Empirical Analysis of Regions, Fund Allocation and Profits for Startups

Zhongjian Wang*

Department of Economics, University of Michigan-Ann Arbor, Ann Arbor, United States

zhongjiw@umich.edu

Abstract. This paper focuses on four indicators for different startups: R&D spending, marketing spending, administration spending, and profit. First, in the multiple linear regression of the original variables on profit, the presence of Collinearity between R&D and market spending only led to a single linear regression between R&D spending and profit. Then, factor analysis concluded two factors that resolved the Collinearity and allowed a better multiple linear regression. Moreover, the significant differences between startups with large and small profits for different spending were analyzed using comparative analysis. It concluded that different R&D and marketing spending made significant differences between startups' profit, but not for administration spending. Finally, the Normal distribution, ANOVA, and Non-parametric tests prove that there is no significant difference between different spending and profits among different regions, which shows the homogeneity of the development of different geographical startups. All these analyses can help startup entrepreneurs to generate profits better.

Keywords: Startups, R&D spending, Marketing spending, Administration spending, Profits

1 Introduction

From the earliest company, Edison General Electric (GE), to the most popular Tesla electric car, to SpaceX, valued at 100.3 billion dollars, startups have always been a gateway to the top commercial stage. Although Covid-19 has dramatically increased the unemployment rate, more and more young people with ideas and skills are starting to increase the number of startups even further. Based on the 2020 National Report on Early-Stage Entrepreneurship in The United States, the percentage of startups dramatically increased; more specific, woman startup entrepreneurs increased from 0.23% in 2019 to 0.30% in 2020, and male startup entrepreneurs increased from 0.38% in 2019 to 0.48% in 2020 [1]. Crow further explains that the most important factor for startups is profit; the primary purpose of a startup is to make a profit, which reduces the need for further financing. [2]. Therefore, in order to better guide startup entrepreneurs, a quantitative analysis of the impact of fund allocation on the startup's profits is necessary. Further, some scholars also have different views on the profit development of the different startups. Bartik argues that the choice of the region by a startup may also affect

H. Mallick et al. (Eds.): ICEMCI 2022, AEBMR 231, pp. 874–882, 2023.
https://doi.org/10.2991/978-94-6463-098-5_100

the maximization of profits [3]. Thus, it is also necessary to quantitatively study the spending and profits of startups in different regions. The study of profits and financial input of startups can also better help angel round investors, to some extent, to better understand the effectiveness of a startup's resource allocation in addition to valuation and skills. As Ball said, traditional investors would be more concerned with the profits of the startup and the factors that can affect it [4].

Multiple linear regression and factor analysis are widely used to quantify a startup's profits and success. For example, Albourini et al. assume that a startup making profits represents success, and he uses variables such as entrepreneurial enthusiasm to predict the success of startups based on Multiple linear regression [5]. Kimuli and Isaac further adopt a simple linear regression model to study the effect of different fund spending on profit for startups, demonstrating there is a significant positive relationship between those two [6]. Sathaworawong et al. moved forward by combining factor analysis to downscale fourteen factors to go with seven to analyze startup funding better using multiple linear regression [7]. Moreover, normal distribution, comparative analysis, and ANOVA are widely used to quantify the impact of regions on the financial inputs and profitability of startups. For example, Scherer initially analyzed the distribution of startups' innovation profits by using distributions such as the normal distribution [8]. Micheli et al. further used the T-Test of comparative analysis, using financial statements to compare the profitability of startups and university spin-offs, among other factors [9]. Shetty also used comparative analysis to analyze the sources of funding and investment in different startups [10]. Moreover, Kartanaitė went to the ANOVA test to identify the financial status of different startup unicorns and the financial efficiency of different regions [11].

This study used factor analysis based on the traditional multiple linear regression to obtain a solid linear relationship between the factors of business and management on the profit of startups. In addition, a comparative analysis was also used to demonstrate further whether different factors influence the size of the profit. This indicates that whether startups can generate high profits is strongly related to the allocation of financial inputs to the business, which has some implications for the development direction of startups and investors' investment. Furthermore, it also used normal distribution and ANOVA to obtain evidence that startups in the three developed cities are not influenced by geography in terms of different variables or factors, showing consistent evidence of startup development in different cities in the US.

