) discarded the importance of macro-variables in explaining returns, while simultaneously showing that factor models based on fundamental characteristic (including the liquidity factor) have strong explanatory power, even at international level. Pastor and Stambaugh (2003) confirmed that aggregate liquidity could be a relevant state variable in asset pricing models. Sloan (1996) focused the attention on the asset side of the balance sheet and the predictive power of accruals and cash flows in describing returns, increasing the vector dimensionality of significant premiums; and Bansal, Dittmar and Lundblad (2005) supported the explanatory power of cash flows by showing the way by which they are directly related aggregate consumption risk.

Berk, Green, Naik (1999) tried to reconcile the amount of CAPM anomalies discovered in the literature by relating firm characteristics to economic theory on real options valuation model: in this framework firm characteristics are a proxy of the embedded growth options that can be realized or demised by companies through their investment choices. Chen, Lakonishok and Sougiannis (2001) utilized as well the real option framework to explain why high intensity R&D firms earn large excess returns; Li (2011) provided similar evidence when after showing that financially constrained firms earn higher returns, he found that the effect strengthens with the level of R&D expenditures. Carlson, Fisher and Giammarino (2004) and Titman, Wei, Xie (2004) kept exploring in a real option framework, studying the effects of corporate and capital investments on stock return dynamics. In 2012 Garleanu, Panageas and Yu returned on the real option framework to relate investment-based and consumption-based asset pricing literature by demonstrating how the diffusion life-cycle of technological innovation which convert growth option into assets in place reduces the risk premiums and how it is validated by empirical data on security prices behavior.

Through the years the literature kept discovering risk premiums, sometimes in apparent paradoxical contradiction the basic pricing model: Goyal, Santa Clara (2003) found that even idiosyncratic volatility could become a significant factor driving pricing assets. Trying to reconcile this paradox, Pastor and Stambaugh (2009) used an equilibrium proposition to study the effects of exogenous innovation adoption in the railroad and internet industries, on the gradual transformation mechanism of risk from idiosyncratic to systemic. Another apparent contradictory result came from the rational/irrational debate: both Harvey and Siddique (2000) from a rational risk aversion approach and Barberis, Jin and Wang (2019) from a behavioral prospect theory perspective that overweight tail distribution, demonstrated how average returns can relate (positively) negatively with (negative conditional) positive skewness in asset returns distribution.

In another attempt to provide economic validity to the pricing anomalies generated by characteristic-based portfolio, Vassalou (2003) resorted to macro-variables as well to demonstrated how GDP growth perform as well as Fama-French model of firm characteristics in explaining the cross-section of returns; while Barro (2006) found that the anticipation of rare-disasters (i.e., the *peso problem*) have a large explanatory power of the cross-section of returns.

Goyal and Welch (2006) tried to summarize the amount of empirical world up-to-date and tested both in-sample and out-of-sample the explanatory power of the factors and methods proposed by literature: macro variables as investment to capital ratio, consumption, wealth; firm characteristics and interest-rate related factors; statistical techniques as the model selection approach. They did not find one variable that had meaningful empirically robust explanatory power, and concluded that – ironically – the best estimate for equity premiums would have been resulted from an

Ordinary LS regression on market return historical averages. The same procedure provided optimal results for estimating firms' beta coefficients in the CAPM as showed in Welch (2019). On the same perspective, Fama and French (2008) showed, with a commendable effort of intellectual honesty, how common firm characteristics does not have sound statistical significance as factors, and that their explanatory power can be the result of data mining.

While maintaining that many of the anomalies introduced by the literature are not statistically solid, Hou, Xue and Zhang (2014) found that a four-factor model based on Q-theory of investment introduced by Kaldor (1966) and Brainard and Tobin (1968) consisting of market, size, profitability and investment largely summarizes the cross-section of returns as well as the more recognized Fama-French (1992) model including momentum. Novy-Marx (2013) anticipated the same discovery when they found that profitable firms earn higher adjusted returns. In the same line of thought, Fama and French (2015) that either a five-factor model combining the former 3 factor of Fama and French (1992) with profitability and an investment factors; as well as the four-factor model proposed by Hou et al. (2014) produce valid models for explaining returns.

The list of factor studies gets longer every year: Harvey, Liu and Zhu estimated that more than 300 factors can be traced on the literature, and just considering the highest ranked journals. Trying to reconcile them is an effort that goes beyond the aim of this project. For this reason, the next section will narrow the subject to a specific subset of asset pricing models, and review the research field of industry dynamics in multifactor models.

2.4 Industry Dynamics

"In an economic system, the realization of profits is the criterion to which [...] firms are selected", this statement from Alchian (1950) brought the attention on the analysis of market power and industry dynamics to explain firm performance.

The discussion relating industry structure effects on firms and consumers (welfare) is rooted in the classical economic literature. Starting from the initial work of Bain (1950), industry concentration costs become a central theme of economics: it was researched whether the degree of competition a company face, could influence the level of profit rates. For Bain (1951), the "average profit (is) higher, higher the concentration", while for Stigler (1964) the number of rivals wasn't a relevant variable in explaining the level of margins.

This debate between the supporters of the Structure-Conduct-Performance paradigm versus the Chicago school's approach built the field of Industrial Organization in economics, and provided the intuition of investigating whether industry structure has an impact on financial returns. The seminal work in this branch of finance refers to King (1966) who showed how indexes based on industry affiliation helped describe firms' performances. In other words, he documented how

the comovement of securities prices can be adequately described by a market and an industry factors model. Elton and Gruber (1973) supported the results, and found that industry indexes have good explanatory power for market returns. Cohen and Pogue (1967) tested the double-index model, and rejected the significance of industry effects. Similarly, Meyers (1973) refuted King's results using a principal component analysis methodology. Lessard (1974) reached the same conclusion as well studying the diversification effects of international portfolio. He showed that the multi-factor stochastic process of returns was specified better when national rather than industries affiliations factors were used. Dickens and Katz (1986) found how wage differentials were better explained by industry factors. Fama and French (1988) studied the role of industry correlations, along with size, in capturing the differences in return behavior. Carrying these results further, Kale, Hakansson and Platt (1992) posited that the industry exposures are "substantially more important" than common fundamentals factors such as Earning-to-Price, Book-to-Market, Size, Historical Beta and Dividends in explaining cross-sectional variance of equity returns. In a global study Roll (1992) found that industries experience different level of volatilities and plays a major role in explaining national stock price indexes behavior, and Beckers, Grinold, Rudd and Stefek (1992) showed how industry effects impact sectors across nations and could account by the proportion of returns left unexplained by the world market factor; Contending Roll's findings, Heston and Rouwenhorst (1994) found that the proportion of global returns explained by industry membership are smaller than country specific effect. The study by Griffin and Karolyi (1995) supported these results. In Fama and French (1997) study over the formal approaches of capital budgeting techniques, showed that industry risk loading plays a relevant role in the valuation problems.
Opposite to this view, Cochrane (1999) refuted the need of industry factors to be part of any multifactor pricing model, since industry portfolio are already explained by traditional CAPM, all showing the same average market return. Moskowitz and Grinblatt (1999) revived the debate by showing how momentum-based trading rules lose statistical significance once they are controlled for industry momentum investment strategies (going long the stocks from winning industries and shorting the losing ones). The same result was achieved by Hoberg and Phillips (2017) using their innovative Text-based Network Industry Classifications (TNIC) to map the network of fims' industry relations. But again Cochrane (2021) showed that industry factor comovements alone should not command risk premium.
When literature offers an equal share of contrasting results, it could be useful to tackle an interesting problem from a different perspective: Khalilzadeh-Shirazi (1974) studied which factors related to industry structure can influence profit rates, and found that product differentiation was a significant variable in explaining price-cost margins. The same intuition was supported by Porter (1979) when he observed that industry leaders – and hence industry concentration – were not profitable at all. He

brilliantly resolved the Bain's *structure-conduct-performance* paradigm with the concentration-efficiency hypothesis of Demsetz (1975) positing the presence of strategic groups (that can be formed by just one firm) based on the similarity in key variables such as product differentiation or mobility barriers, and explaining the different level of profitability across and within industry as the insulation from rivalry provided by the dynamics of strategic groups.

