SCHOOL OF
ECONOMICS AND
MANAGEMENT

LUND UNIVERSITY

Network Effects and Valuation

Examining the platform-mediated business model of international tech companies in order to test for an exponential association between revenue and user base

by

Carl Fredrik Falkner

Oskar Munck af Rosenschöld

May 2021

Bachelor's Programme in International Business

Supervisor: Niklas Lars Hallberg

Abstract

This study examines the association between revenue and user base for companies who operate a platform-mediated business model and benefit from network effects. More specifically, the study incorporates both an exponential and linear function in order to see which one most accurately explains the association. Previous research has theorized the notion of an exponential association between revenue and user base, as the number of connections increases at an exponential rate when a new user joins a platform. Furthermore, as previous research points toward a different return on invested capital across industries, an industry variable will be incorporated. Therefore, the study aims to investigate the theoretical aspects of network effects and Metcalfe, Reed and Odlyzko's law in order to see how these theoretical viewpoints creates the notion of an exponential association between revenue and user base. Moreover, the study aims to test the association across different industries.

By incorporating data on 85 companies into both a linear and exponential regression, two findings have been observed. Firstly, the dataset cannot determine whether an exponential relationship more accurately explains the association between revenue and user base compared to a linear one. This finding therefore states the contrary to existing literature as an exponential association cannot be determined. Secondly, industry does play an important role when studying the association between revenue and user base. This finding shows that some industries are more efficient at extrapolating revenue from its user base.

Keywords: Metcalfe's law, network effects, valuation, platform-mediated business model, shareholder value approach, increasing returns, standards

Word count: 15'795

Acknowledgements

As we are applying the finishing touches to our thesis, we would like to reflect back on the process of writing our thesis paper. It has been a challenging ten weeks and our ambition has led to situations of uncertainty. The process of writing a thesis is complicated and requires determination, vigour and an intellectual mindset. Furthermore, the writing process has ensured us that determination, curiosity and proper guidance are key in order to overcome the challenges. Therefore, we would like to show our gratitude towards our supervisor for providing the necessary guidance to overcome the difficult parts of the research process.

We are immensely thankful to our supervisor Niklas L Hallberg for the support during the writing period. Dr. Hallberg has provided us with the necessary feedback, guidance and support to handle the difficult aspects of our thesis. It was a true privilege to get the opportunity to write a thesis with the help of Dr. Hallberg’s exquisite knowledge within the field of management. Dr. Hallberg’s, valuable insights and professionalism made him an ideal supervisor and future students would be privileged to get the same opportunity.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Problematization

International tech companies who employ a platform-mediated business model are generally pointed out as culprits in terms of overvaluation. These types of companies include Amazon, Facebook and Alphabet who are some of the highest valued firms on the stock market in the midst of the COVID-19 pandemic (Delevingne, 2020). Platform-mediated companies operates a digital business model that connects users via a network. However, one should note that these firms also apply vastly different business models than the traditional tangible asset-based counterparts. This is where the difference between business models that incorporate tangible and intangible assets to generate cash flow becomes apparent. Traditionally, tangible asset-based business models have permeated the world of business, where a company use physical assets, such as equipment and inventories to generate cash-flow (Greco, Crielli, & Grimaldi, 2013). Today however, many of the world's largest companies use intangible assets instead where for example a new technology such as a digital platform that connects users to a network replaces the physical asset to generate cash flow (Eisenmann, 2007; Kolasky, 1999; Greco, Crielli, & Grimaldi, 2013). This in turn creates a problem where the traditional valuation is applied to companies who operate a business model with few tangible assets. Therefore, there exists a risk of mispricing the intangible assets such as the positive externalities brought upon by network effects (De Boer, 2021). These positive externalities could therefore form a new perspective between revenue and user base. Metcalfe (2013) states that the value of a company increases exponentially as each new connection brought on by a new user follows an exponential relationship. In a traditional industry selling a physical product, the sales revenue is a function of the price and units sold (Collier, 2015). Therefore, the relationship between revenue and customer should be linear as the amount of goods sold is a function of number of customers. However, Metcalfe (2013), Reed (1999) and Odlyzko (2005) states that platform-mediated business models cause an exponential development of revenue when user base grows linearly.

Network effects could help analyst forecast future revenue streams as a function of user base when valuing a platform-mediated firm. This is due to the fact that the theory of network effects help estimate a company's value by looking at the value added by each new user to the network in an exponential manner (Katz & Shapiro, 1985; Metcalfe, 2013; Odlyzko 2005; Reed, 1999;). Network effects is the phenomena that the utility acquired from a good or service for one user depends on the number of other users (Katz & Shapiro, 1985). In other words, a user gets more value from a good or service as more users join the network. Therefore, the positive externalities generated by network effects will exponentially increase as more users join the same network (Katz & Shapiro, 1985; Metcalfe, 2013). This growing network could in theory become an intangible asset that generates cash flow and increases the value of a company. A company who incorporates a platform-mediated business model could see its value increase due to the positive externalities brought upon by its users.

The shareholder value approach determines that firms should focus on creating cash flow generating abilities (Rappaport, 1986). To do this, firms should follow the strategy that creates the greatest sustainable competitive advantage (Rappaport, 1986). Gallaugher and Wang (2002) identifies that conquering a large part of the market share in a network market is the single most important variable in creating a competitive advantage in that type of industry. This is because of the large positive externalities, increasing return dynamics and the possibility of setting a standard created by a large user base (Arthur, 1996; Bonardi & Durand, 2003; Eisenmann, Parker & Van Alstyne, 2011; Katz & Shapiro, 1985).

The shareholder value approach is reflected in a discounted cash flow analysis (DCF). According to Koller, Goedhart and Wessels (2015) a DCF quantifies the value of a company by estimating future cash flows and discounting them with a weighted average cost of capital (WACC). The authors state that the "formula [DCF] … represents all there is to valuation. Everything else is just detail" (Koller, Goedhart & Wessels, 2015, p.31). A problem with a DCF however, is its reliance on assumptions to quantify the equity value. The free cash flow is estimated by analysts and then discounted by the discount rate (WACC). The free cash flow can change drastically depending on who is doing the DCF and what assumptions are made. Furthermore, as each year's annual revenue is estimated, the assumptions become less reliable for each forecasted year. If an exponential function is better at forecasting revenue than a linear function, the valuation of a company carried out with a DCF analysis could be refined as free cash flow could be more accurately estimated. Another important aspect reflected upon by Koller, Goedhart and Wessels (2015) is the fact that one needs to take industry into consideration when forecasting cash flows. The authors state that return on invested capital (ROIC) shifts vastly between industries. Furthermore, different industries are able to more efficiently gain a competitive advantage while simultaneously achieving increasing return on capital over time (Koller, Goedhart & Wessels, 2015). It is therefore of importance to take industry into consideration when forecasting growth as the ROIC is vastly different between industries.