The study is organized as follows: Chapter 2: For the original data, the multiple linear regression of the three variables on profit only found a strong linear relationship between R&D spending and profit. Chapter 3: Because the results of chapter 2 were unsatisfactory, two factors were concluded using factor analysis: business and management factors. A significant multiple linear regression relationship was found using those two new factors. Chapter 4: Comparative analysis was used to find significant differences in RD spending and marketing spending, but not in administration spending, between large and small profitable startups. Chapter 5: The ANOVA analysis was conducted on those startups' profits and inputs of three states, which were normally distributed but not significantly different because the significance level of the ANOVA test was greater than 0.05. The last part is the Conclusion.

2 Linear Regression of the Three Original Variables

2.1 The Relationship Between RD Spending and Profit

In order to find out the effect of different startups' funding allocation on profits, multiple linear regression in SPSS was used to analyze using the stepwise approach. It shows that only R&D spending enters the P-value range of 0.05-0.10. After increasing the range to 0.1-0.2, Marketing spending enters the range, and two models exclude Administration spending. According to the following table, the first linear regression model only includes RD spending, which has a 94.7 percent explanation. Furthermore, it has a very strong significance level (0.00) in both F-test and T-test, which shows that those regression and coefficients are significant, and R&D spending and profit have a strong positive correlation.

Table 1. ANOVA

Model 1	F	Sig.
Regression	849.789	0.000

Table 2. Coefficients

Model 1	t	Sig.	95% confidence Interval lower Bound	Upper Bound
Constant	19.320	0.000	43930.115	54135.683
RDspend	29.151	0.000	0.795	0.913

However, when the second model includes market spending, the significant level in the T-test is not significant (0.06). Moreover, the Tolerance of Collinearity Statistics of the multiple linear regression of the second model is also less than 0.5, which indicates that there is collinearity between marketing and R&D spending, which may lead to inaccurate models. Therefore, only one strong linear regression was found using the three variables on profit.

$$Y_1 = 0.854X_{RD} + 49032.899 \quad (1)$$

It is important to understand the relationship between R&D spending and profits. Although investing heavily in innovation may cause a startup to underperform, if a startup wants to make a good product, or a revolutionary product, then R&D spending is essential. Moreover, there is a metric called RORC (Return on Research Capital), which is a metric that investors are interested in, that is, the return on the profit generated by R&D spending. Research and development success represents whether a startup's value will increase, whether the product will be sold, and whether the startup will be profitable.

Furthermore, the collinearity between marketing and R&D spending can be understood in two dimensions. First, generally, startups do not have sufficient initial funds, and even if they receive a large amount of financing, they also need to allocate the capital investment wisely. Because if too much money is invested in the market spending, then the relative investment in R&D will be reduced. Secondly, if the R&D spending is ideal and produces a very efficient product, then appropriately, the marketing spending can be reduced because a good product will attract many investors and consumers.

3 Factor Analysis, Multiple Linear Regression

Since substituting the three variables into the multiple linear regression did not provide significant inference. Moreover, there was a multicollinearity relationship between both R&D and marketing spending. Therefore, factor analysis was attempted to discover the relationship among the three variables and summarize new factors to perform multiple linear regression better.

3.1 Factor Analysis

First, in the Correlation Matrix, it was shown once again that R&D and Marketing spending are strongly related (correlation: 0.724), but they are not strongly correlated with Administration spending.

Also, in Bartlett's Test of Sphericity, the significance level is 0.00, which rejects the null hypothesis and indicates that factor analysis can be done. Moreover, in the Communalities table, the explanatory power of the extracted factors was greater than 85%, and the Total Variance Explained by the two factors was 92.458% (much greater than 80%), which indicates that those two factors explained the variables very well.

Table 3. Total variance explained

Component	Total	Initial Eigenvalues% Variance	Cumulative%
1	1.754	58.481	58.481
2	1.019	33.977	92.458
3	0.226	7.542	100.000

Eventually, according to the Rotated Component Matrix, the factor analysis concluded two factors: Business Spending (RD+Market), and Management Spending (Administration). Here factor analysis can help variables solve collinearity well, and new factors can help to analyze the sample more precisely.

Table 4. Rotated Component Matrix

Variables	Business Factor	Management Factor
MarketingSpend	0.941	
RDspend	0.912	
Adminstration		0.992

3.2 Multiple Linear Regression based on New Factors

Using the stepwise approach with business spending and management spending, both factors entered the p-value range of 0.05 to 0.1, and both had more than 90% interpretation. According to the following table, the multiple linear regression with both factors has a strong significance (0.00 or 0.01) in both F-test and T-test. This indicates that business and management spending have a significant effect on profit. And they do not generate multicollinearity with the help of factor analysis.

