Dean&Francis

ISSN 2959-6157

Research on Stock Market Prediction based on Inverse Probability of Treatment Weighting Method

Jingqi Chen [1, *]

[1] School of Economics, University of Washington, Seattle, 98152, United States

*Corresponding author: jc2001@uw.edu

Abstract:

Although randomized controlled trials (RCTs) are considered the best method to evaluate the impact of interventions, their application in the financial industry is not feasible due to ethical and practical issues. Propensity Score Matching (PSM) minimizes the problem of bias in observational studies. By providing a way to circumvent these limitations. This study uses logistic regression and inverse probability of treatment weighting (IPTW) to evaluate the relationship between market capitalization and important financial indicators, such as revenue and income. Using data from 12 companies, this study identified the key drivers of market capitalization, including revenue, gross profit, and number of employees. The calculation of the average treatment effects (ATE) for Google and Apple showed a good correlation between these financial factors and market capitalization results. The results show that propensity score matching and inverse probability of treatment weighting can provide valuable investment strategies in the stock market by accounting for confounding variables, which are pretty much available for everyone. Though the market is fluctuating in the short run. The average treatment effects estimate the expected return in the long run and might give better market returns and higher Sharpe Ratios for investors.

Keywords: Propensity; average total effect (ATE); IPTW.

1. Introduction

Randomized controlled trials (RCTs) are considered the "gold standard" for comparing intervention effects because of their random distributions in the assignment of units to groups [1]. But there are some limitations for using this type of design. For example, if people want to conduct a trial related to the correlation of crime rate between parents and children. It is not ethical, because conducting those trials might push children to the criminal side that otherwise would probably never happen [2, 3]. The insinuations and being treated as potential criminals are what produces a real criminal. Propensity score offers a quantitative measurement that could avoid conducting RCT and compare the result of effects [4].

The propensity score is a statistical method used in observational studies. Using the score, people can reduce the effect of bias and get a better result. This method was originally developed in subjects related to medical research [5]. But now this method is proved can be applied in various fields such as financial markets. In finance, the propensity score method can be used to analyze the impact of financial related policies, investment strategies [6]. It might also estimate the effects of market events that lead to certain outcomes, taking into account potential confounders that may astray the results. So, the first step is to find propensity scores. Finding propensity scores is a complex matter. It involves estimating the probability that each individual in a study receives the treatment. In a simpler term, the probability that an individual is selected in the treatment group. The more covariates people try to minimize, the more accurate the result would be. The calculation is typically done using logistic regression [7].

However, other methods like machine learning algorithms can also be used. The logistic function is the heart of logistic regression. The process involves using the function to map predicted values to probabilities. And therefore, getting the propensity scores that were intended in the first place. The advantage of logistic regression compared to linear regression is that it produces binary results from 0 to 1, which is exactly the probabilities people are looking for [8]. Once the propensity scores are estimated, units in the treatment group (beneficiaries) are then matched with non-treated members with similar propensity scores or probability of participating in the treatment group. There are a few matching algorithms that can be used. The most common matching algorithms used in PSM include: Nearest-neighbor matching: Each program beneficiary is matched to the non-beneficiary unit with the closest propensity score. Non-beneficiaries for which there are no beneficiaries with a sufficiently similar score are discarded from the sample; the same is true for beneficiaries for which there is no similar non-beneficiary. A variation of nearest-neighbor matching matches multiple (for example, the or five) non-beneficiaries to one single beneficiary. Radius matching: (i.e., 'Caliper' matching): A maximum propensity score radius-a 'caliper'-is established, and all non-beneficiaries within the given radius of a beneficiary are matched to that beneficiary [9].

The main topic of this essay is whether it is possible to employ the propensity scores to predict or at least have a greater chance of beating the market return and achieving a higher sharpe ratio. There are a couple of ways someone can use the propensity score to assess the financial market. People can use the score to evaluate whether the market or specific stock is overestimated or underestimated. By doing a process called IPTW, and gets a score of reasonable market cap given the current financial stats [10].

2. Methods

2.1 Data Source

The data source of this article is from Kaggle, and it was compiled by Rishabh Patil which consists of data from various companies' 10-K annual reports and balance sheets from 2009 to 2023 of all the major public companies. The companies are categorized by industry. This provides a broad basis for analyzing the impact of financial data on company performance across different sectors.