The shareholder value approach and network effects could therefore be combined to emphasize the relationship between revenue and user base. Researchers have previously tried to find ways of pricing network effects, with the most well-known model being Metcalfe's law. Metcalfe (2013) states that the value of a network is proportional to the number of connected users squared, where the proportionally is an unknown constant (see Appendix A). This theory has been rigorously criticised and alternative ideas have been offered in order to explain the value of networks. These alternative laws include Odlyzko's law (2005) who states that value is proportional to the user base but diminishes after a given time. The proportionality is also an unknown constant but compared to Metcalfe, Odlyzko believes that the value of each new user decreases after an undefined network size. The last law, Reed's law (1999), states that the value is proportional to the number of possible subgroups (see Appendix A). The proportionality is yet again an unknown constant where the value of the network is directly proportional to the number of individual groups formed by users on the network. A study by Zhang, Liu and Xu from 2015 tested the fit of the different laws by applying them to the historical values of Facebook and Tencent. The authors came to the conclusion that Metcalfe's law most efficiently explains the value of networks compared to its rivals, however the sample size was restricted to two (Zhang, Liu & Xu, 2015).

Metcalfe's and the above-mentioned laws create an interesting notion of a set proportionality between revenue and user base. The laws have been tested in various case studies (Metcalfe, 2013; Zhang, Liu & Xu, 2015). However, these previous studies are somewhat flawed. Firstly, they were all conducted with a qualitative case design and never applied to a larger sample size than two. This can be due to the fact that they were conducted at a time when public information on platform-based companies was limited (see Metcalfe, 2013; Odlyzko, 2005; Reed, 1999). Secondly, the studies look at the proportional aspects of the association supported by a visual fit or a statistical relationship with a sample size equal to two that cannot establish an association which enables out-of-sample inference. Lastly, existing literature has not tested for industry specific factors whose effect could be significant according to Koller, Goedhart and Wessels (2015). Therefore, there could be an industry variable that effects the association between revenue and user base. Hence, this study ventures to add to the literature by examining the relationship between user base and revenue making use of a statistical method with a sample size of 85 and adding an industry variable.

1.2 Research Purpose

This paper seeks to test both an exponential and linear relationship between user base and revenue, in order to see which function most accurately describes the association between user base and revenue. Additionally, the study wants to find out if any specific industry is more or less reliant on its user base for revenue growth. If an exponential association between users and revenue can be observed, it can refine the way revenues are forecasted in platform-mediated companies, creating a more coherent and realistic method to help in valuation. This is due to the fact that future revenue could be more accurately estimated with help of the user base coefficient which would reduce the amount of assumptions needed when forecasting free cash flow. Furthermore, if an industry difference is observed then industry should be taken into consideration when forecasting revenue growth. Thus, our research question is as follows: *How does the user base of international platform-mediated companies affect revenue?*

The study is conducted with quantitative data collected from investor relations material of 85 companies on three variables: revenue, user base and industry. A regression analysis is used to test whether an exponential or linear relationship most accurately describes the data measured by Akaike information criterion that evaluates the best model fit. Variables taking into account the industry profile are then added to the regression that most accurately fits the data in order to test for industry differences. Our research model is illustrated in figure 1.1.

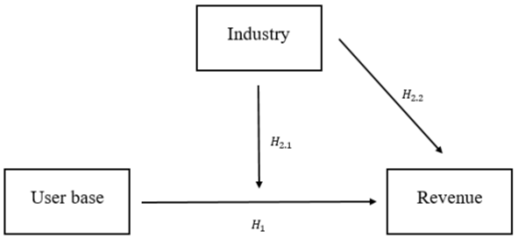


Figure **1.1**: Research Model

2 Theory and Hypotheses

2.1 Shareholder Value Approach

Rappaport (1986) defines that the core purpose of a running business is its ability to generate cash flow and that this is also the best metric to use in valuation. The idea of relying on free cash flow to derive a fair value of a company, called the shareholder value approach, was introduced by Rappaport in 1986. He defines that shareholder value is derived from planned actions by management investing shareholders equity in order for them to make a superior return compared to what they could have earned by investing in assets having the same amount of risk themselves. Shareholders derive their value of an investment from the current dividend and, as residual claimants, the firm's ability to generate future cash dividends which is reflected in the market price. They do this through estimating future cash flow and discounting it by the weighted average cost of capital. Rappaport define that the strategy generating the greatest sustainable competitive advantage will be the one to generate the best return to shareholders. In doing this the management is going to succeed in satisfying other stakeholders as well such as employees, debtholders, customers and suppliers since they also depend on the company's ability to generate cash flow. With this rationale, he means that business strategies should be evaluated on their ability to maximize cash flow and hence increase return to shareholders. Rappaport describes that through estimating the value of a company with the shareholder value approach, the investor incorporates future prospects of the business, minimizes the effect of alternative accounting standards and take into account the risk and time value of money that other performance metrics lack.

Since the introduction on Rappaport's ideas a shift in management focus has occurred putting cash flow and revenue in focus (Francis & Minchington, 2002). Additionally, Bayón, Gutsche and Bauer (2002) state that about 70 percent of stock value is located in investor's belief of how much future return a company can generate from its intangible assets and if that return will be greater than the cost of capital. Investors thus expect regular cash flow reporting and customer lifetime value calculations in order to assess an investment (Bayón, Gutsche & Bauer, 2002). In digital platform companies, whose cash flow generating ability greatly depend on the intangible assets, taking the form of the platform along with its specifications and brand value, this is even more relevant than in traditional tangible asset-based companies.

