A Work Project presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

A Market-Based Approach to Identify Conglomerates

Steinar Skagestad

Work project carried out under the supervision of:

Fernando Anjos

22-05-2020

Abstract

The paper takes on a marked-based approach and creates a model to classify which industries a particular company participates in. The marked-based classification is compared with company reported accounting-based segments to look for discrepancies. My main result is that there is a significant discrepancy. Finally, the model is applied to try to predict reported entrances in new industries and to study whether the diversification discount changes when the marked-based approach is used. It is not possible to infer anything specific from the prediction analysis, but the diversification discount decreases when the market-based approach is applied.

Keywords: Corporate diversification, Industry reporting, Efficient Market Hypothesis, Conglomerates

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Contents

I. Introduction & Background

According to the Financial Accounting Standards Boards (FASB), investor surveys regarding GAAP conveys that segment reporting is an area in need of improvement. "*Investors normally model a company at the segment level rather than at the consolidated level*" (Buesser, 2019). Moreover, this information is compared to different industry benchmarks, which can have a direct impact on investment decisions. Furthermore, historical segment reporting is commonly employed in conglomerate research and is frequently used to investigate or try to explain the conglomerate discount. One example is Schneider and Spalt's (2016) research on how the CEO long shot bias is more present in "smaller segments". This is just a few of many examples on how accurate segment reporting is valuable, and how improper segment reporting could have severe consequences to investment decisions and research.

As the world have developed away from a simple agriculture- and manufacturing- based economy, segment reporting has become increasingly complex. Hence, it is has become relevant to investigate, and question the accuracy of segment reporting by companies. This lays the foundation for the study done in this paper.

II. Literature Review

This paper aims to explore how accurate segments are reported from a market point of view. Segment reporting in GAAP is done by companies and can easily be subjective. Consequently, it is relevant to ask,

"Are there any discrepancies between how markets seemingly classify industries that companies operate in and how companies classify what industries they operate in themselves?"

Here, the market is defined as the financial markets and its behavior, as stocks in companies and industries are bought and sold. Eugene Fama's studies about the efficient market hypotheses declare that markets are very effective allocators of public information and can process it very well (Fama, 1970). Given that this is true, markets should also be able to process what industries specific companies are invested in, which again will be reflected in its valuations and prices. To illustrate; Any company that have invested in selling crude oil will, to some degree, depend on prices in the oil market and then also movements in the oil industry. Consequently, it is compelling to try and proxy the market's perspective on which industries companies operate in and compare the results with the industries companies report that they belong to.

When a model that allocates industries to companies is developed, it can be applied numerous ways. As stated above, the efficient market hypothesis argues that markets process public information very efficiently. One implication is that markets should be able recognize if a company enters into a new industry at an early stage. Investments into new industries can often be announced ahead of time, and if markets believe that a company is committed to enter, it should also influence the value of the company. As perspectives on value are altered, price and return should respond. However, segment reporting is not done continuously. Thus, lag effects may occur, as industries are reported after a companies have initiated their operations in a particular industry. Consequently, the model will be applied to examine whether it is able to predict future reported industries for companies.

Industry classification is also important for research on corporate diversification and conglomerates. There is an extensive amount of research on this topic, and the "conglomerate discount" is a subject of particular interest (Maksimovic and Phillips, 2013). Traditionally, This research has been driven by the desire to explore the efficiency of internal capital markets. Maksimovic and Phillips (2013) define the conglomerate discount as "*when the market value of the diversified firm is lower than a comparable portfolio of single-segment firms*". One of the more prominent papers on the topic, written by Lang & Stulz (1994), tries to prove this phenomenon by showing that Tobin's Q is negatively related to company diversification.

However, more recent literature has started to dispute the presence of a conglomerate discount, as some argue that a conglomerate premium is more likely to appear. One example is research published by Hund, Monk and Tice (2012). Here, data from COMPUSTAT and CRSP is used to argue that conglomerates trades at a premium when value weights, in opposition equal weights, is used to value companies. Additionally, Anjos and Fracassi (2015) explore how conglomerates with a central position and connection to many companies, can create value through informational advantages.

Despite these conflicting views, the conglomerate discount is a well-established term in relation to research on corporate diversification. Yet, industry classification is an important part of the conglomerate discount equation. If company reported industry classifications are skewed somehow, this can create a bias when the conglomerate discount is calculated. Hence, this paper is also aiming to use the market-based industry classification approach to look at and compare how a conglomerate discount(or premium) potentially can differ from when company reported industries are employed.

To summarize and hypothesize, the research project will be split into 3 parts: First, the paper design a market-based industry classification model. The output from this model will be compared to reported industries by companies to look for potential discrepancies. This is particularly relevant as it can indicate that inaccurate reporting potentially creates bias in other fields. A field of particular interest here is conglomerate research. Next the model will be applied to perform two different analyses. One to predict future reported industry entrances, and one to evaluate if the conglomerate discount also is present when the marked-based approach allocates industries to companies. As the market tends to be a very efficient communicator and processor of information, one can hypothesize that a market based model will be able to predict future reported entrances into new industries.