It was not concentration to have an effect on profit and prices, but product differentiation and barriers.

Notwithstanding the evidence on strategic groups relevancy, literature moved the attention on industry effects on profitability, as in Schmalensee (1985) and Wernerfelt and Montgomery (1988) when they tested the explanatory power of industry concentration in driving accounting rate of returns; or in Lewin, Cowen and Mowery (1985) where they showed how the persistence of profitability of firms in concentrated industries allows for higher R&D investments. In line with the industrial organization literature, Hou and Robinson (2006) investigated whether industry concentration is a priced factor in a multifactor pricing model through two distinct channels: barriers to entry insulates firms from distressed risk as indicated by Schumpeter, and firms with more market power are less risky because they engage in less innovation, as postulated by Bain. They confirmed that firms in highly concentrated industries earn lower returns, even after controlling for traditional firm characteristic premiums as size, value, momentum and others. In a similar work Peress (2010) showed that concentration generated lower returns because of the ability of monopolist firms in insulating their profits by passing shocks onto their customers. Aguerrevere (2009) investigated in a framework of real growth options the effects of market concentration and found that risk varies with the level of demand: competitive industries are riskier when demand is low, while the opposite is true when demand is high, because growth options are more (less) riskier in good (bad) times. Hoberg and Phillips (2010) confirmed the results, observing that competitive industries become riskier in periods when product demand decreases – i.e., during recessions.

On the effect of concentration on returns, a recent study by DeLoecker and Eeckhout (2017) provided more ambiguous results: they investigated the effects of the increase of markups (i.e., revenues minus total variable costs) in the last 40 years and found that higher markups translated in higher dividends, higher profits and then higher market power. Consequently, an environment characterized by stronger levels of market power implies an overvalued equity market with respect to a more competitive economies, refuting the idea that concentration results in lower returns. Similarly, Syverson (2019) explained how the literature industry effects subsumed from concentration measures is misleading since concentration is "worse than a noisy barometer". Lastly, Pellegrino (2021) studied the effect of increased market power and industry concentration on total surplus, and measured the welfare loss caused by increased industry concentration. From an asset pricing perspective, the implicit consequence of a systemic loss in welfare could slide downward investors preferences along the risk aversion curve, imposing a generalized increase on the risk premiums. Literature on industry and concentration effects on returns leaves too much room for opposite evidence in the results. It may be more productive to leave industry considerations on the background and focus towards the intuition provided by Porter regarding the dynamics regulating strategic groups and the more specific aspect of market dynamics, which is precisely the role that product differentiation may have in explaining the cross-section of returns.

2.5 Product Market Differentiation and Network Effects

In a study relating competition and company performance, Nickell (1996) showed that competition is associated to higher productivity growth rates, and that productivity benefits eventually accrues only to the fittest among competitors: productivity gains are enjoyed only to firm that survive the competition mechanism. Since the intensity of rivalry is inversely related to the chances of survival, competitive environments are riskier, and firms facing higher levels of competition should command higher returns. This is the main proposition of this paper.

In a study considering how managerial incentives are affected by competition, Schmidt (1997) identified another channel through which competition could affect return riskiness: by reducing the company profitability, competition reduces the firm's distance from default, especially during recession periods. That is precisely the kind of dynamics that could impact the investor's marginal utility of consumption and alter its intertemporal rates of substitution, as theorized by Cochrane (2009) in the equilibrium framework of the consumption asset pricing model. Similarly, Raith (2003) found that product substitutability impact firm riskiness: the rise in the elasticity of demand caused by increased competition, increase the variance of firm profits. Building on the same logic, Gaspar and Massa (2006) found that profits of firms having more rigid demands varies less than firms facing higher elasticity over their products, and when this kind idiosyncratic risk is priced, investors will require higher rate of returns for their holdings facing higher competitive pressures. Analogously, Bustamante and Donangelo (2017) found that the combination of higher operating leverage combined with threat of entries drives the observed positive relation between competition and rate of returns. Another mechanism supporting the correlation of product market competition and asset returns is illustrated by Valta (2012). He shows that since firms operating in competitive industries face higher risk of defaults and lower liquidation values, they suffer from significantly higher cost of debt, while leading firms have access to cheaper financing.

From a corporate finance perspective, Hoberg and Phillips (2014) found that company that sells more unique products have higher stock valuations than companies that makes products easier to replicate. In financial terms higher valuations means lower returns for more differentiated companies. Gu (2016) studied the relation of R&D intensity and product market competition in a real options framework, and found that among R&D intensive firms, the higher the level of competition a firm face, the higher the rate of returns earned. With a Mertonian state-contingent hedging logic, Hoberg and Phillips (2016) showed that firms react to negative demand shocks by investing into differentiated product markets.

It can be noted that this area of research relating levels of competition and rate of returns have produced more consistent results than literature on industry indexes, or again the contradictions found in studies on industry concentration effects. Because of this consistency, it will be interesting to further investigate the competition-return dynamics, and provide new evidence on differentiation effects: the isolation from competition provided by product differentiation can be the lever through which companies hedge against profit variability, productivity shocks, and risk of default, while increasing at the same time the value of their growth options. On the contrary, the more the company similarity to other companies, the lesser it will be insulated from those risks. If the nature of these risks is idiosyncratic, there should be no effect on equity values because of the diversification mechanism indicated by Markowitz (1952). If those risks manifest themselves in a systemic fashion, especially during periods of high marginal utility of consumption (i.e., recessions), the more intense the rivalry a firm face, the higher the premium will be required by rational investors for holding that company's equity.

The view where idiosyncratic risks translates into systemic ones can be traced in macroeconomics studies on network effects: Gabaix (2011) explained how idiosyncratic firm-level shocks have the potential to translate into aggregate shocks to GDP and financial markets; Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012) showed how, in connected economies, intersectoral input-output closely-linked networks translate microeconomic sector specific shocks into increase of aggregate volatility through cascade effects. Similar intuitions are provided by Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) in the study of propagation of negative shocks throughout different financial network interconnectedness configurations; and by Carvalho and Tahbaz-Salehi (2019) describing the propagation of shocks through production networks from a specific industry to the rest of the economy. A limitation in attributing cascade effects from idiosyncratic competition risk to equilibrium returns is the orientation of these dynamics: while cascade effects emerge mainly from vertical relatedness, competition impact companies' horizontal relations, as pointed out by Hoberg and Phillips (2016). From Hoberg and Phillips is drawn the main tool to test the proposition of this study: since theories and empirical evidence shows that risk is positively related with the intensity of competition each firm faces, it will be tested whether network similarity scores can be used as a proxy to measure the negative relation between product differentiation and equity returns.

3. Data

3.1 Data Sources

This study tries test whether the degree of competition helps explaining returns. To test the model, data will be retrieved from different sources: companies monthly stock returns from 1989 to 2019 are retrieved from the *CRSP-Compustat* Merged database; while monthly risk-free returns, *market*, *size* and *value* premium are found in Professor French Data Library website[1]. Finally, to produce a measure indicating the level of competition a firm face, it is adopted the similarity score computed by Hoberg and Phillips (2016) and located in their online Data Library[2].
They used Natural Language Processing techniques on companies SEC mandatory 10-K filings to computed product total similarity score for each company. The differentiation score can then be used to rank companies with respect to their vocabulary closeness to other companies, and so to define their level of distance/centrality with respect to other companies in the market network.