2.2 Network effects

In order to estimate future cash flows a broad range of estimations has to be drawn about the current and future actions and operations of a company. Usually, analysts from different banks and financial institutions track the actions of companies and can thus make informed guesses about their future in order to arrive at a present value. However, because the future is uncertain, estimates tend to vary between analysts. Hence, more information on what affects companies in the market and what the dynamics of competition looks like leads to more educated guesses and estimates. Companies with platform-mediated business models must take into account the network externalities that their own and their competitors platforms generate when formulating strategies. In turn, analysts take this into account when drawing assumptions about the future of a platform-mediated company and estimating their value.

Katz and Shapiro (1985) define that network effects arise from the positive externalities of products where the utility derived from one user using the product correlates positively with the number of other users using the same product. In other words, connecting a customer (defined as a node of the network in the original article) to the network will increase the utility derived from the other nodes. This mechanism is called demand side economies of scale which increase switching costs, users' willingness to pay and barriers to entry (Eisenmann, Parker & Van Alstyne, 2006; Eisenmann, Parker & Van Alstyne, 2011).

Katz and Shapiro (1985) define three versions of network effects: direct, indirect and two-sided mutually reinforcing networks. Direct effects are externalities where the value derived from a product is directly correlated with how many people that are using it. An example of where direct network effect can be found is on the social media company Facebook. The utility that a user derives from the network is directly correlated to how many connections he can make and interact with. Secondly, the authors define indirect effects as the utility that is indirectly gained from other users using the same product. Returning to the Facebook example, indirect effects can be demonstrated through how many other products or services connect themselves to Facebook's API. In this example, the utility increase as the user is able to use the product in other applications. This effect is dependent on how many users that are connected to Facebook since the larger the user base is, the more attractive a connection to Facebook's API becomes. The third effect, two-sided network effects are described as the positive network effects that arise when a platform is used by two groups of users that interact with each other. In this type of network, the utility derived by a user group is dependent on the number and quality of the other. Amazon, the global e-commerce giant, is a good example of this type of network effect since the two sides (consumers and merchants) are dependent on the number of nodes on the other side. Here, Amazon is acting like a market intermediary through enabling and simplifying the connection between the two sides and hence reduce searching costs.

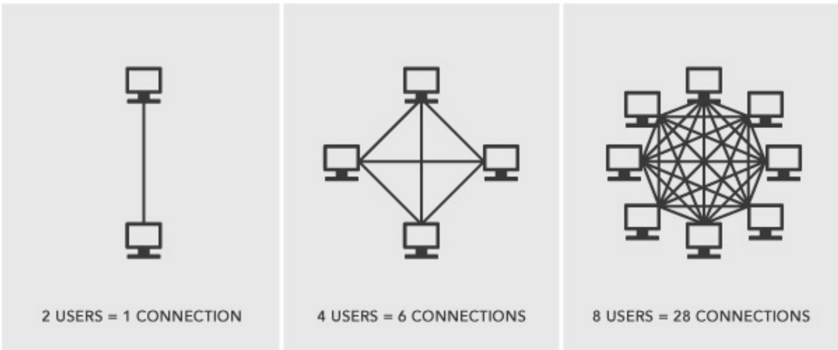


Figure 2.1: *When 4 plus 1 equals 10 (Busse, 2012, n.p.). Example of direct network effects.*

2.3 Dynamics of Competition

The very idea of a platform-mediated business model is inherently based on scale economics and positive feedback systems. Bresnahan (1998) explain that as the network grow the user base learns new ways of using the platform generating an information asset to the company that guides them in what direction to innovate. As the ideas get integrated the platform becomes more attractive and more users join. This creates a virtuous cycle that generates returns for both sellers and buyers and is mutually reinforcing (Bresnahan, 1998). If successfully done, this creates a strong market position where the virtuous cycle of positive feedback begins. Sellers and buyers tend to make long term platform specific investments that leads to platform longevity and provides excellent barriers to entry against competition (Bresnahan, 1998). In two-sided platforms, another virtuous cycle between the number of sellers and buyers has to be commenced and maintained. Both sides of the platform stimulate growth in each other and therefore, companies must figure out which one to price and which one to subsidize in order to stimulate growth (Eisenmann, Parker & Van Alstyne, 2006). These mechanisms make virtuous cycles easy to maintain but hard to start and stop. Bonardi and Durand (2003) states that a dominant platform who operates in a market benefiting from network effects can be defined as an industry standard in terms of the specifications of the platform once it has aggregated the critical mass of users (Bonardi & Durand, 2003). Typically, only a few players can co-exist in each market competing on the same standard (Bresnahan, 1998).

Defining a standard is a battle between two or more incompatible technologies for market share where the products cater to the same need (Shapiro & Varian, 1999). This battle can be referred to as a standards war that can determine the survival of a company (Shapiro & Varian, 1999). Standards wars often end in one of three outcomes: a truce where a combination of the technologies are adopted and shared, a duopoly where more than one standard is used, or a fight to the death with one victorious firm. The last alternative is prevalent in markets with strong network effects and positive feedback loops (Shapiro & Varian, 1999).

Gallaugher and Wang (2002) state that it is not always the best product that succeeds in building a network that generates a virtuous cycle. This condition stems from the effects generated by network externalities that increases the utility of a product with the size of the network making the product and the size of the network equally important in the pursuit of market victory (Katz & Shapiro, 1985; Shapiro & Varian, 1999). As an example, the QWERTY keyboard standard was developed in the late 19th century because of its ability to avoid typewriter typebars jamming and clashing if struck in a fast succession which was a large problem at the time (David, 1985). This problem was later removed when the typebars were exchanged to down-stroke, forward-stroke and electric typewriters. Competing standards began to emerge promising faster typing and added efficiency. One of them placed the letters DIATHENSOR on the home row which enabled typists to write 70 percent of the words in the English language with these ten letters. However, when individuals and companies chose what standard they would learn, most chose the QWERTY keyboard because of the proportionally large availability of hardware effectively placing other standards behind (David, 1985). This example signifies the importance of network size of a product.

Sometimes companies collaborate on a standard, developing products and applications that integrate with each other allowing users to communicate with a user base that use goods from different firms (Katz & Shapiro, 1985). Companies often employ this strategy in order to reach the critical mass rapidly at which the network can function without artificial help. For example, the mobile phone operating system Android is shared among firms including but not limited to Samsung, Huawei, Nokia and Blueberry. This means that app developers can sell their apps to users regardless of hardware developer. Companies become less inclined to share a network user base the larger they become even though public welfare increase by sharing (Katz & Shapiro, 1985).