III. Methodology

3.1 Setting up the Model:

To investigate discrepancies about segment classifications between the market approach and company reporting, data is gathered from different sources. Wharton Research Data Services is used to access most of the data. First, information regarding reported operating segments by companies are collected from COMPUSTAT historical segment files. This database is again filtered by business segment, which returns "*An industry segment or product line reported by a company*" (WRDS, 2020).

Next, this company data is matched with and categorized by its appropriate Kenneth R. French 49 industry based on its respective SIC codes. To perform this study, it is necessary to create a universe of different industries. Professor Kenneth R. French have categorized companies into different industries based on their SIC codes. This information is publicly available on his personal website, and will be used to classify companies in different industries (French, 2020). According to U.S. Securities and Exchange Commission, SIC codes that appear in company filings, "*indicate the company's type of business*"(SEC, 2020). When companies are matched and put into its industries for a specific year, a number of "company reported industries" can be observed. Here, every observation can be viewed as "$R_{i,t}$". The two subscripts represents industry(i) and year(t), while "R" represents a specific company.

After company reported industries are gathered and properly categorized, market-based information is collected. First, daily company returns are withdrawn from the Center for Research in Security prices' (CRSP) database. In addition, daily industry returns for the 49 industry portfolios are gathered from Kenneth R. French's website.

This study is based on data of S&P 500 companies as of April 10th 2020, and is observed over a timespan from 1990 through 2018. Data from COMPUSTAT segment files is self-reported and about 70% of North American companies use this platform to report segment information . Consequently, segment data about some S&P 500 companies is missing (one example is Bank of America). Thus, industry reported data for 470 of the S&P 500 companies is obtained. Moreover, some of the companies do not have returns for the entire period of study, resulting in less observations in certain years. This can potentially have an effect on the significance of the results for certain parts of the outcome of the analysis.

When the data is sorted and tidied it is possible to create a model that portrays the markets perspective on which of the 49 industries that the S&P500 companies are invested in. The model is designed to measure the industries on an annual basis for every company, as companies also report which industries they operate in once a year to COMPUSTAT. This is done by running a linear multiple regression of a company's daily return against the daily returns of every industry in the 49-industry portfolio for the calendar year. Namely, daily industry returns over a year will serve as independent variables against company returns which is the dependent variable. This process returns several observations. One can interpret any of those observations as "xy" where x is a specific company x and y is a specific year. The entire output from the regression model can be envisioned as 3-dimensional matrix with 49 coefficients for every company year (xy), where each of the 49 coefficients tries to explain daily marginal changes in company stock prices based on changes in respective industry returns for a given year. Thus, the equation for the model will be:

$$xy = b_0 + b_1 industry_1 + b_2 industry_2 \ldots. + b_{49} industry_{49} + \varepsilon \quad (1)$$

When the regressions are calculated, the model determines which industries companies are associated with by identifying the largest coefficients from the regression. These coefficients are industries that have greater impact or weight on the returns of the company. Consequently, the largest coefficients represent the industries the model estimates that a company is more likely to operate in.

Initially, the model will assume that every company operate in either one or two industries. Hence, for every company year xy, the two highest coefficients of every regression are filtered out and isolated. These two coefficients are assumed to represent the industries the market believe a company is invested in.

Subsequently, the industries of the two coefficients that are filtered out will be compared to COMPUSTAT reported industries for a company for the same year($R_{i,t}$). It is natural to assume that many companies only operate in one industry. Hence, if a company only reports that it is part of one specific industry for a given year, the second highest coefficient from the market regression model is also removed.

Lastly, the industries the model assigns to companies are compared to $R_{i,t}$. This is done for every company year within the Kenneth French 49 industry portfolio universe. Then, it is possible to examine discrepancies in industry classifications from the two different approaches.

3.2 Predicting Reported Entrances Into New Industries:

After the model is set up and examined, further analysis can be done. Initially, the model will be used to see if it can predict reported entrances into new industries by companies, which is done the following way: First, every COMPUSTAT reported entrance into a new industry (given that the company already operates in another industry) is located. This will give a number of different observations, $R_{i,t}$, where for every observation, industry i is a new industry for the company even though there are other observations of the same company where i is different and where t is smaller.

Next, we use the regression model as a predictor. For every $R_{i,t}$ that is detected, the market approach is used to obtain the two industry coefficients that represents that particular industry i but where t = t-1 and t-2. This is the relative weights of those industries for a company the year before the reported entrance and two years before the reported entrance. When the coefficients are calculated, we subtract the t-2 coefficient from the t-1 coefficient, which will return the change in the industry coefficient over that given period. If the change in the coefficient is >0, the industry has gained a stronger marginal influence on the returns for the given company relatively to other industries. This indicates that the new industry has a stronger influence on the company in year t-1 compared to t-2. If the change is significant and relatively different than the change of other industry weights for the same period, it could very well communicate that markets has started to believe in the company's commitment to invest in the new industry. Hence, one can infer that the models work as a predictor for reported company entrances in new industries in year t.