2.2 Variable Selection

The financial performance of a company is influenced by a variety of metrics. These metrics can reflect different aspects of its operational efficiency. The goal of this research is finding the best predicting factor among the given types, which acts as the best indicator for predicting the value of the company. Hence, the predictor of the share price, since holding the numbers of shares equal, the market cap is proportional to the share price. The dataset consists of a total of 12 companies' financial reports, the dataset is very thorough. The dataset has 20 variables in total. The prime factors are the net income, gross profit, revenue and cash flow.

2.3 Method Introduction

This article is using IPTW to achieve the prediction desired. Inverse Probability of Treatment Weighting (IPTW) is a powerful statistical technique that helps mitigate this bias. Allowing researchers to draw more valid causal inferences from observational data. To address confounding, IPTW leverages propensity scores. IPTW is applied by assigning each participant a weight based on the inverse of the probability that they received the treatment they actually received. For example, if someone conducts research on heart disease, factors like age, history of diagnoses, smokers, alcohol users and people with higher BMI. In this instance, the older and people with bad habits subjects are more likely to have heart disease than the younger and healthier ones. Thus, those elderly also tend to have higher propensity scores, holding other covariates equal. If the older subjects are in the treatment group, they will have less weight, while if they are in the control group, they would have much more weight. The subjects that

ISSN 2959-6157

have propensity scores in the middle usually weigh more than the ones in the two ends. Hence, this balances out the covariates between treatment groups in observational studies.

3. Results and Discussion

3.1 Data Processing

Figure 1 demonstrates propensity scores and the frequency they occurred. The higher propensity ones are believed to have higher market caps. In common sense, a rise in the revenue and earnings imply that the company is operating better than previous year, or at least without adjustment for inflation. And the propensity is relying on the regression fit model, so it is reasonable to assume those with higher propensity scores achieve higher market caps. But the goal here is to predict the value of the company with better accuracy, so one of the factors has to be better than the other one. Therefore one more method is needed to add into the model, the AIC, which helps people to locate the best sets of variables to be used in prediction (this could be multiple covariates construct. AIC is used to compare multiple models. The model with the lowest AIC is typically preferred because it represents the best trade-off between fit and complexity. This data is only used to compare whether revenue or earnings weigh more on market cap. After having tested multiple AIC pairs to determine which has the lowest AIC score, the results showing significantly low revenue and gross profit score and employee counts score indicate that they are the closest fit overall.