Another dimension of competition that play a role in network markets is the departure from the principle of diminishing marginal productivity (diminishing returns). Competition under diminishing returns means that a market over time will experience decreasing returns constituted by squeezed margins either from price pressure or from rising costs (Brue, 1993). The theory states that marginal output will decline as a single factor of production is linearly increased in the production process (Brue, 1993). Thus, a market equilibrium given by price and quantity can be easily predicted (Arthur, 1996). Arthur states that competition in network markets with network-based businesses shift economic behaviour from an environment of diminishing returns to one of increasing returns. In an increasing returns milieu, price and quantity is not determined on the same terms as in a diminishing returns environment. Rather, because of network effects and what Arthur calls customer groove-in (defined as the costs in time and learning a user spend to enable use of a product) products become more attractive and offers higher value to the customer when they age, experience positive feedback virtuous cycles and adds individuals to their user base.

What this means for competition is that products that gain advantage will probably gain further advantage and firms losing advantage stand to lose further advantage (Arthur, 1996). Gaining advantage will in most cases lead to a larger market share which is according to Gallaugher and Wang (2002) the single most important driver of competitive price advantage in these types of markets. As a firm's market share grows larger it will generate more network externalities and thus be able to charge a higher price (Gallaugher & Wang, 2002). Kolasky (1999) adds that increasing returns are especially evident in markets that have low variable costs, as many digital products and services do. He uses the example of software to illustrate this mechanism where the development cost is independent from the amount of software sold and therefore the scalability becomes theoretically unlimited.

In this tough setting of competition, companies need to use non-traditional strategies to entry a market. One strategy is to out-innovate the competition offering a radically superior product or service, defined as Schumpeterian innovation (Eisenmann, Parker & Van Alstyne, 2011; Shapiro & Varian, 1999). Providing a superior solution and managing consumer expectations leads to an opportunity to define a new standard in the market. In order to enter as a new player, Shapiro and Varian (1999) defines that innovation capabilities with short product design cycles are imperative. Additionally, the company needs to manage consumer expectations to profile its own standard as the winning one (Shapiro & Varian, 1999). Pursuing these two strategies gives the challenger a good chance of coming out on top.

A second strategy is through what Eisenmann, Parker and van Alstyne (2011) refer to as platform envelopment. The foundation of this idea is that companies having a position in a platform market can envelop a platform user base in an adjacent market through developing a platform with similar functions as the target platform. The authors develop three different scenarios: (1) two platforms where the user bases are largely overlapping and the platforms complement each other, (2) platforms are weak substitutes and user bases has a small overlap, (3) the platforms are unrelated and there is an asymmetric relationship in the overlapping of users (see figure 2.2) (Eisenmann, Parker & Van Alstyne, 2011).

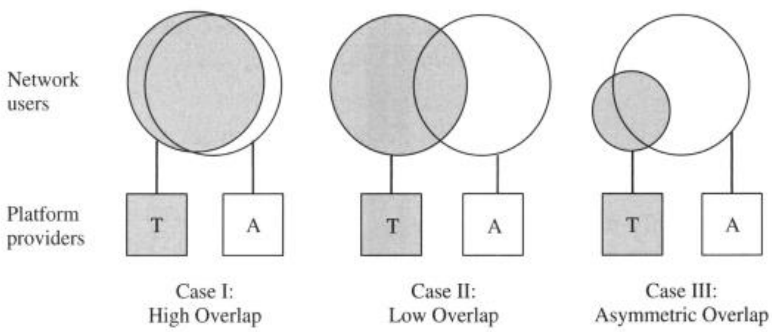


***Figure 2.2** "User base overlap between attacker (A) and target (T) platform" (Eisenmann, Parker & Van Alstyne, 2011, p. 1278)*

Under the first scenario of market condition, the challenger can develop a product with the same functionality as the target and bundle the products, pricing them near the sum of the platforms sold separately. Under the second market condition scenario, the challenging company can develop a similar product as the target and bundle the two platforms if the addition offers significant economies of scope. If that condition is met, offering the two products at a deep discount relative to their optimal stand-alone price will offer a good chance of success. Under the third market condition scenario, the challenger can be successful if the target user bases overlap significantly with a part of the attacker user base and the additional product offers high economies of scope (Eisenmann, Parker & Van Alstyne, 2011). Common for the three scenario strategies is the strategy of bundling products that effectively limits the targets access to overlapping customers (Eisenmann, Parker & Van Alstyne, 2011). Even if this approach offers a way that does not imply radical innovation and leapfrogging incumbent players, the strategy is primarily applicable in large companies already enjoying success in a platform market. The examples that are given in the article to a large part include Google, Apple and Microsoft enveloping smaller platforms in adjacent markets. Thus, what this points to is yet another dimension of competition that favours scale of network, large user base and companies possessing a large market share.

2.4 Estimating the Value of the User Base

Stephen and Toubia (2010) compared two-sided mutually reinforcing platforms allowing sellers to interact with other sellers to platforms that only allowed interaction between sellers and buyers. They found that all types of connections contribute significantly to revenue creation. When sellers could connect with each other in a platform to create paths and interconnections, the buyers had a simpler time navigating between virtual shops to find the perfect suited product to their needs. Stephen and Toubia likened an e-commerce platform that allow connections between sellers to a shopping mall in contrast to geographically dispersed shops where customers have to travel a longer distance in order to compare and look at different products. The study concluded that all types of user connections contribute significantly to revenue generation.

Other attempts that try to put an actual value on a network has been made by a number of scholars. The most well-known laws define a proportional relationship between user base and value using revenue as proxy. The most renowned of the laws defining value of network effects is Metcalfe's law. Metcalfe (2013) states that the value of a network is defined by the number of connections that are possible within it. Therefore, he reasoned that the value of a network should be proportional to the user base to the power of two (see appendix A).