3. 3 Analysis of the Diversification Discount:

Finally, the regression model will be used to examine whether the diversification discount is different when companies' industries are decided by the model, in contrast to the classical COMPUSTAT industry classifications. To perform this analysis, a valuation measure is needed to compare the value of diversified firms to non-diversified firms. Here, Tobin's Q will be used. Tobin's Q measures the ratio between firm value of a company and its assets replacement costs, and is defined as:

$$\text{Tobin's Q} = \text{Enterprise Value of Firm} / \text{Total assets of firm} \tag{2}$$

COMPUSTAT is used to gather company-specific data on enterprise value and total assets. Subsequently, Tobin's Q is measured for every company year available for the S&P 500 companies. Afterwards, the diversification discount is computed with the regular COMPUSTAT industry classification. Here, all companies are assembled in a data frame for every year they are present. This creates a set of observations. One specific observation can be named $Q_{r,t}$ where r refers to a specific company, while t is a specific year. Furthermore, 50 dummies are added for all observations ($Q_{r,t}$). 49 of the dummies represents each of the 49 portfolio industries and will equal 1 if COMPUSTAT reports that the company is part of a specific industry. There is also a dummy called "diversified" that is added to the model. This dummy will equal 1 for any $Q_{r,t}$ that has more than one industry dummy equal to 1. In other words, a company is diversified if it is part of more than one industry.

When the data is tidied and set up properly one can estimate whether there is a diversification discount or not. This is done through a pooled regression of all observations, where Tobin's Q serves as the dependent variable and all the 50 dummies serves as independent variables. Hence, the equation for this regression become:

$$\text{Tobin's Q} = b_0 + b_1\text{Diversified} + B_2\text{Industry}_1 + \ldots B_{50}\text{Industry}_{49} + \varepsilon \qquad (3)$$

b_1, the coefficient for the diversified dummy will be the most important variable. The number this coefficient returns is going to display whether there is a diversification discount or diversification premium present. A negative coefficient indicates that there is a diversification discount present as Tobin's Q will be lower for diversified firms. A positive coefficient will on the other hand indicate that there is a diversification premium present.

The methodology will almost be identical when the diversification discount is estimated with the market-based approach. The main goal is still to estimate the discount through a diversification dummy where Tobin's Q is the dependent (valuation) variable. However, this time the model disregards the idea of a 1-2 industry universe. Instead, if the coefficient that represent the respective industry has a higher weight than a certain cutoff limit, an industry will be present for that company year. As it is not totally clear what this limit should be, the cutoff is based on average reported industries/firm. Thus, both the market-based approach and the model that is based on COMPUSTAT data will have the same number of industries present. Yet, these industries will potentially be distributed differently among the companies. The pooled regression of Tobin's Q on the 50 different dummies is calculated using the marked-based approach when the steps above is done. Additionally, company- and year fixed effects will be added to the regression model to control for omitted variable bias. The results should provide evidence of a potential diversification discount or diversification premium.

Lastly, one can compare the results for the two different approaches. If industries are distributed differently for the market-based approach one should expect a certain difference in the results. However, it is interesting to examine if it is sizeable or not as this can reveal signs of a bias in the "classical" conglomerate discount.

IV. Presentation of Results

4.1 The Model:

The market based approach produced an output of 11524 regressions of company year observations for the S&P 500 companies between 1990 and 2018.