Origins of textual analysis literature can be traced since Cowles (1933) and the ability analyst advice to forecast stock performance. More recently, with the advent of more and more pervasive media, Antweiler and Frank (2004) investigated the role of internet stock message boards in conveying relevant financial information and predicting returns. Advancements in computing power allowed Tetlock (2007) to combine a quantitative content analysis to principal component factor analysis to study the interaction between media textual content and market activity, and again in Tetlock, Saar-Tsechansky and Macskassy (2008) to found that financial press negative textual contents predicted low accounting earnings. Li (2008) focused on official financial report, studying whether lexical properties of 10-K filings explain performance and earnings persistence. On a similar approach, Loughran and Mc Donald (2011) studied how textual tone in 10-K filings is linked to firm financial characteristics (trading volume, return volatility and others) and can act as a proxy for other information that drives returns. Baker, Bloom and Davis (2016) developed an automated language processing tool to use the press news coverage to estimate an index of economic policy conditions. Grentzkow, Kelly and Taddy (2019) presented the many fields in which text became a relevant input in economic research: due to the developments in machine learning statistical techniques, content analysis can be used in finance to study return predictability, in macroeconomics to forecast trend in relevant variables, in political economy to analyze political agendas, in marketing to study the drivers of consumer behavior and in industrial organization to build appropriate definitions of product markets.

Hoberg and Phillips NLP categorization fits in this last branch of analysis: from EDGAR website, they gather SEC mandatory annual 10-K filings from 50,673

[1] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[2] https://hobergphillips.tuck.dartmouth.edu/

companies. Data was collected from 1989 to 2019. Within each document, they performed a natural language processing (NLP) on the business description section of 10-K filings. The output of the text-parsing task allows the authors to build the set W including all unique words used by companies in the 10-K Product Description section. Through this technique, each company vocabulary can be represented by a word vector $\mathbf{p}_i$ of length $\boldsymbol{W}$, "with each (row) being populated by number one if firm *i* uses the given word, and zero if it does not", and then normalized into a vector $\mathbf{v}_i$,

$$\mathbf{v}_i = \frac{\mathbf{p}_i}{\sqrt{\mathbf{p}_i \cdot \mathbf{p}_i}}$$

Because of this normalization, firm's vectors will reside in a $\boldsymbol{W}$-dimensional unit sphere. For data completeness, the HP dataset will be merged with CRSP database through the company GVKEY (unique company identifier), and companies having less than 50 observations throughout the 30-years' time period are removed from the sample to improve the robustness of the analysis. Following this arbitrary cutoff, the dataset is populated by 11,183 firms.

The second building block of the analysis is a measure that can map network relation in terms of closeness or distance of firms in product space, year by year: the *Product Cosine Similarity* matrix $\mathbf{M}_t$ is obtained by the dot-product of each vector $\mathbf{v}_i$, and measures the angles between any given pair of vectors on a unit sphere. This is the measure of closeness or distance between each pair of companies, and is computed as

$$Product\ Cosine\ Similarity_{i,j} = (\mathbf{v}_i \cdot \mathbf{v}_j)$$

And

$$PCS\ Matrix = \mathbf{M}_t = \begin{bmatrix} \mathbf{v}_{i,t} \cdot \mathbf{v}_{i,t} & \cdots & \mathbf{v}_{i,t} \cdot \mathbf{v}_{n,t} \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{n,t} \cdot \mathbf{v}_{i,t} & \cdots & \mathbf{v}_{n,t} \cdot \mathbf{v}_{n,t} \end{bmatrix}$$

The matrix is a common tool for quantitative information processing studies (Sebastiani 2002) and has the virtue of allowing intuitive graphical representation of network relations.

Figure 1: Example of a Product Space

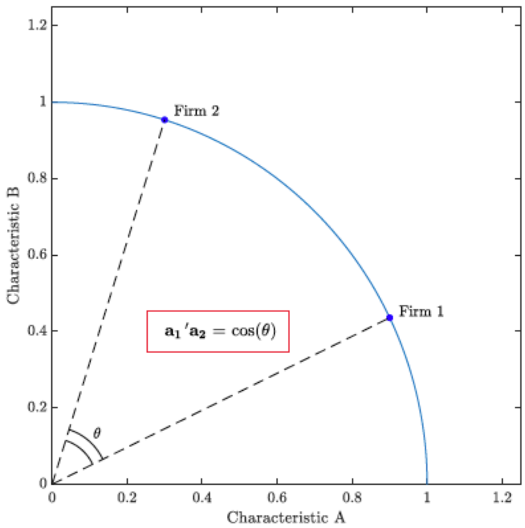

Figure Notes: The figure is from Pellegrino (2021) and represents the simple case of two firms and two words/characteristics. Each firm is a vector on the unit hypersphere (in this case, a circle). The dot product of two vectors equals the cosine of the angle θ. The tighter the angle, the higher the similarity.

Figure 2: Visualization of the Product Space

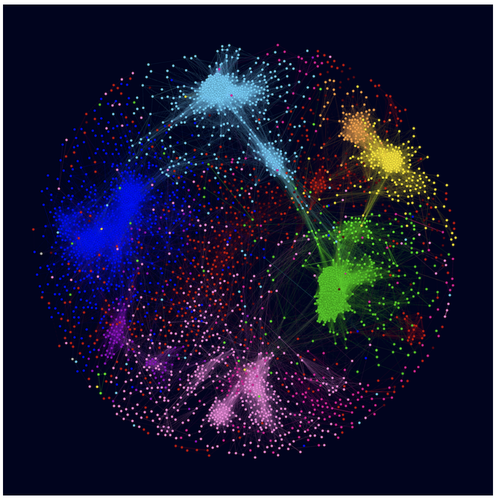

Figure Notes: The diagram is a two-dimensional representation of the network of product similarities and it is designed by Pellegrino (2021) through a dimensional-reducing algorithm used in social network analysis. Firm pairs that have thicker links are closer in the product market space. Different colors identify different industries.

Network literature (Newman, 2018) helps interpreting the similarity matrix $\mathbf{M}_t$ as the network adjacency matrix $\mathbf{A}_t$, where values are scaled node links between pair of firms. Vectorizing the matrix $\mathbf{M}_t$, we can obtain a vector $s_{TSS,t}$ indicating in each row the network scaled degree of node centrality for each company. In other words, the firm's vector similarity scores are the proxy used in the analysis to indicate the degree of competition that each company holds in the multidimensional network of product markets: the higher the score, more intense the competition the company faces; the lower the value the more differentiated a company is

$$\mathbf{s}_{TSS,t} = \mathbf{M}_t \times \mathbf{1^n}$$

To identify the degree of differentiation that each company has with respect to the market, it will be used the *nabla* symbol ∇ because of its relatedness with the mathematical concept of divergence - i.e., tendency to converge toward or diverge from a point. In the empirical procedure, higher the degree of ∇-score, the higher the higher the similarity and the lower the differentiation of company *i* in the year *t*.
To provide robustness to the results of analysis, the ∇-score will be computed in two different ways: an average of similarity for each pairwise score, called $\nabla_{AVG,t}$ that is obtained in this paper by computing every year for each firm the average value of pairwise cosine similarity,

$$\boldsymbol{\nabla}_{AVG,t} = \frac{1}{\mathrm{n}} \times (\mathbf{M}_t \times \mathbf{1^n} - \mathbf{I})$$

Where $\mathbf{M}_t$ is the year-t n×n *Product Similarity Matrix*, $\mathbf{1^n}$ is a vector of n×1 ones, and $\mathbf{I}$ is the identity matrix.
The second way uses a total similarity score computed as the sum of similarity scores that each firm have in the Hoberg and Phillips cosine similarity matrix. This second value will be called $\nabla_{TSS,t}$, and it is computed as $\mathrm{s}_{TSS,t}$,

$$\boldsymbol{\nabla}_{TSS,t} = (\mathbf{M}_t \times \mathbf{1^n})$$

It is called *Total Sum of Similarity* and it is computed as the total similarity score provided by the Hoberg and Phillips in their web Data Library minus one – i.e., the similarity of the firm with itself it is not considered in the computation of this paper, and consequently the difference between HP score and ∇_{TSS} is equal to one.