Table 5. ANOVA

Model 2	F	Sig.
Regression	158.618	0.000

Table 6. Coefficient

Model 2	t	Sig.	95% confidence Interval lower Bound	Upper Bound
Constant	53.576	0.000	107806.673	116218.605
Business	17.431	0.000	32564.510	41061.845
Management	3.660	0.001	3480.838	11978.173

Therefore, here we arrive at a significant multiple linear regression formula.

$$Y_2 = 36813.178X_B + 7729.506X_M + 112012.639 \quad (2)$$

After the factor analysis removed the collinearity, the p-value of the management factor also became smaller and could better explain the profit. In general, the business factor: R&D spending can help startups to develop good products, and marketing spending can help startups to promote their products; while the management factor: can help startups to motivate their employees better to work overtime and better increase startups' profit. From the coefficients of X_B and X_M, it can also conclude that the magnitude of the change in the business factor is much greater than the change in the management factor, so this can be a preliminary indication that the business spending can better determine the profitability of the startup. Moreover, in order to present a more

accurate result determination of which factor most affects profit, it needs to be further analyzed with another method, such as comparative analysis.

4 Comparative Analysis - Different Profit Scale Startups

The multiple linear regression for profits shows what affects startup development. In order to investigate more deeply the effect of different spending on profits, the startups in the sample above and below the median profit are divided into two categories to perform a comparative analysis.

4.1 The Importance of R&D and Marketing Spending

First, run a normality test for large and small profit startups; if the distribution is normal, then a T-test can be done; if not normally distributed, a Non-parametric test (Mann Whitney U test) is needed. According to the Shapiro-Wilk test, Administration spending, marketing spending, and management factor all failed to reject the normality null hypothesis, and their significance level was greater than 0.05. However, the significance of R&D spending and business factor was less than 0.05 and rejected the normality hypothesis. For variables and factors that pass the normality test, one can continue to compare startups with different profit sizes by independent sample test. According to the table below, the significance of the spending and factors of management of startups with different profit sizes is greater than 0.05, indicating no significant difference in the management spending of startups with different profit sizes. However, the two-tailed significance of market spending is 0.000, indicating a statistically significant difference between the market spending of large profit startups and small profit startups.

Table 7. Independent Samples Test--T-test for Equality of Means

Variable	t	Sig.(2-tailed)
Administration	-0.907	0.369
	-0.907	0.370
Marketing Spend	-5.671	0.000
	-5.671	0.000
Management Factor	-0.845	0.402
	-0.845	0.403

The Non-parametric Independent-Sample Mann-Whitney U test was for variables and factors which are not a normal distribution. According to the following table, it was found that the significance of both R&D spending and business factor was 0.000, which rejected the null hypothesis (the distribution is the same across both profit size startups), respectively, which indicates that there is a significant difference in R&D spending and business factor between startups with large and small profits.

Table 8. Non-parametric Independent-Sample Mann Whitney U test

Variable	Sig.	Decision
RD Spend	0.000	Reject the null
Business Factor	0.000	Reject the null

This reinforces the magnitude of the multiple linear regression coefficients. The R&D and marketing spending included in the business factor will largely affect whether the startup is more profitable or less profitable. The administration spending and factor will affect the startup's profit but will not have a significant effect relatively to business spending. This relationship is quite understandable, because, as a startup, developing products and promoting them is the most important thing. In many cases, startups' profits are also taken into R&D, and employees may not earn a lot of income or overtime fees, and many of them keep going by their imagination and passion for the future. Moreover, management spending can keep employees, but putting money into research and development of the product is the future of startups. Thus, the business factor (R&D+Marketing) may influence the size of a startup's profits to a greater extent.

5 Normal Distribution and ANOVA

In order to find out whether different cities affect startups' spending and profit, tests of normality are first used to check whether all variables could be analyzed by t-test. It is shown that all the variables and two factors are normally distributed (>0.05), which can be further tested by t-test, except for Administration spending, which should use Non-parametric Tests.

Based on the normality test results, all the normally distributed variables were first analyzed by the Test of Homogeneity of Variances. Moreover, all the significance is greater than 0.05. Therefore, it demonstrates that the P-value fails to reject the null hypothesis, indicating that the variance is homogeneity across different cities, which F-test can further test.

Based on the one-way ANOVA test with LSD, all significantly greater than 0.05, indicate the failure to reject the null hypothesis. Therefore, it can be stated that there are no significant differences in RD spending, marketing spending, profit, and business and management factors across different cities. Furthermore, combined with the Non-parametric test of administration, the results fail to reject the null hypothesis(sig.>0.05), which indicates that there is no significant difference in the administration costs of startups in each region.