Figure 1 Top 10 Industries according to Market Model

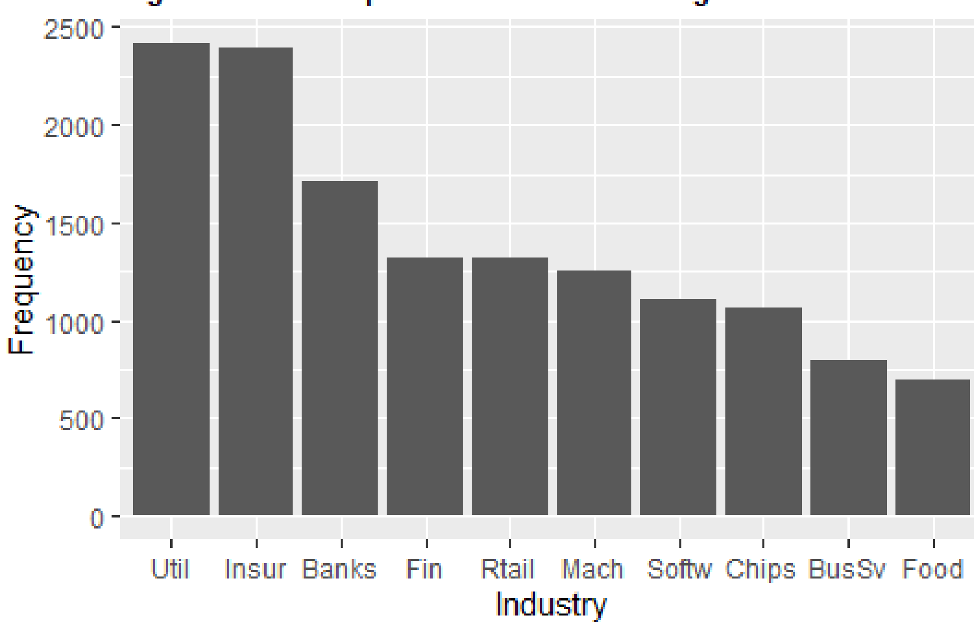
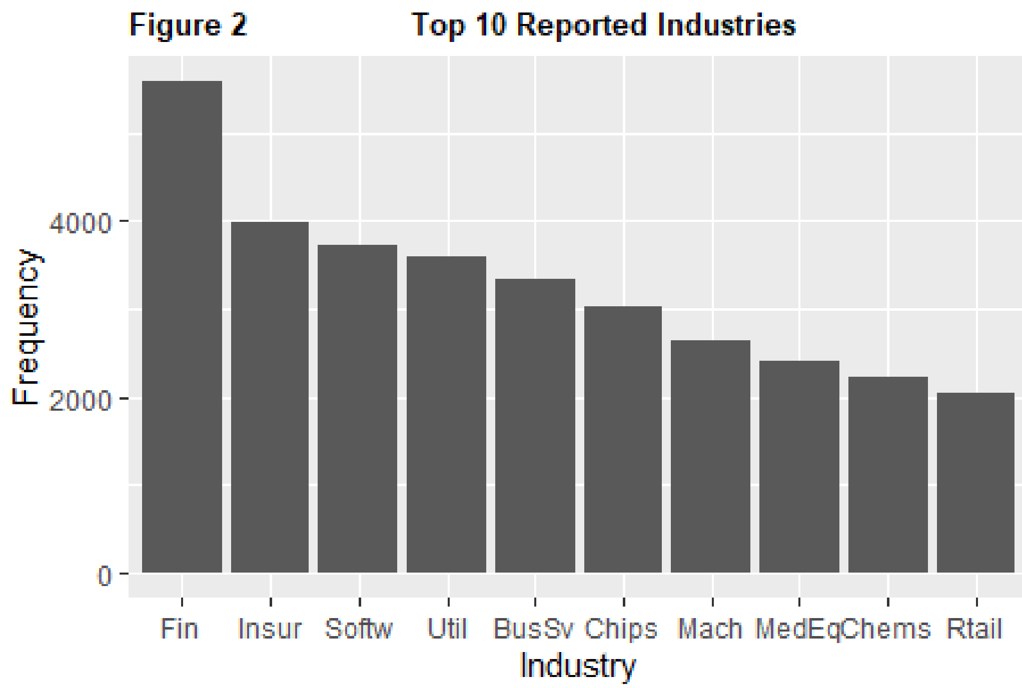


Figure 1 visualizes the ten most frequent industries for company years between 1990 and 2018 for the S&P500 companies, according to the estimations made by the model. Utilities, insurance, and banks are the 3 industries that occurs most frequently, as shown in the figure.

Figure 2



Source: Data from WRDS

Figure 2 on the other hand portrays the ten most frequent industries that are reported by the companies for company years between 1990 and 2010 in COMPUSTAT. Here, the trading, insurance and software industry occur most frequently. In this case, banking is not even among the top ten industries. However, insurance and banking are industries with many common properties, which could be the reason why it appeared so frequently for the model in opposition to COMPUSTAT. In sum, the two industry classification methods seem to share a lot of common traits regarding commonly observed industries.

Table 1 Matching of Industries

	Industry 1	**Industry 2**
TRUE	3711	1110
FALSE	5172	6404
Missing Data	2641	4010

Table 1 summarizes the results of the matching process between COMPUSTAT reported segment data and the industries given by the market-based approach. “TRUE” represents instances where industries match for the two different industry classifications methods for the same company year. “FALSE”, on the other hand, presents the number of instances where no match is found . As shown, there is more observations that returns a “FALSE” result than a “TRUE” result. This result suggests that there are discrepancies between the market and companies in terms of perspectives on which industries companies are operating in. Furthermore, the table shows that there are 1369(4010-2641) additional instances of missing data for industry 2. This is due to some companies reporting that they only are part of one industry, and it occurs in 11,8% of the observed company years. The 2641 missing observations in column 1 are instances where a company does not report to be part of any industry for a given year. If this happens, a comparison with the market based-approach is meaningless, which is reflected in the matching results. Additionally, the distribution of industries into “Industry 1” and “Industry 2” is based on the largest and second largest coefficient that is given form the marked-based approach.

There is a clear difference between the results for the two industries. This is expected. As "Industry 1" is the industry with the highest coefficient, it is natural that it matches with reported industries more frequently. The market-based approach agrees with the reported industries in about 41,8 % of the instances for the main industry, while 14,8% of cases are equal for industry 2. As there are a total of 49 different industries for every company year, over 40% matching is a significant result. Nevertheless, there is a huge gap between reported industries and the industries companies are connected to according to the market. Thus, it is reasonable to argue that the answer to the first question asked is that there are discrepancies between market measured industries and reported industries for companies.