Odlyzko's law was derived initially by Odlyzko and Tilly (2005) as a response to Metcalfe's proposed method of approximating the value of a network. In their article they outline heavy criticism to the properties of Metcalfe's law and offers a contesting view. Odlyzko and Tilly denies the implied assumption of Metcalfe that all connections in a network has the same value and that therefore, the value calculation should not reflect such an aggressive assumption. Instead, Odlyzko and Tilly argue that the marginal utility of added possible connections are diminishing. In addition, to back up their argument, they reason about what the equal value of connections would mean in the market. They mean that if that was the case, networking companies would merge to a larger extent that what they do today. They mean that, with the logic of Metcalfe, the value of any networking company is n^2 and thus the value of two stand-alone networking companies is $2n^2$. If the two companies were to merge under the effect of Metcalfe's law the user base becomes 2n and the value would double without adding any users since $(2n)^2 = 4n^2$. Odlyzko and Tilly establish that networking companies do not exhibit the high merger rate that this would imply. Therefore, the authors offer the derivation of network value as proportional to the user base times the log user base (see appendix A). Under this law the marginal value of an added connection diminishes with size. This also means that if companies were to merge the value increases merely five percent which would explain the low rate of mergers, acquisitions and interconnections of networks (Odlyzko & Tilly, 2005).

The third law found in the literature estimating network value is Reed's law that approximates the value more aggressively than Odlyzko and Metcalfe. Reed states that the value of a network is proportional to the number of subgroups that can be formed within the network (Reed, 1999). Assuming that the smallest group consists of two users, the value of a network would then be proportional to two to the power of the user base (see appendix A).

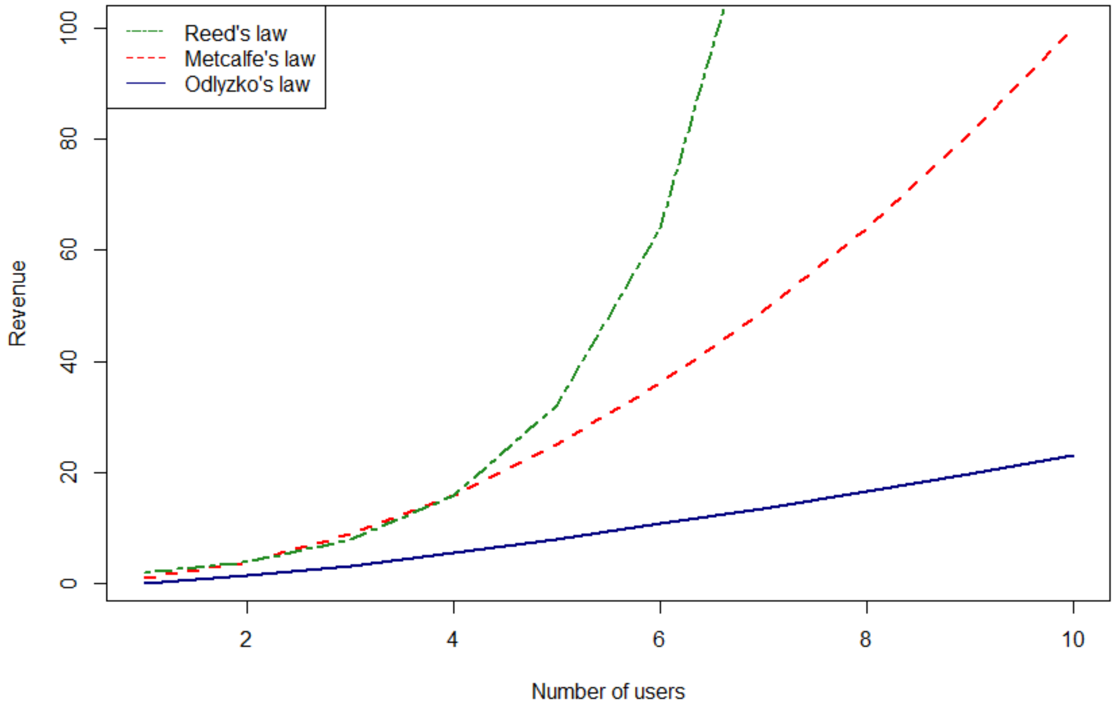


Figure 2.3*: Visualisation of network value development under Metcalfe's, Odlyzko's and Reed's law. Hypothetical values.*

A pervading problem with the laws represented in the literature is that they have not been empirically tested. Visual comparisons have been made regarding the three laws compared to the revenue development of two companies, Facebook and Tencent, however, no statistical method was used to evaluate the results of the analysis (Metcalfe, 2013; Zhang, Liu & Xu, 2015). Additionally, Metcalfe, Reed and Odlyzko's prior test has never been applied to more than two companies.

Given the theoretical exponential relationship between revenue and user base set out by Metcalfe, Reed and Odlyzko, an exponential function should explain the variance of revenue regressed on user base more efficiently than linear function. Our first hypothesis thus becomes:

> *Hypothesis 1: The association between revenue and user base in platform mediated companies can be better explained by an exponential relationship than a linear relationship.*

2.5 Platform-Mediated Companies

Schreieck et al. (2016 in Schweiger et al. 2016) state that there are two approaches a platform-mediated company can take, either technology oriented or market oriented. The technology-oriented perspective reflects the approach taken by companies that provide technological building blocks that make up a part of other companies offering (Schreieck et al. 2016 in Schweiger et al. 2016). Examples of this approach are Amazon web services, Alibaba cloud or Google Vertex AI who make up important parts of other companies' infrastructure. The unique selling point of these companies is their consistent focus on offering the most advanced infrastructure on the market to the best price, often dispensed via software-as-a-service models (SaaS) (Ojala, 2013). The market-oriented approach reflects companies focusing on developing a software ecosystem putting user interactions in focus by enabling as many connections as possible on a common platform (Eisenmann, Parker & Van Alstyne, 2011). This paper will focus on market-oriented platform-mediated companies.

Market oriented platform-mediated companies operate in different industries. According to Koller, Goedhart and Wessels (2015), the median return on invested capital (ROIC) shifts vastly between industries. The authors calculated the median industry ROIC by collecting data on companies between 1965 and 2013. The results yielded varying ROIC in different industries. Pharmaceuticals and IT-services were some of the most profitable industries and were able to achieve a significantly higher ROIC by operating in industries with high barriers to entry, evident by for example patents according to the authors. Other aspects that formed varying ROIC were customer lock-in, loyalty to brands, increasing return to scale and sustainable competitive advantages. High ROIC companies were able to lock in customers by achieving strong competitive advantages that resulted in increasing returns. Low ROIC companies on the other hand, tended to struggle with achieving a price premium or cost advantages evident by the worst performing industry being utilities. The final aspect reflected on by Koller, Goedhart and Wessels (2015) was the extreme variety of median ROIC within the industries themselves. Low ROIC industries tended have a low varying median between quartiles in comparison to high ROIC industries. Pharmaceuticals for example had a median variation about ten times that of utilities (Koller, Goedhart & Wessels, 2015).