Table 9. ANOVA

Variable	F	Sig.
RD spending	0.532	0.591
Marketing Spending	1.194	0.312
Business Spending	0.940	0.398

Management Spending	0.008	0.992
Profit	0.575	0.567

Table 10. Hypothesis Test Summary

Model	Null Hypothesis	Test	Sig.
1	Same across state	Kruskal-Wallis	0.815

Finally, the significance of the two-by-two comparisons for each dependent variable in the multiple comparisons is greater than 0.05. This further indicates no significant differences across cities for different inputs and profits. Thus, based on ANOVA and Non-parametric tests, those analyses indicate that the differences in inputs and profit of startups in different cities in the sample data are caused by random errors and cannot make a significant inference about the overall startups.

Moreover, the reason that profits do not have significant differences across regions, it is also possible that rational entrepreneurs of different startups will choose the geography of their business by the agglomeration of industries. For example, if an entrepreneur wanted to start a high-tech business, he would choose to go to Silicon Valley in California. If the entrepreneur wants to start a financial business, he may choose to go to New York, where there are more financial support facilities. If an individual wants to start a retail or tourism business, that individual can go to Florida. Because startups always involve greater risk, a rational entrepreneur will not blindly choose to develop in a region unsuitable for a particular industry. This may be a reason why the profits of startups are relatively indistinguishable from one region to another.

6 Conclusion

At first, only one significant relationship was found by substituting three variables into the multiple linear regression. Subsequently, factor analysis was conducted to summarize the new factors. This allowed for a better analysis of the correlation between different spending and profit, which found that the business and management factors had a very significant relationship with profit. Moreover, a comparative analysis of big and small profit size startups reveals that R&D spending, marketing spending, and business factor are significantly different for different profits size startups, but administration spending and factor are not significantly different. This is quite understandable, because product development and promotion is the best way for startups to generate profits. Finally, the analysis of the normality distribution test, ANOVA, and Non-parametric tests did not reveal that startups in different geographic areas would have different profits, and spending. So this would allow startup entrepreneurs to understand that the difference in profit and expense is not significantly different among those cities. Furthermore, this allows startups to choose a location to start a business by looking at other factors such as, the ease of bank lending, government support policies, and investment activity in different locations.

References

1. Fairlie, Robert W., and Sameeksha Desai. "National report on early-stage entrepreneurship in the United States: 2020." Available at SSRN 3810193 (2021).
2. Crow, David. "Valuing usability for startups." In Cost-Justifying Usability, pp. 165-184. Morgan Kaufmann, 2005.
3. Bartik, Timothy J. "Small business start-ups in the United States: Estimates of the effects of characteristics of states." Southern economic journal (1989): 1004-1018.
4. Ball, Stephen J. "Serial entrepreneurs, angel investors, and capex light edu-business start-ups in India: Philanthropy, impact investing, and systemic educational change." In Researching the global education industry, pp. 23-46. Palgrave Macmillan, Cham, 2019.
5. Albourini, Firas, Alaeddin Ahmad, Mohammad Abuhashesh, and Nawras Nusairat. "The effect of networking behaviors on the success of entrepreneurial startups." Management Science Letters 10, no. 11 (2020): 2521-2532.
6. Kimuli, Isaac. "The Effect of Startup Capital of the Owner of Small Business Enterprises to their Performance in Uganda: Case Study Kawempe." PhD diss., Makerere University, 2019.
7. Sathaworawong, Pimolrat, Natcha Thawesaengskulthai, and Kanis Saengchote. "Determinant of Startups' Fund-raising Value: Entrepreneur and Firm Characteristic." In 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1309-1314. IEEE, 2018.
8. Scherer, Frederic M. "The size distribution of profits from innovation." In The economics and econometrics of innovation, pp. 473-494. Springer, Boston, MA, 2000.
9. Micheli, Anna Paola, Carmelo Intrisano, and Anna Maria Calce. "Italian university spinoffs vs Italian innovative companies: a comparative analysis of profitability, liquidity and efficiency profile." (2021).
10. Shetty, Kishan Kumar. "A comparative study on impact of venture capital financing on startups in India." In International Conference on Education, Humanities and Management (ICEHM-17) March, pp. 14-15. 2017.
11. Kartanaitė, Inga, and Rytis Krušinskas. "Financial Efficiency of Unicorns: Regional and Sector Related Aspects." Engineering Economics 33, no. 2 (2022): 200-214.