This paper does not try answer why these discrepancies exist, it merely points out that there are differences. Yet, as these discrepancies seems to be present, several new questions arise, and further new analysis can be done. If industry reporting by companies is inaccurate, bias can be present at many different levels. Moving on, this paper will try to analyze some of these questions.

4.2 Predicting Reported Entrances into New Industries:

The next step of this paper is to look closer at the results from the prediction analysis, which tried to analyze whether the market-based approach could predict future COMPUSTAT reporting in new industries for companies. There were 367 instances of companies that entered new industries under these conditions for the time period. It makes sense that this number is relatively low as companies as most companies tend to keep their business within their business line. However, some companies, like for example Berkshire Hathaway (which reported that it entered a new industry 8 times between 1990 and 2018) seems to drive this number up.

Table 2 Descriptive Statistics

1st Quantile	Median	Mean	3rd Quantile
-0.25571	-0.02954	-0.02626	0.22999

Table 2 delivers descriptive statistics about the change in the relevant industry coefficient from two to one years before a company reports to enter a new industry. The change in the variable is close to zero and somewhat surprisingly also slightly negative, both in terms of mean and median numbers. This result contradicts the hypothesis and expectation that the change in coefficients would be positive. A positive change was expected as one would assume the market to start treating companies more like the industry they are about to enter ahead of time, which would resulting in a higher weight in the respective variable connected to that industry. At the same time it is important to note that even though the average change in the coefficients are slightly negative, they are very close to zero.

There may be several reasons why the change described above is close to zero rather than positive. It is likely that many of the entrances into new industries are of smaller proportion. Hence, new industry entrances may, in many cases, not significantly affect the value of a company. Another potential reason may be lagged reporting of entrances into new industries. This may cancel out the expected change in coefficients leading up to the entrance in the industry. Markets may also react differently on the entrance into new industries on a company-to-company basis. Some companies may show and reflect that they are committed to invest in new industries at a very early stage. Other companies will in contrast not convince markets that they are committed to stay and operate in a new industry. All of these examples can create “noise” in the prediction analysis, which may explain the results that were found.

4.3 Analysis of the Diversification Discount:

Table 3 Descriptive Statistics

Total industry – company - year	Number of Industries Reported	Mean industry/company year
8122	10791	1.32862

Finally, we take a closer look at how the diversification discount reacts when the market-based approach is used to allocate industries for companies in contrast to reporting by companies from COMPUSTAT. Summary statistics is presented in table 3 above. There is a total of 8122 observations where Tobin’s Q is present. For these observations, companies reported to be part of 10791 industries. Hence, there is an average of 1.33 industries per company year in the dataset for the entire period. Additionally, 30.9 % of the company years that are gathered from COMPUSTAT are categorized as “diversified”. This is slightly more than the estimations made by the market-based approach, where 25.4% of the company years are estimated to be diversified.

Table 4 Conglomerate Discount Results

	Tobin's Q			
	Reported Industries	Market-Approach	Reported Industries	Market-Approach
	(1)	(2)	(3)	(4)
diversified	-0.927***	-0.227**	-0.911***	-0.253***
	(0.102)	(0.097)	(0.101)	(0.097)
Agric	0.319	0.409	0.375	0.402
	(1.693)	(0.420)	(1.676)	(0.417)
Food	-0.096	0.746***	-0.200	0.705***
	(0.193)	(0.145)	(0.191)	(0.144)
Soda	1.755***	-0.515	1.729***	-0.475
	(0.276)	(0.406)	(0.273)	(0.403)
Beer	-0.110	-0.659**	-0.255	-0.667**
	(0.470)	(0.310)	(0.465)	(0.307)
Smoke	1.422***	2.502***	1.530***	2.513***
	(0.455)	(0.414)	(0.450)	(0.411)
Toys	0.071	4.510***	0.119	4.614***
	(0.457)	(0.359)	(0.452)	(0.357)
Fun	0.202	-0.455**	0.215	-0.485**
	(0.255)	(0.212)	(0.253)	(0.210)
Books	0.339	-1.516***	0.327	-1.509***
	(0.357)	(0.281)	(0.353)	(0.279)
Hshld	0.518***	0.719***	0.472***	0.749***
	(0.184)	(0.187)	(0.182)	(0.187)
Clths	0.510*	-0.819***	0.523**	-0.762***
	(0.264)	(0.205)	(0.262)	(0.204)
Whlsl	-0.117	0.056	-0.151	0.066
	(0.133)	(0.155)	(0.131)	(0.154)
Rtail	0.662***	0.409***	0.620***	0.433***
	(0.118)	(0.100)	(0.117)	(0.100)
Meals	0.587**	0.054	0.557**	0.047
	(0.242)	(0.189)	(0.240)	(0.188)
Banks	-0.078	0.385***	-0.099	0.210**
	(0.144)	(0.100)	(0.143)	(0.104)
Insur	-0.413**	0.019	-0.495***	-0.051
	(0.164)	(0.095)	(0.163)	(0.095)
RlEst	0.105	-1.585***	0.075	-1.634***
	(0.270)	(0.473)	(0.267)	(0.469)
Fin	-0.136	0.142	-0.208**	0.086
	(0.106)	(0.125)	(0.105)	(0.124)
Other	0.078	-0.794***	-0.015	-0.585**
	(0.221)	(0.280)	(0.219)	(0.278)
Constant	2.261***	2.121***		
	(0.047)	(0.034)		
Observations	8,122	8,122	8,122	8,122
R^2	0.095	0.123	0.100	0.123
Adjusted R^2	0.090	0.117	0.092	0.115
Residual Std. Error (df = 8071)	2.377	2.340		
F Statistic	17.007*** (df = 50; 8071)	22.581*** (df = 50; 8071)	17.828*** (df = 50; 8051)	22.512*** (df = 50; 8051)
Note:				*p<0.1; **p<0.05; ***p<0.01