This paper focuses on digital finance, social network and e-commerce industries as platform-mediated business models are overrepresented in these categories. Digital finance companies have transformed the way banking is managed. Many of the firms having enjoyed success in this industry have effectively managed to disintermediate the markets with lower transaction costs as result, provided peer to peer lending platforms that enable lending without going to a bank, and developing digital financial advisors that can advise individuals in their investment and private finances situation at a low cost and at any time (Dhar & Stein, 2017). In the digital finance industry, business model innovation is an important driver of revenue as incumbent players often struggle to keep up with cutting edge solutions. This generates a better functioning platform that builds user base which makes it gain from network effects as the amount of trading information that a user can extract increases with the amount of capital under management on the individual platform (Velu, 2015). With the same motivation, interactions between platforms increase their chances of survival in the market (Velu, 2015). Revenue is mostly generated by fees and interest stemming from the use of products and services offered depending on company.

Social networks derive value through offering a new way of networking as opposed to the traditional means of social interactions. Enders et al. (2008) apply the concept of the long tail, previously used to describe e-commerce businesses, to the business of social networks. By doing that, they state that social networks let us make connections with people who we would never have met in real life. More importantly, we can keep in contact with people we have developed weak ties to as opposed to the strong ties that humans develop to their closest friends. Granovetter (1973) states that these weaker ties are used far more than strong ties when it comes to job search and advice-seeking. Thus, social networks create value for their users by letting them maintain weak ties contacts that can potentially become valuable in the future (Enders et al. 2008). Enders et al. (2008) discuss different revenue models in social network companies. They state that social networking sites (called SNS) generally can go about value capture through deploying an advertising model, subscription model or a transaction model. Often SNS use more than one of these models on their platforms. The advertising model, also the most common one of the three since users in most cases demand cost free platform use, relies on a large user base in order to enable targeted marketing of third-party goods to a subset of the user base. Thus, the company has to have a large user base in order to be able to offer advertisers a large enough segment of their users that corresponds with the target group of the advertiser. The subscription model relies on the value of the offer made on the platform. The larger the value, the larger the consumers’ willingness to pay for it. It generates revenue through a subscription fee paid by the user on a regular basis. A transaction model is deployed by some SNS acting as market intermediaries connecting buyers and sellers. This model relies on the reputation of the platform and the trust consumers have in it. The SNS takes a transaction fee on each good sold (Enders et al. 2008).

The fundamental property that differentiates e-commerce from regular retail is that the product offering lies on the long tail (Brynjolfsson, Hu & Smith, 2006). The long tail describes products in a product category that are unpopular and sell relatively bad in comparison to the popular products. This makes them unprofitable for a physical retailer to carry in their assortment. However, since e-commerce platforms enable matchmaking between sellers and buyers over large geographical distances, selling the long tail becomes a viable business strategy. Brynjolfsson, Hu & Smith (2006) states that trade patterns on e-commerce platforms are much less centralized around popular products compared to physical retail stores and that they are distinguished by the sale of obscure products. As such, the value creation potential in e-commerce firms rely on four factors: complementarities, efficiency, novelty and lock-in according to Amit and Zott (2002). Complementarities refer to bundled products from which a customer can derive more value from if he owns them both rather than separately. E-commerce platforms can offer complementarities as products or services to the product segment that is sold on the site. For example, Amit and Zott bring up the example of the online travel booker 'E-booker' that offer weather, currency exchange information and contact to vaccinators on their site. Efficiency refers to the decline in transaction costs, defined broadly as time consumed, searching costs and information asymmetry. The smaller the transaction costs, the grater the efficiency. Novelty means that the company either address consumer needs that are latent or create new markets for previously not traded products. Lock-in refers to the ability of the company to motivate their users to come back. The majority of E-commerce companies use the transaction revenue model outlined in the preceding paragraph. They act as market intermediaries that connect buyers and sellers in order to facilitate commerce in a match-making process (Laudon & Traver, 2007 in Enders et al. 2008).

Platform-mediated business models are employed in a broad range of settings. Furthermore, the literature determines that different industries employ different monetisation and business strategies to increase either user base or revenue. This mean that some industries should exhibit a higher association between revenue and user base as some business models are more efficient at extrapolating revenue than others. However, as revenue models can differ within industries as well, the differences should not be severe. Thus, we develop our second hypothesis:

> *Hypothesis 2: The impact of user base on revenue in platform-mediated companies varies between the e-commerce, digital finance and social network industries.*

3 Methodology

3.1 Research Design

The study follows a hypothetico-deductive design. A deductive approach is “developing a hypothesis (or hypotheses) based on existing theory, and then designing a research strategy to test the hypothesis” (Wilson, 2010, p.7). Bryman and Bell (2011) states that the deductive approach typically represents a very common view in regard to the relationship between theory and research. A hypothetico-deductive method according to Bougie and Sekaran (2020) provides a scientific approach for generating knowledge. The authors also state that such a method tests a theory with specific observations which ultimately can be used to reject or not reject the null hypothesis. The purpose with the study was to test two different hypotheses which have been derived from well-established theories of economics and management studies. These include network effects, Metcalfe (and the alternative) laws, standards, shareholder value approach as well as increasing returns. Existing theory has had a central role in the formulation of the hypotheses, and why the study is conducted from the hypothetical-deductive approach.

A cross-sectional study was formulated, in order to test the hypothesis. A cross-sectional is according to Bougie and Sekaran (2020) a time horizon study where data is gathered at a single point in time. A cross-sectional study is conducted in order to quantify and analyse the association of two or more variables (Bryman & Bell, 2011). A problem with a cross-sectional study however, is that according to Bryman and Bell (2011) it can only detect patterns of association between two variables and not any causal relationship. Longitudinal studies on the other hand are according to Bougie and Sekaran (2020) studies in which data points are collected over a set period of time. These studies are according to the authors, popular when tracking a certain factor over a period of time, such as advertising effectiveness. The purpose of the study was to test the hypothesized exponential association between revenue and user base. A cross section study was therefore deemed the most appropriate as it analyses the association between variables. However, as stated before a cross sectional study only looks at the association and therefore no inference about the causal relation can be drawn. Therefore, the study can only draw conclusions regarding the association between revenue and user base and not an inferred causal relationship.