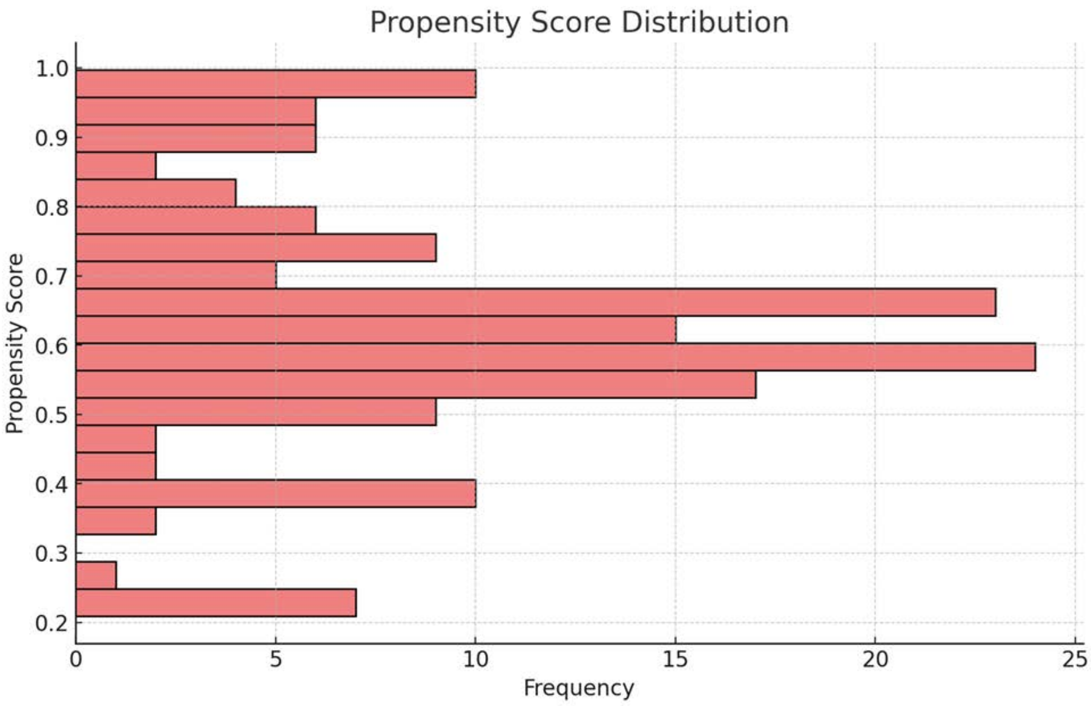

Fig. 1 Propensity score distribution

Note that Revenue & Gross Profit is 6.00, Number of Employees & Revenue is 6.00, Net Income & Debt/Equity Ratio is 12.59, Number of Employees & Gross Profit is 19.49, Revenue & Net Income is 21.76, Revenue & Debt/Equity Ratio is 22.05, Number of Employees & Net Income is 22.34, Net Income & Gross Profit is 22.38, Gross Profit & Debt/Equity Ratio is 22.38, Number of Employees & Debt/Equity Ratio is 22.80.

The dataset contains lots of information that could be misleading in the model, since they are not from the same company. So the wise way to do the comparison is isolate each company's data from the others, and do the logistic regression one company at a time, to prevent extreme value that makes the whole thing meaningless. The goal for this research is to determine what factor affects the market cap the greatest. When comparing the financial data of Apple (AAPL) and Google (GOOG) through logistic regression and inverse probability treatment weighting (IPTW) methods, it is evident that both companies' high market capitalizations are driven by their significant revenues and profits. This analysis is to find how these financial factors such as revenue, profit, and number of employees affect their market cap. By calculating the average treatment effect (ATE) to understand the impact of these financial drivers on market capitalization.

3.2 Model Evaluation

For Apple, the data shows that revenue, gross profit, and number of employees are important indicators of its market cap. The ATE for Apple's market capitalization (market capitalization is considered high or low here) is 1.0, if people are doing a binary treatment effect, by setting the market cap to be large(1) or small(0). The 1.0 indicates that there is a clear difference in market capitalization results based on the company's financial performance. When market capitalization is analyzed as a continuous variable, the ATE between the treatment group and the control group shows a huge difference, with an ATE of $973.76 billion. This reflects Apple's extraordinary market capitalization relative to other companies, which is driven by its strong revenues and profit margins (Table 1).

Table 1. Apple's financial report with propensity score and IPTW

Year	Revenue	Gross Profit	Employees	Market Cap	Propensity Score	IPTW
2022	394328	170782	164000	2066.94	0.7745594753	1.291056442
2021	365817	152836	154000	2913.28	0.9361529903	1.06820147
2020	274515	104956	147000	2255.97	0.6316018488	1.583275923
2019	260174	98392	137000	1304.76	0.7690572483	1.300293317
2018	265595	101839	132000	748.54	0.8735084966	1.144808555
2017	229234	88186	123000	868.87	0.5592114252	1.78823242
2016	215639	84263	116000	617.59	0.4395348334	1.784232205
2015	233715	93626	110000	586.86	0.7870813929	4.696630387
2014	182795	70537	97000	647.36	0.5693416212	1.756414713
2013	170910	64304	84400	504.79	0.848331762	6.593338282
2012	156508	68662	76100	500.61	0.1952190929	1.24257421
2011	108249	43818	63300	376.4	0.2041230786	1.256475685
2010	65225	25684	49400	296.89	0.09574065544	1.10587743
2009	42905	17222	36800	189.8	0.09822541574	1.108924578

A similar analysis conducted on Google shows a similar pattern. Google's propensity score and IPTW results indicate that its market capitalization is similarly strongly influenced by revenue, gross profit, and employee count. As one of the largest technology companies in the world, Google's high market capitalization is supported by these fundamental financial factors. The calculated Average Treatment Effect (ATE) for Google's market capitalization, using the number of employees as the covariate and Inverse Probability of Treatment Weighting (IPTW), is approximately 264.29 billion. Though less than what Apple has, still showing great correlation of market cap with the revenue (Table 2).