Data From WRDS

Table 4 gives a snapshot of the most important parts of the pooled regressions of Tobin's Q for all $Q_{r,t}$. Only 18/49 industry variables are included here due to size of the entire regression table. Let us first focus on the OLS regressions, namely regression (1) and (2). The topmost estimator, "diversified", which represents whether firms are diversified or not, is the most important variable. The table reveals that a diversification discount is present for both industry classifications. Value seems to deteriorate when a firm diversify its operations, as Tobin's Q decreases. However, there are visible differences with regards to the magnitude of the diversification discount. While the reported industry perspective provides evidence of a rather strong diversification discount (Tobin's Q decreases with 0.927 when a firm is diversified), the market-based approach predicts a decrease that is a lot less severe (a decrease in Tobin's Q of 0.227) . Moreover, while the diversification discount clearly is significant for the reporting method($p<0.01$), the result is a little bit more doubtful for the market-based approach. Additionally, the fit of the marked based regression is higher. This indicates that the marked based approach generally is better to capture variations in Tobin's Q. Fixed effects are also added to the two approaches in regression (3) and (4). If we compare these results to regression (1) and (2), the diversification discount converge a little bit. However, the result still seems the same as before.

Even though both models compute that corporate diversification tends to be less valuable, there is a large gap between the results from the two approaches, which adds an interesting perspective to research on the diversification discount. The results from this analysis indicates that conglomerate discount is exaggerated. If this is true and one assumes that the market-base approach portrays a more accurate view on how industries are allocated, there is a significant bias present in conglomerate research. As stated earlier, COMPUSTAT is frequently used to gather industry data and study corporate diversification. If reporting done by companies is inaccurate, it will spread to research that use this data.

While the results from the model and analysis done above is thought provoking, it is also important to be aware of the limitations and caveats that might be present. The first limitation is the scope of the data. The analysis is done with daily return data on S&P 500 as of early April 2020, which means that the it is limited to this "universe". Furthermore, some company years does not have return data for a full year. The consequence is that these company years gives might not be as accurate(this would only be true for a small subset of the data).

Another limitation is that the model does not have a very specific method to decide how many industries companies should be a part of. The paper approaches this issue two different and mutually exclusive ways in the first and last part of the analysis (as the goal of the two parts was somewhat different). When the model is searching for industry related discrepancies, it is limited to a one or two industries per company year. This might not necessarily be true for all companies, as some of them are operating in more than two industries. In the latter part of the analysis on the other hand, industries is included based on a cutoff for the industry related regression coefficients (final cutoff was 0.6). Consequently, there is a few company years that ended up not being connected to any industry at all as all of the industry coefficients was below the cutoff. Even though this only happened for a few companies, it is imprecise as all of the companies should be connected to at least one industry. One should also mention that thousands of regressions are performed for this analysis which creates certain statistical challenges. Multicollinearity is probable to occur between the different variables in many different cases which again may affect the variance of the coefficients, and limit the reliance of the model. Tests for multicollinearity have been performed and it have been present in some cases. However, it is problematic to remove industry coefficients for the sake of the analysis.

Another important limitation of this paper is that the 49 industry portfolios acquired from professor Kenneth French's website uses an accounting-based approach. Companies report which industries they are part of and then this data is assembled into industry portfolios. This could potentially make the study more inaccurate as this potential bias is integrated in the independent variables of the model. Nevertheless, the model still stands to reason under the assumption that most firms accurately reports their segments. Hence, the industry portfolios as a whole should deliver sensible information for this study. Yet for future reference, this is an issue that could be looked closer at. One way to possibly explore this is to examine whether one can create other marked based indexes for every single industry that can serve as independent variables.

Additional topics could possibly also be explored in future research. The prediction analysis brought up several questions for future research. First off, this paper did not answer why there are discrepancies between the market-based approach and the reported industries. Given that the market-based approach is accurate, do companies have any motives to report that they are part of certain industries? Furthermore, it is also possible to explore the prediction analysis closer. A natural take would be to try to do the prediction analysis over a different, and probably longer, timespan. Comparing "t-1" coefficients with t-5 or t-3 coefficients could for example be interesting to look at.