3.2 Data Collection Method

In order to answer the research question and test the hypotheses, data in regard to revenue and user base have been compiled from a variety of primary and secondary sources. In total 85 companies have been selected for the study, formulated by a set of criteria. The information regarding revenue and user base has been collected from either a primary or secondary source. Primary sources include annual, quarterly and investor relations reports and is deemed reliable as it comes directly from the reporting firms. Secondary sources primarily include estimates done by analysts from a variety of sources such as Slack's monthly active users complied by GP Bullhound. The data points taken from secondary sources have been limited to high-quality sources which include the likes of GP Bullhound, Statista, Forbes and Reuters. However, as in all cases with secondary data there will be a larger residual risk of skewed result as the estimates calculated by analysts can be incorrect. However, taking into consideration the large number of over 250 data points, collected on 85 companies, the estimates should not significantly skew the results or effect the study negatively.

In some cases the primary sources could not be used as the reporting firms used other metrics than what is used in this study. User base for example is typically measured in MAU, DAU or YAU and there is no industry standard to measure active users. Therefore, some primary data observations hade to be ignored and taken from secondary sources in order to keep the metric unchanged.

3.2.1 Selection of Firms and Sampling

The cases incorporated into the regression were collected as a sample from the larger population. It was therefore of great importance to obtain a sample size that can reliably estimate the population parameters. This is because a sample can make statistical inferences about the population (Berenson, Levine & Szabat, 2014). There are two main components of sampling that according to Berenson, Levine and Szabat (2014) increases its validity, these are sample size and distribution of samples. Having a sample mean equal to that of all possible sample means (sampling distribution of the mean) creates a sample that is unbiased. Moreover, a larger sample size decreases the error of the mean by a factor that is equal of the square root of the sample size (Berenson, Levine & Szabat, 2014). The first step in constructing the sample was therefore to create a sample frame that data could be drawn from. In this study the sample frame was formulated by a set of criteria. These criteria are as follows:

- Operate a platform-mediated business model.
- Serve an international audience.
- Conduct operations for the year 2019.
- Companies have a market-orientated approach.
- Be active in the e-commerce, social network or digital finance industries.

These criteria needed to be fulfilled in order minimize the residual risk of certain data points disproportionally influencing the outcome. The idea behind the criteria was to make sure only firms operating in the appropriate environment was incorporated into the study. Each case needed to be proportional to the population and because the studies tests for the association between revenue and user base, each case needs to be appropriate to minimize the risk of skewed results. The first criteria were constructed in order to filter out hybrid business models as the study only tests platform-mediated business models. The second criteria were introduced to remove firms operating solely on a domestic level in order to provide fair and comparable statistics. The third criteria was added as the cross-sectional study takes its data from a certain period of time. This meant that all cases in the sample had to have active operations during the year 2019 in order to be eligible for the study, as this was the most appropriate time period. This meant that firms founded later than 2019 were not eligible. Furthermore, platforms that closed down before the end of 2019, such as Google plus, who ceased operations in the second quarter of 2019 (Google, n.d.) were not eligible for the sample. The fourth criteria states that the company must have a market-oriented approach in contrast to a technology-oriented approach as outlined in chapter 2.5. The last criteria that needed to be fulfilled was that the firms needed to be active in one of the three tested industries. These industries are present in order to test the significance of an industry variable and in turn see if a specific industry was more reliant on its user base for turnover growth than another. These three industries will be defined and discussed in more detail below. After the sampling frame was formulated, the sample needed to be collected. In the case of this study, all cases that had reliable public information were added into the sample. In total, 85 companies were deemed reliable and therefore incorporated into the sample. The relatively small sample size has consequences for the result in terms of reliability.

E-Commerce

E-commerce is according to the Cambridge dictionary, "the business of buying and selling goods and services on the internet" (Cambridge Dictionary, 2021a, n.p.). For this study, online retailers, food ordering/delivering platforms and rating platforms have been defined as e-commerce. This category will be formed by firms such as Wish.com, Etsy and Trivago who all operate by using a digital platform that sells a good or service via the internet. A caveat however, is that the study is limited to firms who operate based solely on a digital platform-mediated business model (not a hybrid). Therefore, companies that do not separate online and physical users or online and in-store sales, will not be included in this study. These kinds of firms include Macy's, Walmart and Best Buy even though they are the largest online e-commerce retailers in the US (Peters, 2021). In total, 30 companies were included in the e-commerce category.

Social Network

A social network is according to the Cambridge dictionary "a website or computer program that allows people to communicate and share information on the internet using a computer or mobile phone" (Cambridge Dictionary, 2021b, n.p.). For this study, social media platforms, messenger applications, live streaming services and interactive gaming platforms have been defined as a social network. This industry will therefore include firms or platforms such as Instagram and Twitch. This is because all of the above-mentioned examples operate a platform where communication and information sharing is key to the business model. However, a general problem with these platforms/companies is that they are in some cases part of a conglomerate or parent company. This means that the revenue is not specific for the platform as it is generally a part of a parent company's income statement. Therefore, platforms such as QQ (Chinas second largest messenger application) have been removed from the sample (Thomala, 2021). This is due to the fact that QQ is a part of Tencent who reports the MAU, but not how much of its revenue is attributed to QQ. In some cases however, parent companies do specify how much of its revenue is generated from a specific platform or this information is accessible from a credible secondary source such as in Statista's Instagram estimates (Tankovska, 2021). If this is the case, then the firm is eligible for the study. In total, 28 companies were included in the social network category.

Digital Finance

Digital finance is according to the European Commission (EC) a "term used to describe the impact of new technologies on the financial services industry" (European Commission, n.d. n.p.). The EC continues to state that digital finance includes products, applications and business models that have transformed traditional banking and financial services (European Commission, n.d.). For this study, digital payment, online trading, digital financial services and Fintech platforms have been defined as digital finance. This includes eToro, PayPal, Revolut, Venmo and Apple Pay. These examples fit the criteria due to the fact that all have implemented new technologies to the financial service industry such as user-based platforms. However, a general problem with these types of companies is that they tend to be a part of a parent company or heavily funded by venture capital. Just as in the aforementioned section on social network, digital finance companies are limited in available data. Therefore, the firms needed to have available data in regard to user base and revenue in order to be incorporated into the study. For example, large incumbent banks and institutions do not report what part of the revenue is directly generated from a digital platform. This means that digital banking platforms operated by the large incumbent banks could not be included in the study as the data points could skew the results by over or underrepresenting the revenue and/or user base. In total, 27 companies were included in the digital finance category.