3.2 Assumptions

Since consistency is the main element that guarantees validity in hypothesis testing, it is better to start the analysis of the data by defining as clear as possible the assumptions of the model. The first-order condition (Fama, 1970; Lucas, 1978; Cochrane, 2009) to be imposed on an agent holding a risky asset for two subsequent time periods is:

$$\mathbb{E}[(1 + R_{i,t+1})m_{t+1}|\boldsymbol{\Omega}_t] = 1$$

where $(1 + R_{i,t+1})$is the total return on asset *i*, $m_{t+1}$is the agent's marginal rate of substitution between the two periods, and $\boldsymbol{\Omega_t}$ is the information set available to the investor at time *t*. The marginal rate of substitution $\boldsymbol{m}_{t+1}$ can be defined as the pricing kernel, or stochastic discount factor that prices each future payoff realizations of the risky assets.
Since the marginal rate of substitution is not observable, in order obtain testable estimates from first-order condition, it is necessary to assume observable proxies for

the marginal rate of substitution m_{t+1}. From this indefiniteness arouse the richness in variety and amount of asset pricing models, since each author speculate about the adequate proxies that can describe the marginal rate of substitution m_{t+1}. As indicated in the literature survey section, the proxies can be either computed directly from observed covariations portfolio returns or derived from nonmarket variables (e.g., inflation as in Rosenberg, 1974 or aggregate consumption as in Breeden and Litzenberger, 1978).

A second element to be considered is the joint simultaneous specification of marginal rate of substitution, statistical distribution of the proxy, and agents' preferences. The first limitation in this paper is that there will be no assumptions regarding investors preferences.

While preferences are not modeled, the assumption that agents' beliefs are unbiased is included. In this way, ex ante expectations of returns can coincide with true expectations, and first and second moments of returns can be inferred correctly from empirical frequencies of data (Bossaerts, 2001). It has been noted by many authors, as Ghysels (1998), Ang, Chen (2007) that betas of unconditional multifactor models show significant time variation: while it could be useful to investigate whether conditional pricing models would reduce estimation biases, this analysis assume time stationarity for the beta coefficients as well.

Another assumption states that the data generating process of the pricing kernel proxies (market return, consumption, firm characteristics) is time-invariant so that the ergodic theorem could hold. While Dhrymes (1984) imposed the same assumptions in its formulation of a multifactor model, abundant evidence against premiums stationarity was found in the literature as in Cochrane (2007) and Gagliardini, Ossola and Scaillet (2016). Still, in this paper will be maintained that premia and residuals have time-invariant first and second moments, in order to guarantee the unbiasedness of the OLS estimation. Lastly, as pointed out by Connor (2010) the vector of regression residuals must be cross-sectionally uncorrelated.

The marginal rate of substitution will be assumed to have a linear specification with respect to the multifactor model:

$$\boldsymbol{m}_{t+1} = \mathbf{a}_t + \mathbf{b}_t \mathbf{F}_t$$

Where the intercept $\mathbf{a}_t$, the vector $\mathbf{b}_t$ of coefficients, and the risk-premium vector $\mathbf{F}_t$ are functions of the period-*t* information set (i.e., lagged values). By expanding expectation of the marginal rate of substitution, after doing computation and assuming the existence of a risk-free asset, the following specification results:

$$\mathbb{E}[r_{i,t+1}] = \frac{Cov_t[r_{i,t+1} r_{\mathbf{F},t}]}{Var_t[r_{\mathbf{F},t}]} \mathbb{E}[r_{\mathbf{F,t}}]$$

or

$$\mathbb{E}[r_{i,t+1}] = \boldsymbol{\beta}_{i,t}\mathbb{E}[r_{\mathbf{F,t}}]$$

where the vector of coefficients $\boldsymbol{\beta}_{i,t}$ is a function of the period-*t* information set, and r_i represents the net excess return over the risk-free rate of the firm *i*. This expression shows how the expected excess return are obtained from the product of each firm beta coefficient and the vector of risk premiums.
The estimation techniques (Fama, Macbeth 1973) of the pricing model requires two steps: first, it is run a time-series regression for each firm-*i* realized excess returns against the $\boldsymbol{\lambda}$-vector premiums, and tested whether: the intercept α_i is zero, and the $\boldsymbol{\beta}_i$ coefficients are significant. The second step involves a cross-sectional regression of time-*t* excess returns against the average of estimated $\boldsymbol{\beta}_i$ coefficients as explanatory variable, and this time the t-test is performed to check whether the vector of risk premiums $\boldsymbol{\lambda}_t$ is significantly different from zero.

3.3 Test

The estimation process of beta coefficients and risk premiums will follow the two-step procedure explained by Fama, Macbeth (1973). If the model would have been an equilibrium one, a different procedure must have been used. The first step involves the estimation of each firm's beta coefficient through a times series OLS regression of excess returns against the factor premium vector. The beta estimates for fim *i* will be the average of the time-series estimates. Hence, there will be estimated a beta coefficient (factor loading) for each company. In vector notation

$$\mathbf{r}_i = \mathbf{b}_i\mathbf{F} + \mathbf{u}_i$$

Where $\mathbf{r}_i$ is the excess returns vector of each company excess return (over risk-free) time-series. The $\hat{\mathbf{b}}_i$ are the averages of the time-series estimates of the beta coefficients; the $\mathbf{F}$ is a vector of factor premiums that includes the market premium, Fama-French *value* and *size* premiums, and the differentiation factor; and $\mathbf{u}_i$ is a vector of i.i.d. residuals. Since this study theorize the positive relation between risk and competition, this implies that the level of expected returns increases linearly with the degree of competition a company must confront with. For this reason, the differentiation factor ∇ is calculated as the return on an investment strategy that goes long the average returns of the fifth quintile and shorts the first quintile companies of the ∇-scores distribution.
Once betas are estimated, the second step consists in the estimation of the factor premia. A cross-sectional regression is performed using the vector of betas as explanatory variable, while the lambda premiums are the coefficients to be estimated. The annual factor premium is computed as the average of the cross-sectional estimates for lambda: in other words, there will be a factor premium for each year t. In vector notation

$$\mathbf{r}_t = \boldsymbol{\lambda}_t\hat{\mathbf{b}}_t + \mathbf{u}_t$$

Where $\mathbf{r}_t$ is the excess returns vector representing excess returns for each year; whereas $\boldsymbol{\lambda}$ is the vector of differentiation premium estimates $\widehat{\lambda_t}$, In this second step, excess returns will have a different meaning depending on the multifactor model specification: it can then be interpreted as excess return over risk-free, or risk free and market premium, or risk-free and the Fama-French 3 factor. The reason why it's computed in this fashion is because the factor others than ∇ are already computed and published through the Professor French website (the library includes: one-month risk-free rate; market, size and value premiums). The betas are the time-series average of coefficients estimated for each company in the first step of the regression, and the factor premium is now the annual average of the lambda coefficient obtained through the cross-sectional OLS regression against estimated beta.
In order to test the explanatory power of product differentiation, six different specifications will be tested. The specifications are divided into two groups: the first will test the explanatory power of ∇_{AVG} - the average level of differentiation – over excess returns over risk free, to observe the explanatory without any equilibrium model restriction; then over CAPM excess returns; and lastly over the FF 3-factor model. Even if the literature offers widespread evidence against the *market* model, it is of practical interest because it reflects average investor holdings in the real world, as pointed out by Roll and Ross (1994). For a robustness check, a second group of regression will be performed using ∇_{TSS} score in place of the average score, again with respect of excess returns over the three aforementioned specifications.