V. Conclusion

This paper has explored industry- and conglomerate related subjects. First and foremost, the paper has investigated, and tried to answer whether there are any discrepancies between how markets seem to classify industries and how they are reported by companies themselves. By creating a multiple OLS model based on industry- and company specific returns, industries have been allocated to companies and compared with reported industry classification to look for differences. The model's calculated industries matched with reported industries in about 40% of the cases after they were categorized based on Kenneth French's 49- industry portfolio. The result shows that there are clear similarities, but nevertheless also discrepancies in terms of industry classification between the two methods.

Next, the market-based industry classification approach have been used to complete two additional analyses. First the model was applied to investigate if it could predict future reported entrances into new industries for companies. This was done by looking at the change in relevant industry coefficients prior to the reported entrance. The result did not give any reason to believe that the model was able to make these predictions. Ultimately it may not be that surprising due to different kind of noise related to these entrances. Lastly, the paper estimated the conglomerate discount (or premium) by using a marked-based industry classification approach, and compared it to results from reported industries. Both methods found a conglomerate discount. However, the discount was noticeably smaller when the model was used to decide industries for relevant companies. These results are thought provoking as they indicate that there can be a bias present when conglomerate discounts are calculated.

Bibliography

Anjos, Fernando, and Cesare Fracassi. 2015. "Shopping For Information? Diversification And The Network Of Industries". *Management Science* 61 (1): 161-183. doi:10.1287/mnsc.2014.2060.

Buesser, Gary. 2019. "For The Investor: Segment Reporting". *Fasb.Org*. https://www.fasb.org/cs/ContentServer?c=Page&cid=1176172449742&d=&pagename=FASB%2FPage%2FSectionPage.

"Division Of Corporation Finance SIC Code List". 2020. *Sec.Gov*. https://www.sec.gov/info/edgar/siccodes.htm.

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review Of Theory And Empirical Work". *The Journal Of Finance* 25 (2): 416. doi:10.2307/2325486.

French, Kenneth R. 2020. "Kenneth R. French - Detail For 49 Industry Portfolios". *Mba.Tuck.Dartmouth.Edu*. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html

Hund, John, Donald Monk, and Sheri Tice. 2012. "Apples To Apples: The Economic Benefit Of Corporate Diversification". *SSRN Electronic Journal*. doi:10.2139/ssrn.2023786.

Lang, Larry H. P., and René M. Stulz. 1994. "Tobin's Q, Corporate Diversification, And Firm Performance". *Journal Of Political Economy* Vol. 102 (No. 6): 1251.

Maksimovic, Vojislav, and Gordon M. Phillips. 2013. "Conglomerate Firms, Internal Capital Markets, And The Theory Of The Firm". *Annual Review Of Financial Economics* 5 (1): 226-229. doi:10.1146/annurev-financial-110112-120933.

Schneider, Christoph, and Oliver Spalt. 2016. "Conglomerate Investment, Skewness, And The CEO Long-Shot Bias."

"WRDS Overview Of Compustat Segments Data". 2020. *Wrds-Wharton.Upenn.Edu*. Accessed May 1. https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/compustat/segments/wrds-overview-compustat-segments-data/.

APPENDIX:

Diversification discount regression, with all variables included:

Table 4 Conglomerate Discount Results

	Tobin's Q			
	Reported Industries	Market-Approach	Reported Industries	Market-Approach
	(1)	(2)	(3)	(4)
diversified	-0.927***	-0.227**	-0.911***	-0.253***
	(0.102)	(0.097)	(0.101)	(0.097)
Agric	0.319	0.409	0.375	0.402
	(1.693)	(0.420)	(1.676)	(0.417)
Food	-0.096	0.746***	-0.200	0.705***
	(0.193)	(0.145)	(0.191)	(0.144)
Soda	1.755***	-0.515	1.729***	-0.475
	(0.276)	(0.406)	(0.273)	(0.403)
Beer	-0.110	-0.659**	-0.255	-0.667**
	(0.470)	(0.310)	(0.465)	(0.307)
Smoke	1.422***	2.502***	1.530***	2.513***
	(0.455)	(0.414)	(0.450)	(0.411)
Toys	0.071	4.510***	0.119	4.614***
	(0.457)	(0.359)	(0.452)	(0.357)
Fun	0.202	-0.455**	0.215	-0.485**
	(0.255)	(0.212)	(0.253)	(0.210)
Books	0.339	-1.516***	0.327	-1.509***
	(0.357)	(0.281)	(0.353)	(0.279)
Hshld	0.518***	0.719***	0.472***	0.749***
	(0.184)	(0.187)	(0.182)	(0.187)
Clths	0.510*	-0.819***	0.523**	-0.762***
	(0.264)	(0.205)	(0.262)	(0.204)
Hlth	-0.167	-0.133	-0.217	-0.059
	(0.247)	(0.206)	(0.244)	(0.205)
MedEq	0.782***	0.660***	0.775***	0.671***
	(0.123)	(0.158)	(0.122)	(0.157)
Drugs	1.559***	1.547***	1.543***	1.634***
	(0.133)	(0.169)	(0.132)	(0.168)