ISSN 2959-6157

Table 2. Google's financial report with propensity score and IPTW

Year	Revenue	Gross Profit	Employees	Market Cap	Propensity Score	IPTW
2022	257637	137737	156500	1736.72	0.79	1.27
2021	256731	136819	149500	1962.81	0.81	1.23
2020	182527	99437	135301	1215.45	0.67	1.49
2019	161857	89737	118899	920.34	0.71	1.41
2018	136819	76539	103459	820.72	0.68	1.47
2017	110855	59839	98777	729.45	0.65	1.54
2016	90053	48729	88541	648.73	0.63	1.59
2015	74639	40129	81511	527.83	0.58	1.72

3.3 Discussion

Both Apple and Google exhibit strong positive correlations between revenue and earnings with market cap. The logistic regression and propensity score modeling suggest that both companies have a high likelihood to reach a higher market cap based on their revenues and profit margins. Furthermore, considering their market position, they almost certainly would grow in the long term, regardless of the economic cycles in between. Since these two factors hold extraordinary weight on the market cap of both companies, it is reasonable to assume that the growth in revenue and profit is an indicator of market cap growth. However, the lack of quarterly reports makes the research incomplete. With the more detailed data, the results would be much more compelling than what the author have now. This is unsurprising given the scale and profitability of both companies. The ATE for Apple, at $973.76 billion is still shocking, underscores just how substantial the difference in market cap can be between treated and control groups, representing the impact of financial success on market cap.

4. Conclusion

This study primarily demonstrates the way inverse probability treatment weighting (IPTW) and propensity score matching (PSM) can be used in the financial market. When people try to evaluate how financial issues affect market capitalization. Using these techniques on the financial data of Google and Apple, the study finds that important metrics like revenues, gross profit, and workforce size significantly affect market capitalization. The average treatment effect (ATE) was determined to be $264.29 billion for Google and $973.76 billion for Apple, showing the significant influence of these financial variables on a company's market capitalization. PSM offers a more reliable tool for making judgments that are more often to be correct, particularly when attempting to comprehend intricate relationships in financial markets. Since it allows people to reduce the effects of confounding variables and focusing on some of the variables that are intended as study priority.

While the results showed significant effects of these variables, there are some limitations to this study. The lack of quarterly data leaves a hole that could compromise the analysis's accuracy, only dozens of reports can't forge a strong and thorough analysis. A greater knowledge of how variations in financial performance over shorter time periods impact market capitalization can be obtained with more regular and comprehensive data. To increase the accuracy of the findings, future research should concentrate on broadening the study's scope, including more financial variables, and obtaining more frequent data reporting. The knowledge gathered from this study, however, offers a strong basis for applying PSM in financial research. It gives investors a means to enhance market performance projections. With the techniques that are used in this research, people can ignore the less correlated variables and make better selections and investments.

References

[1] Maria K, Christian R, Monika S, Maria B. Randomized Controlled Trials. Dtsch Arztebl Int, 2011, 108(39): 663-668.

[2] Cecilia N. The ethics of clinical trials. Ecancermedical science, 2014, 8: 387.

[3] Phyllis S, Mary M C, Jeffrey D. Ethical Considerations of Randomized Controlled Trials. Oxford Scholarship Online, 2009, 19-44.

[4] Ameneh E V, Leila J. A brief guide to propensity score analysis. Med J Islam Repub Iran, 2018, 32: 122.

[5] Emily G, Tim W, Jamie C S, Mark L. A review of the use of propensity score diagnostics in papers published in high-ranking medical journals. BMC Medical Research Methodology, 2020.

[6] Jonathan E S, Quinn T S, Robert L W. Propensity Score Matching in Accounting Research. The Accounting Review,

2017, 92 (1): 213-244.
[7] Todd G N, Kathleen M C. Logistic regression. Methods Mol Biol, 2007, 404: 273-301
[8] Horacio M C, Juliana C F. Linear and logistic regression models: when to use and how to interpret them? J Bras Pneumol, 2022, 48(6): 20220439.
[9] Peter C A. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res, 2011, 46(3): 399-424.
[10] Peter C A, Elizabeth A S. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med, 2015, 34(28): 3661-3679.