3.4 Limitations

Two important caveat of this paper refers to the fact that the model does not imply any sense of causality between the factors vector $\mathbf{F}$ and the vector of excess returns $\mathbf{r}$. Furthermore, the chosen factors may or may not be the real risks driving stock returns, as pointed out by Roll, Ross, Elton, Gruber and Grinold (1994).
Furthermore, given the large number of factors that have been tested within and outside the literature, in order to provide valid statistical significance to the estimated premiums and avoid Type I errors, as stressed by Linnainmaa and Roberts (2018), a t-statistic cutoff of 3.0 will be imposed to the analysis, following the recommendation of Harvey, Liu and Zhu (2015) and the methodology of Green, Hand and Zhang (2016). Giglio, Liao and Xiu (2020) solved the issue of data snooping (i.e., Type I error) in linear asset pricing models taking advantage of modern machine learning tools: these techniques are built to improve the validity of estimates by replacing traditional OLS with advanced shrinking methods on discount factors estimates (Kozak, Nagel and Santosh 2019).
While researchers in finance and in many other fields rely increasingly on machine learning to apply dimension-reduction procedures on the data – the LASSO technique developed by Tibshirani (1996), used by Rapach, Strauss, Zhou (2013) to investigate equity relationship at international level; the tests on the *zoo of factors* in

Feng, Giglio and Xiu (2020); the elastic net formalized by Zou and Hastie (2005); and PCA and random forests by Breiman (2001); the advanced nature of these techniques and tools is beyond the scope of this investigation.
Lastly, the ultimate test of validity will be performed after the publication of the results, as pointed out by McLean and Pontiff (2015) whom after studying a large amount of pricing explanatory variables, found that almost all the anomalies lost their significant predictive power.

4. Results

Three different regressions were performed against the two differentiation factor ∇_{AVG} and ∇_{TSS}, for each year, for three differently computed excess return: $\mathbf{r}_{e}$ shows monthly excess returns over the risk-free rate; the second return $\mathbf{r}_{e+CAPM}$ are the excess return over risk-free and CAPM returns; the last specification $\mathbf{r}_{e+3FF}$ refers to excess returns obtained subtracting from the monthly returns the monthly risk-free rate and the monthly *market*, *size*, and *value* returns indicated in Fama and French (1992). The values of ∇_{AVG} and ∇_{TSS} are the sample means of the vector $\hat{\boldsymbol{\lambda}}_t$ of annual estimated premiums. Data on security returns are provided by the CRSP-Compustat database; risk-free, market, value and size premiums are retrieved from Professor French Data Library; differentiation scores are computed from Hoberg and Phillips NLP analysis of EDGAR 10-K filings. Results relates to the mean value obtained for the differentiation factors in the three different specifications.

Panel A. Differentiation factor ∇ in three different specifications.

	∇_{AVG}	*t*-value	∇_{TSS}	*t*-value
$\mathbf{r}_{e}$	0,2630**	237,27	-0,1720**	-1,7e+02
	(0,0011)		(0,0010)	
$\mathbf{r}_{e+CAPM}$	0,2278**	207,72	-0,1483**	-1,5e+02
	(0,0011)		(0,0010)	
$\mathbf{r}_{e+3FF}$	0,3316**	290,46	-0,2085**	-2,1e+02
	(0,0011)		(0,0010)	

** denotes t-statistics significant at 5 percent level

The first element to be noticed is the wedge drawn between the two measures of differentiation. Instead of providing a tool for checking the validity of the differentiation score, it seems the two premiums measures different dimension of competition. While the firm *average similarity* score has a positive relation with excess returns, with an increasing premium when more factors are subtracted from net returns; it seems that *total similarity* shows a negative relation with returns. Both factors, in all the regression are significantly different from zero, and the null-hypothesis of zero effect can be soundly rejected even following the more stringent prescription of Linnainmaa and Roberts (2018).

One reason that could explain the opposite effect of the two factors may refer to the fact that similarity *averages* and total sums describe different phenomena. While the ∇_{AVG} indicates the average level of competitive pressures a company must face; the total sum of similarity ∇_{TSS} may group together firms with high-degree of rivalry among few competitors and firms with a large amount of more differentiated competitors. If this intuition would be proven correct, the average similarity premium ∇_{AVG} would be a more consistent proxy for measuring competition intensity effects on returns.
While maintaining the opposite results for the two factors, it can be interesting to investigate whether these premiums offer diversification benefits in a portfolio perspective.

Panel B. Correlation coefficients between ∇_{AVG} and FF factors.

	∇_{AVG}	*Market*	*Size*	*Value*
∇_{AVG}	1,0000			
Market	-0,0698	1,0000		
Size	0,0271	0,1824	1,0000	
Value	-0,1154	-0,2195	-0,1418	1,0000

Panel C. Correlation coefficients between ∇_{TSS} and FF factors.

	∇_{TSS}	*Market*	*Size*	*Value*
∇_{TSS}	1,0000			
Market	-0,0473	1,0000		
Size	0,0168	0,1824	1,0000	
Value	-0,0688	-0,2195	-0,1418	1,0000

In an investment strategy combining the four factors, the ∇-factor provide a small diversification effect versus *CAPM* and *value* portfolios, while increasing a little the riskiness of the portfolio when combined with a *size*-oriented strategy. This small positive correlation is supported by both school of industrial organization literature, whose posits that the degree of competitiveness a firm face can be impact its size, as described by Smirlock, Gilligan and Marshall (1984).
As previously observed in regression results, the average similarity factor has stronger negative and positive correlations with Fama-French factors, supporting the hypothesis of its stronger descriptive power of product diversification effects.
One test to further investigate the validity of the proposed factor model concern the time-series of estimated residuals. A test of autocorrelation on the lambda regression residuals checks the assumption of independent and stationary residuals and, in case of test rejection, it signals the potential presence of omitted pricing variables. A conventional technique used in the literature is the test performed by Shanken (1992) and is beyond the scope of this project.

5. Conclusions

This paper analyzed the prediction in classical literature concerning the effect that competition has on asset prices, and contributed to the debate by formulating the hypothesis of a positive relation between the degree of product market differentiation and asset returns. Using the NLP data from Hoberg and Phillips (2016) analysis, two measures of company differentiation have been computed.

While the *total sum of similarity* score ∇_{TSS} that presented an opposite effect with respect to the hypothesis; the average similarity score ∇_{AVG} resulted to be a significant proxy for the differentiation premium, where firms with higher product market similarity (i.e., less differentiated) maintain a positive relation with expected returns. The presented results contribute to the market structure debate, providing some evidence in favor of the presence of a systemic return premium increasing linearly with the level of competition; and supporting from an investment theory perspective the negative relation found in the literature between level of competition and size of the companies. More sophisticated tests on the significance of the results should be performed to support the validity of these findings.

NLP and advanced machine learning techniques offer the opportunity to deepen the knowledge of financial market dynamics. Starting from the preliminary results of this paper, research could expand into two further directions: it could be tested whether the average similarity score ∇_{AVG} supports the results of multifactor pricing theories relating returns, market rivalry and R&D expenditures described by Chen, Lakonishok and Sougiannis (2001) and Li (2011); another interesting aspect would be the exploration of network vertical effects, if NLP techniques could effectively quantify the level of vertical integration among firms, and help predict networks' cascade effects studied in Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012).

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. "The network origins of aggregate fluctuations." *Econometrica* 80, no. 5 (2012): 1977-2016.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. "Systemic risk and stability in financial networks." *American Economic Review* 105, no. 2 (2015): 564-608.