Chems	-0.014	-0.680***	-0.035	-0.649***
	(0.150)	(0.159)	(0.149)	(0.158)
Rubbr	0.088	0.687**	0.043	0.662**
	(0.266)	(0.338)	(0.263)	(0.336)
Txtls	0.317	-0.782*	0.287	-0.984**
	(0.378)	(0.437)	(0.373)	(0.435)
BldMt	0.074	0.900***	0.073	0.783***
	(0.157)	(0.173)	(0.155)	(0.173)
Cnstr	-0.133	-0.395*	-0.250	-0.382*
	(0.221)	(0.205)	(0.219)	(0.203)
Steel	0.127	0.155	0.107	0.056
	(0.261)	(0.221)	(0.259)	(0.220)
FabPr	0.148	2.679***	0.003	2.542***
	(0.463)	(0.502)	(0.458)	(0.498)
Mach	0.144	0.578***	0.140	0.588***
	(0.133)	(0.111)	(0.132)	(0.111)
ElcEq	0.168	0.375	0.142	0.295
	(0.207)	(0.250)	(0.205)	(0.248)
Autos	0.065	0.108	0.021	0.025
	(0.201)	(0.184)	(0.198)	(0.183)
Aero	-0.260	-0.603**	-0.277	-0.572**
	(0.218)	(0.266)	(0.215)	(0.264)
Ships	-0.501	-0.006	-0.494	-0.020
	(0.330)	(0.360)	(0.327)	(0.357)
Guns	-0.093	-0.636	-0.029	-1.066**
	(0.428)	(0.445)	(0.424)	(0.444)
Gold	-0.745	-1.203*	-1.335	-1.039
	(1.189)	(0.682)	(1.177)	(0.677)
Mines	0.471	-1.174***	0.365	-1.108***
	(0.339)	(0.433)	(0.335)	(0.430)
Coal	0.877	-1.070	0.179	-0.814
	(1.377)	(0.750)	(1.368)	(0.746)
Oil	-0.313**	-0.692***	-0.351***	-0.651***
	(0.132)	(0.139)	(0.130)	(0.138)
Util	-0.353**	-0.558***	-0.324**	-0.608***
	(0.138)	(0.086)	(0.137)	(0.086)
Telcm	-0.389**	-0.875***	-0.389**	-0.921***
	(0.167)	(0.200)	(0.166)	(0.198)
PerSv	0.056	-1.035***	-0.056	-1.060***
	(0.411)	(0.269)	(0.406)	(0.267)
BusSv	0.551***	0.870***	0.526***	0.816***
	(0.110)	(0.138)	(0.109)	(0.138)
Hardw	0.541***	1.317***	0.524***	1.396***
	(0.169)	(0.170)	(0.167)	(0.169)
Softw	1.607***	1.324***	1.607***	1.411***
	(0.104)	(0.112)	(0.103)	(0.112)
Chips	0.715***	0.561***	0.683***	0.611***
	(0.121)	(0.108)	(0.119)	(0.107)
LabEq	0.438***	0.117	0.444***	0.155
	(0.149)	(0.177)	(0.147)	(0.176)
Paper	-0.061	-0.583***	-0.074	-0.558***
	(0.189)	(0.175)	(0.187)	(0.174)
Boxes	-0.389	-1.079***	-0.341	-1.100***
	(0.316)	(0.288)	(0.313)	(0.286)
Trans	-0.265**	0.278**	-0.273**	0.184
	(0.123)	(0.136)	(0.122)	(0.136)

Whlsl	-0.117	0.056	-0.151	0.066
	(0.133)	(0.155)	(0.131)	(0.154)
Rtail	0.662***	0.409***	0.620***	0.433***
	(0.118)	(0.100)	(0.117)	(0.100)
Meals	0.587**	0.054	0.557**	0.047
	(0.242)	(0.189)	(0.240)	(0.188)
Banks	-0.078	0.385***	-0.099	0.210**
	(0.144)	(0.100)	(0.143)	(0.104)
Insur	-0.413**	0.019	-0.495***	-0.051
	(0.164)	(0.095)	(0.163)	(0.095)
RlEst	0.105	-1.585***	0.075	-1.634***
	(0.270)	(0.473)	(0.267)	(0.469)
Fin	-0.136	0.142	-0.208**	0.086
	(0.106)	(0.125)	(0.105)	(0.124)
Other	0.078	-0.794***	-0.015	-0.585**
	(0.221)	(0.280)	(0.219)	(0.278)
Constant	2.261***	2.121***		
	(0.047)	(0.034)		
Observations	8,122	8,122	8,122	8,122
R^2	0.095	0.123	0.100	0.123
Adjusted R^2	0.090	0.117	0.092	0.115
Residual Std. Error (df = 8071)	2.377	2.340		
F Statistic	17.007*** (df = 50; 8071)	22.581*** (df = 50; 8071)	17.828*** (df = 50; 8051)	22.512*** (df = 50; 8051)
Note:				*p<0.1; **p<0.05; ***p<0.01