Aguerrevere, Felipe L. "Real options, product market competition, and asset returns." *The Journal of Finance* 64, no. 2 (2009): 957-983.

Alchian, Armen A. "Uncertainty, evolution, and economic theory." *Journal of Political Economy* 58, no. 3 (1950): 211-221.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. "The cross-section of volatility and expected returns." *The Journal of Finance* 61, no. 1 (2006): 259-299.

Ang, Andrew, and Joseph Chen. "CAPM over the long run: 1926–2001." *Journal of Empirical Finance* 14, no. 1 (2007): 1-40.

Antweiler, Werner, and Murray Z. Frank. "Is all that talk just noise? The information content of internet stock message boards." *The Journal of Finance* 59, no. 3 (2004): 1259-1294.

Bachelier, Louis. "Théorie de la speculation." *Annales scientifiques de l' É.N.S. 3e série*, tome 17 (1900): 21-86.

Bain, Joe S. "Workable competition in oligopoly: theoretical considerations and some empirical evidence." *The American Economic Review* 40, no. 2 (1950): 35-47.

Bain, Joe S. "Relation of profit rate to industry concentration: American manufacturing, 1936–1940." *The Quarterly Journal of Economics* 65, no. 3 (1951): 293-324.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. "Measuring economic policy uncertainty." *The Quarterly Journal of Economics* 131, no. 4 (2016): 1593-1636.

Ball, Ray. "Anomalies in relationships between securities' yields and yield-surrogates." *Journal of financial economics* 6, no. 2-3 (1978): 103-126.

Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad. "Consumption, dividends, and the cross section of equity returns." *The Journal of Finance* 60, no. 4 (2005): 1639-1672.

Banz, Rolf W. "The relationship between return and market value of common stocks." *Journal of Financial Economics* 9, no. 1 (1981): 3-18.

Barberis, Nicholas, Lawrence J. Jin, and Baolian Wang. "Prospect theory and stock market anomalies." *The Journal of Finance* 76, no. 5 (2021): 2639-2687.

Barro, Robert J. "Rare disasters and asset markets in the twentieth century." *The Quarterly Journal of Economics* 121, no. 3 (2006): 823-866.

Basu, Sanjoy. "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence." *Journal of Financial Economics* 12, no. 1 (1983): 129-156.

Beckers, Stan, Richard Grinold, Andrew Rudd, and Dan Stefek. "The relative importance of common factors across the European equity markets." *Journal of Banking & Finance* 16, no. 1 (1992): 75-95.

Berk, Jonathan B., Richard C. Green, and Vasant Naik. "Optimal investment, growth options, and security returns." *The Journal of Finance* 54, no. 5 (1999): 1553-1607.

Bhandari, Laxmi Chand. "Debt/equity ratio and expected common stock returns: Empirical evidence." *The Journal of Finance* 43, no. 2 (1988): 507-528.

Black, Fischer. "Capital market equilibrium with restricted borrowing." *The Journal of Business* 45, no. 3 (1972): 444-455.

Bossaerts, Peter. *The paradox of asset pricing*. Princeton University Press, 2013.

Brainard, William C., and James Tobin. "Pitfalls in financial model building." *The American Economic Review* 58, no. 2 (1968): 99-122.

Breeden, Douglas T. "An intertemporal asset pricing model with stochastic consumption and investment opportunities." *Journal of Financial Economics* 7 (1979): 265-296.

Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger. "Empirical tests of the consumption-oriented CAPM." *The Journal of Finance* 44, no. 2 (1989): 231-262.

Breeden, Douglas T., and Robert H. Litzenberger. "Prices of state-contingent claims implicit in option prices." *Journal of Business* 51, No.4 (1978): 621-651.

Burmeister, Edwin, Richard Roll, Stephen A. Ross, Edwin J. Elton, Martin J. Gruber, Richard Grinold, and Ronald N. Kahn. A *Practitioner's guide to factor models*. The Research Foundation of the Institute of Chartered Financial Analysts (1994).

Bustamante, M. Cecilia, and Andres Donangelo. "Product market competition and industry returns." *The Review of Financial Studies* 30, no. 12 (2017): 4216-4266.

Campbell, John Y., and John H. Cochrane. "By force of habit: A consumption-based explanation of aggregate stock market behavior." *Journal of Political Economy* 107, no. 2 (1999): 205-251.

Campbell, John Y., and Robert J. Shiller. "Stock prices, earnings, and expected dividends." *The Journal of Finance* 43, no. 3 (1988): 661-676.

Carhart, Mark M. "On persistence in mutual fund performance." *The Journal of Finance* 52, no. 1 (1997): 57-82.

Carlson, Murray, Adlai Fisher, and Ron Giammarino. "Corporate investment and asset price dynamics: Implications for the cross-section of returns." *The Journal of Finance* 59, no. 6 (2004): 2577-2603.

Carvalho, Vasco M., and Alireza Tahbaz-Salehi. "Production networks: A primer." *Annual Review of Economics* 11 (2019): 635-663.

Chan, Kevin C., Nai-fu Chen, and David A. Hsieh. "An exploratory investigation of the firm size effect." *Journal of Financial Economics* 14, no. 3 (1985): 451-471.

Chan, Louis KC, Yasushi Hamao, and Josef Lakonishok. "Fundamentals and stock returns in Japan." *The Journal of Finance* 46, no. 5 (1991): 1739-1764.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis. "The stock market valuation of research and development expenditures." *The Journal of Finance* 56, no. 6 (2001): 2431-2456.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross. "Economic forces and the stock market." *Journal of Business* 59, No.3 (1986): 383-403.

Cochrane, John H. "A cross-sectional test of an investment-based asset pricing model." *Journal of Political Economy* 104, no. 3 (1996): 572-621.

Cochrane, John H. *Asset pricing: Revised edition*. Princeton University Press, 2009.

Cochrane, John H. "Portfolio advice for a multifactor world." *Economic Perspectives,* Vol. 23, 3rd, No. 3, (1999).

Cochrane, John H. "Financial markets and the real economy." *Foundations and Trends in Finance*, Vol. 1, No. 1 (2005): 1-101.

Cochrane, John H. "The dog that did not bark: A defense of return predictability." *The Review of Financial Studies* 21, no. 4 (2008): 1533-1575.

Cochrane, John H. "Presidential address: Discount rates." *The Journal of Finance* 66, no. 4 (2011): 1047-1108.

Cochrane, John H. "Portfolios for long-term investors." *Review of Finance* 26, no. 1 (2022): 1-42.

Cohen, Kalman J., and Jerry A. Pogue. "An empirical evaluation of alternative portfolio-selection models." *The Journal of Business* 40, no. 2 (1967): 166-193.

Connor, Gregory. "The three types of factor models: A comparison of their explanatory power." *Financial Analysts Journal* 51, no. 3 (1995): 42-46.

Connor, Gregory, and Robert A. Korajczyk. "Factor models of asset returns." *Encyclopedia of Quantitative Finance, Rama Cont, ed., Chicester: Wiley* (2010).

Connor, Gregory, and Robert A. Korajczyk. "Factor models in portfolio and asset pricing theory." In *Handbook of Portfolio Construction*, pp. 401-418. Springer (2010).

Cowles 3rd, Alfred. "Can stock market forecasters forecast?" *Econometrica, Journal of the Econometric Society* (1933): 309-324.

Cowles 3rd, Alfred, and Herbert E. Jones. "Some a posteriori probabilities in stock market action." *Econometrica, Journal of the Econometric Society* (1937): 280-294.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. "The rise of market power and the macroeconomic implications." *The Quarterly Journal of Economics* 135, no. 2 (2020): 561-644.

Dhrymes, Phoebus J., Irwin Friend, and N. Bulent Gultekin. "A critical reexamination of the empirical evidence on the arbitrage pricing theory." *The Journal of Finance* 39, no. 2 (1984): 323-346.

Dickens, William, and Lawrence F. Katz. "Interindustry wage differences and industry characteristics." (1986).

Elton, Edwin J., and Martin J. Gruber. "Modern portfolio theory, 1950 to date." *Journal of Banking & Finance* 21, no. 11-12 (1997): 1743-1759.

Elton, Edwin J., and Martin J. Gruber. "Estimating the dependence structure of share prices--implications for portfolio selection." *The Journal of Finance* 28, no. 5 (1973): 1203-1232.

Goetzmann, William N., Stephen J. Brown, Martin J. Gruber, and Edwin J. Elton. "Modern portfolio theory and investment analysis." *John Wiley & Sons* 237 (2014).

Fama, Eugene F. "Multiperiod consumption-investment decisions: A correction." *American Economic Review* 66, no. 4 (1976): 723-724.

Fama, Eugene F. "Stock returns, real activity, inflation, and money." *The American Economic Review* 71, no. 4 (1981): 545-565.

Fama, Eugene F. "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance* 25, no. 2 (1970): 383-417.

Fama, Eugene F. "Efficient capital markets: II." *The Journal of Finance* 46, no. 5 (1991): 1575-1617.

Fama, Eugene F. "Multifactor portfolio efficiency and multifactor asset pricing." *Journal of Financial and Quantitative Analysis* 31, no. 4 (1996): 441-465.

Fama, Eugene F., and Kenneth R. French. "Permanent and temporary components of stock prices." *Journal of Political Economy* 96, no. 2 (1988): 246-273.

Fama, Eugene F., and Kenneth R. French. "The cross-section of expected stock returns." *the Journal of Finance* 47, no. 2 (1992): 427-465.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33, no. 1 (1993): 3-56.

Fama, Eugene F., and Kenneth R. French. "Industry costs of equity." *Journal of Financial Economics* 43, no. 2 (1997): 153-193.

Fama, Eugene F., and Kenneth R. French. "Dissecting anomalies." *The Journal of Finance* 63, no. 4 (2008): 1653-1678.

Fama, Eugene F., and Kenneth R. French. "A five-factor asset pricing model." *Journal of Financial Economics* 116, no. 1 (2015): 1-22.

Fama, Eugene F., and James D. MacBeth. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81, no. 3 (1973): 607-636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. "Taming the factor zoo: A test of new factors." *The Journal of Finance* 75, no. 3 (2020): 1327-1370.

Friedman, Milton. *A theory of the consumption function*. Princeton University Press, 1957.

Friend, Irwin, and Marshall Blume. "Measurement of portfolio performance under uncertainty." *The American Economic Review* 60, no. 4 (1970): 561-575.

Gabaix, Xavier. "The granular origins of aggregate fluctuations." *Econometrica* 79, no. 3 (2011): 733-772.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet. "Time-varying risk premium in large cross-sectional equity data sets." *Econometrica* 84, no. 3 (2016): 985-1046.

Garleanu, Nicolae, Stavros Panageas, and Jianfeng Yu. "Technological growth and asset pricing." *The Journal of Finance* 67, no. 4 (2012): 1265-1292.

Gaspar, José-Miguel, and Massimo Massa. "Idiosyncratic volatility and product market competition." *The Journal of Business* 79, no. 6 (2006): 3125-3152.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. "Text as data." *Journal of Economic Literature* 57, no. 3 (2019): 535-74.

Ghysels, Eric. "On stable factor structures in the pricing of risk: do time-varying betas help or hurt?" *The Journal of Finance* 53, no. 2 (1998): 549-573.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu. "Thousands of alpha tests." *The Review of Financial Studies* 34, no. 7 (2021): 3456-3496.

Goyal, Amit, and Pedro Santa-Clara. "Idiosyncratic risk matters!" *The Journal of Finance* 58, no. 3 (2003): 975-1007.

Green, Jeremiah, John RM Hand, and X. Frank Zhang. "The characteristics that provide independent information about average US monthly stock returns." *The Review of Financial Studies* 30, no. 12 (2017): 4389-4436.

Griffin, John M., and G. Andrew Karolyi. "Another look at the role of the industrial structure of markets for international diversification strategies." *Journal of Financial Economics* 50, no. 3 (1998): 351-373.

Gu, Lifeng. "Product market competition, R&D investment, and stock returns." *Journal of Financial Economics* 119, no. 2 (2016): 441-455.

Hakansson, Nils H. "Capital growth and the mean-variance approach to portfolio selection." *Journal of Financial and Quantitative Analysis* 6, no. 1 (1971): 517-557.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. "… and the cross-section of expected returns." *The Review of Financial Studies* 29, no. 1 (2016): 5-68.

Harvey, Campbell R., and Akhtar Siddique. "Conditional skewness in asset pricing tests." *The Journal of Finance* 55, no. 3 (2000): 1263-1295.

Haugen, Robert A., and Nardin L. Baker. "Commonality in the determinants of expected stock returns." *Journal of Financial Economics* 41, no. 3 (1996): 401-439.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser. "Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988." *The Journal of Finance* 48, no. 1 (1993): 93-130.

Heston, Steven L., and K. Geert Rouwenhorst. "Does industrial structure explain the benefits of international diversification?" *Journal of Financial Economics* 36, no. 1 (1994): 3-27.

Hoberg, Gerard, and Gordon Phillips. "Real and financial industry booms and busts." *The Journal of Finance* 65, no. 1 (2010): 45-86.

Hoberg, Gerard, and Gordon Phillips. "Text-based network industries and endogenous product differentiation." *Journal of Political Economy* 124, no. 5 (2016): 1423-1465.

Hoberg, Gerard, and Gordon M. Phillips. "Text-based industry momentum." *Journal of Financial and Quantitative Analysis* 53, no. 6 (2018): 2355-2388.

Hoberg, Gerard, and Gordon M. Phillips. "Product market uniqueness, organizational form and stock market valuations." *Organizational Form and Stock Market Valuations (August 31, 2014)* (2014).

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. "What factors drive global stock returns?" *The Review of Financial Studies* 24, no. 8 (2011): 2527-2574.

Hou, Kewei, Chen Xue, and Lu Zhang. "Digesting anomalies: An investment approach." *The Review of Financial Studies* 28, no. 3 (2015): 650-705.

Hou, Kewei, and David T. Robinson. "Industry concentration and average stock returns." *The Journal of Finance* 61, no. 4 (2006): 1927-1956.

Jegadeesh, Narasimhan, and Sheridan Titman. "Returns to buying winners and selling losers: Implications for stock market efficiency." *The Journal of Finance* 48, no. 1 (1993): 65-91.

Jensen, Michael C., Fischer Black, and Myron S. Scholes. "The capital asset pricing model: Some empirical tests." (1972): 79.

Jurczenko, Emmanuel, ed. *Factor investing: From traditional to alternative risk premia*. Elsevier, 2017.

Jurczenko, Emmanuel, ed. *Machine Learning for Asset Management: New Developments and Financial Applications*. John Wiley & Sons, 2020.

Kaldor, Nicholas. "Marginal productivity and the macro-economic theories of distribution: Comment on Samuelson and Modigliani." *The Review of Economic Studies* 33, no. 4 (1966): 309-319.

Hakansson, Jivendra K. Kale Nils H., and Gerald W. Platt. "Industry vs. Other Factors in Risk Prediction." No. RPF-201. University of California at Berkeley, 1991.

Khalilzadeh-Shirazi, Javad. "Market structure and price-cost margins in United Kingdom manufacturing industries." *The Review of Economics and Statistics* (1974): 67-76.

King, Benjiamin F. "Industry factors in stock price behavior." *Journal of Business* 39, No.1(1996). 139-190.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. "Shrinking the cross-section." *Journal of Financial Economics* 135, no. 2 (2020): 271-292

Lessard, Donald R. "World, national, and industry factors in equity returns." *The Journal of Finance* 29, no. 2 (1974): 379-391.

Lettau, Martin, and Sydney Ludvigson. "Consumption, aggregate wealth, and expected stock returns." *The Journal of Finance* 56, no. 3 (2001): 815-849.

Levin, Richard C., Wesley M. Cohen, and David C. Mowery. "R & D appropriability, opportunity, and market structure: new evidence on some Schumpeterian hypotheses." *The American Economic Review* 75, no. 2 (1985): 20-24.

Li, Feng. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics* 45, no. 2-3 (2008): 221-247.

Li, Dongmei. "Financial constraints, R&D investment, and stock returns." *The Review of Financial Studies* 24, no. 9 (2011): 2974-3007.

Linnainmaa, Juhani T., and Michael R. Roberts. "The history of the cross-section of stock returns." *The Review of Financial Studies* 31, no. 7 (2018): 2606-2649.

Lintner, John. "Security prices, risk, and maximal gains from diversification." *The Journal of Finance* 20, no. 4 (1965): 587-615.

Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance* 66, no. 1 (2011): 35-65.

Lucas Jr, Robert E. "Asset prices in an exchange economy." *Econometrica: Journal of the Econometric Society* (1978): 1429-1445.

Harry, M. "Markowitz. Portfolio Selection." *Journal of Finance*, 7, no. l." (1952).

McLean, R. David, and Jeffrey Pontiff. "Does academic research destroy stock return predictability?" *The Journal of Finance* 71, no. 1 (2016): 5-32.

Merton, Robert C. "An intertemporal capital asset pricing model." *Econometrica: Journal of the Econometric Society* (1973): 867-887.

Meyers, Stephen L. "A re-examination of market and industry factors in stock price behavior." *The Journal of Finance* 28, no. 3 (1973): 695-705.

Modigliani, Franco, and Merton H. Miller. "The cost of capital, corporation finance and the theory of investment." *The American Economic Review* 48, no. 3 (1958): 261-297.

Moskowitz, Tobias J., and Mark Grinblatt. "Do industries explain momentum?" *The Journal of Finance* 54, no. 4 (1999): 1249-1290.

Mossin, Jan. "Equilibrium in a capital asset market." *Econometrica: Journal of the Econometric Society* (1966): 768-783.

Muth, John F. "Rational expectations and the theory of price movements." *Econometrica: Journal of the Econometric Society* (1961): 315-335.

Newman, Mark. *Networks*. Oxford University Press, 2018.

Nickell, Stephen J. "Competition and corporate performance." *Journal of Political Economy* 104, no. 4 (1996): 724-746.

Novy-Marx, Robert. "The other side of value: The gross profitability premium." *Journal of Financial Economics* 108, no. 1 (2013): 1-28.

Pástor, Ľuboš, and Robert F. Stambaugh. "Liquidity risk and expected stock returns." *Journal of Political Economy* 111, no. 3 (2003): 642-685.

Pástor, Ľuboš, and Pietro Veronesi. "Technological revolutions and stock prices." *American Economic Review* 99, no. 4 (2009): 1451-83.

Pellegrino, Bruno. "Product differentiation and oligopoly: a network approach." *WRDS Research Paper* (2019).

Peress, Joel. "Product market competition, insider trading, and stock market efficiency." *The Journal of Finance* 65, no. 1 (2010): 1-43.

Porter, Michael E. "The structure within industries and companies' performance." *The Review of Economics and Statistics* (1979): 214-227.

Raith, Michael. "Competition, risk, and managerial incentives." *American Economic Review* 93, no. 4 (2003): 1425-1436.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. "International stock return predictability: What is the role of the United States?" *The Journal of Finance* 68, no. 4 (2013): 1633-1662.

Regnault, Jules. "Calcul des chances et philosophie de la bourse." Pilloy, 1863.

Reinganum, Marc R. "Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values." *Journal of Financial Economics* 9, no. 1 (1981): 19-46.

Reinganum, Jennifer F. "Innovation and industry evolution." *The Quarterly Journal of Economics* 100, no. 1 (1985): 81-99.

Roll, Richard. "Industrial structure and the comparative behavior of international stock market indices." The *Journal of Finance* 47, no. 1 (1992): 3-41.

Roll, Richard, and Stephen A. Ross. "An empirical investigation of the arbitrage pricing theory." *The Journal of Finance* 35, no. 5 (1980): 1073-1103.

Roll, Richard, and Stephen A. Ross. "On the cross-sectional relation between expected returns and betas." *The Journal of Finance* 49, no. 1 (1994): 101-121.

Rosenberg, Barr. "Extra-market components of covariance in security returns." *Journal of Financial and Quantitative Analysis* 9, no. 2 (1974): 263-274.

Ross, Stephen A. "The arbitrage theory of capital asset pricing." *Journal of Economic Theory* 13 (1976): 341-360.

Rubinstein, Mark. "The valuation of uncertain income streams and the pricing of options." *The Bell Journal of Economics* (1976): 407-425.

Schmalensee, Richard. "Do markets differ much?" *The American Economic Review* 75, no. 3 (1985): 341-351.

Schmidt, Klaus M. "Managerial incentives and product market competition." *The Review of Economic Studies* 64, no. 2 (1997): 191-213.

Sebastiani, Fabrizio. "Machine learning in automated text categorization." *ACM computing surveys (CSUR)* 34, no. 1 (2002): 1-47.

Shanken, Jay. "On the estimation of beta-pricing models." *The Review of Financial Studies* 5, no. 1 (1992): 1-33.

Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk." *The Journal of Finance* 19, no. 3 (1964): 425-442.

Sloan, Richard G. "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *Accounting Review* (1996): 289-315.

Smirlock, Michael, Thomas Gilligan, and William Marshall. "Tobin's q and the Structure-Performance Relationship." *The American Economic Review* 74, no. 5 (1984): 1051-1060.

Stigler, George J. "A theory of oligopoly." *Journal of Political Economy* 72, no. 1 (1964): 44-61.

Stiglitz, Joseph E. "A consumption-oriented theory of the demand for financial assets and the term structure of interest rates." *The Review of Economic Studies* 37, no. 3 (1970): 321-351.

Syverson, Chad. "Macroeconomics and market power: Context, implications, and open questions." *Journal of Economic Perspectives* 33, no. 3 (2019): 23-43.

Tetlock, Paul C. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance* 62, no. 3 (2007): 1139-1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. "More than words: Quantifying language to measure firms' fundamentals." *The Journal of Finance* 63, no. 3 (2008): 1437-1467.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B* (Methodological) 58, no. 1 (1996): 267-288.

Titman, Sheridan, KC John Wei, and Feixue Xie. "Capital investments and stock returns." *Journal of Financial and Quantitative Analysis* 39, no. 4 (2004): 677-700.

Treynor, Jack L. "Toward a theory of market value of risky assets." Available at SSRN 628187 (1962).

Valta, Philip. "Competition and the cost of debt." *Journal of Financial Economics* 105, no. 3 (2012): 661-682.

Vassalou, Maria. "News related to future GDP growth as a risk factor in equity returns." *Journal of Financial Economics* 68, no. 1 (2003): 47-73.

Welch, Ivo. "Simpler, better market betas." *National Bureau of Economic Research* (2019).

Welch, Ivo, and Amit Goyal. "A comprehensive look at the empirical performance of equity premium prediction." *The Review of Financial Studies* 21, no. 4 (2008): 1455-1508.

Wernerfelt, Birger, and Cynthia A. Montgomery. "Tobin's q and the importance of focus in firm performance." *The American Economic Review* (1988): 246-250.

Zou, Hui, and Trevor Hastie. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: series B* (statistical methodology) 67, no. 2 (2005): 301-320.