3.2.2 Selection of Variables

As the study was conducted using three different types of regression, outlined in 3.3.1, the independent and dependent variables needed to be defined. In this study three variables have been chosen in order to test the hypotheses. The variables that will be used in the regression are as follows:

- Revenue (represented by net income)
- User base (represented by monthly active users)
- Industry (a dummy variable to test for industry variability)

These variables were deemed the most appropriate to use in the regression in order to test the research question. The following section will explain the rationale behind the variables and how each is quantified.

Dependent variable: Revenue

In all the given tests the value of the firm will be represented by its net revenue. The rationale behind the variable stems from the theory of shareholder valuation. This theory states that shareholder value increase when a company achieves a higher return on invested capital than its weighted cost of capital (Rappaport, 1986). This strategy can be obtained through two primary strategies:

- Revenue growth
- Increased capital efficiency

Revenue was deemed to be the most suitable variable evident by a set of different factors. Firstly, revenue is the variable that has been used in previous research to represent network value (Metcalfe, 2013; Zhang, Liu & Xu, 2015). Therefore, the test conducted in this paper will be comparable to previous research by using the same variables. Secondly, compared to increased capital efficiency, revenue is more easily quantifiable since capital efficiency could be affected by alternative accounting standards (Rappaport, 1986). Moreover, revenue has been found most closely connected to the positive externalities generated from network effects and has therefore been assumed to be the best proxy of value.

Market capitalisation was a variable that could have represented value; however, it came with complications that eventually led to it being removed from consideration. A general problem with international tech companies that work with a user base business model, is their tendency to be private. Fintech firms are for example generally highly financed by venture capital and therefore are reluctant to go public in the early stages. Furthermore, a large proportion of these types of firms are subsidiaries which makes it hard to estimate a stock price. The stock value would exist internally, but this kind of information is not given to the public. Therefore, information regarding market capitalisation is extremely limited and the majority of data points would have to be removed. Furthermore, as stated above, all previous research has used revenue as a proxy for value. Therefore, revenue would be the optimal variable to represent value as it makes this study more comparable to previous literature and more data points can be incorporated.

Independent variable: User base

The first independent variable that will be incorporated into the study is user base. A user in our study has been defined as a customer who also generate positive network externalities or an individual who uses the platform without generating direct cash flows to the company itself. Therefore, a user can both be an individual active on a social network or a customer on an e-commerce platform that benefits from the network effects created by the user base. However, users on platforms that incorporate a one-sided business model that does not generate positive network externalities are not included in the study. This is because each new user on a platform does not increase the utility of the other users as no connections are made.

The size of the user base is usually quantified in three different manners, daily active user (DAU) monthly active user (MAU) or yearly active users (YAU). An XAU is defined as a user or customer who has made a purchase, logged in, sent a message, etc. depending on the functions of the platform at least once in the given time frame. The study has used MAU as the variable for user base in order to have comparable results across all data points. MAU can be quantified in two different ways. In a one-sided business model that generates positive network externalities, the MAU is defined as the number of users on the platform. In a two-sided business model however, such as with the e-commerce platform Etsy, one needs to take into account the number of sellers on the platform as well. Therefore, both the buy and sell side have been combined in order to get the total amount of users.

Independent variable: Industry

The industry variable was measured over three of the largest industries where platform-mediated business models have emerged. These three are e-commerce, social networks and digital finance as defined and described in 3.2.1. The industry variable is measured in order to determine industry variability in terms of the user base's impact on revenue. The industry variable was defined as two dummy variables in order to allow for variance between the different categories (Princeton University Library, 2007; James et al. 2013). E-commerce have been deemed the most appropriate reference variable since this industry contains the most observations out of the three in the dataset (James et al. 2013). Thus, the first dummy variable gives a parameter value for social media and the second for financial services. Discarding the dummy variables in the result gives us the values for the e-commerce category. The variables are defined as follows:

$$D_{SN} = \begin{cases} 1: Social\ network \\ 0: E - commerce \end{cases}$$

$$D_{DF} = \begin{cases} 1: Digital\ finance \\ 0: E - commerce \end{cases}$$

3.3 Data Analysis

3.3.1 Multiple Linear and Exponential Regression

For this study a multiple linear regression was conducted to test all of the three hypotheses. James et al. (2013) describes that a linear regression is a useful tool when one tries to quantify a predicted response. The authors state that a regression is an approach that has stood the test of time and is the topic of numerous textbooks. A linear regression according to the authors, simply tests for association between an independent and dependent variable(s). A linear regression can be split into two categories: simple and multiple. A simple linear regression is a straightforward approach that quantifies a predictive response for Y on the basis of a single variable, X (James et al. 2013). James et al. stated that a multiple regression on the other hand tries to quantify a predictive response for Y one the basis of more than one X. The authors state that one could theoretically conduct a simple linear regression for each variable. However, this tends to come with two different problems. Firstly, it is hard to make a general prediction of Y given three different regressions since each dependent variable is associated with a different regression (James et al. 2013). Secondly, the authors state that each regression ignores the other when estimating the regression coefficients. Therefore, a multiple linear regression is the better alternative when an association between an independent and more than one dependent variable ought to be examined. See appendix A for a multiple linear regression's general equation.

For hypothesis 1, the goal was to examine whether the data follows a linear or exponential association. Hence, in addition to the multiple linear regression, an exponential regression was used to test for an exponential association. An exponential regression is similar to that of the linear regression in terms of fit and output, however is used to test a non-linear relationship that is diminishing or appreciating at a certain rate (Abramson, 2021). In appendix A the general equation is outlined.

For this study, four values computed from the multiple regression was of importance. These four values were p-value, F-stat, regression coefficients and R squared (R^2). P-value test for significance and is used to reject or not reject the null hypothesis. A small p-value indicates that there is an unlikely probability that the association between variables is random (James et al. 2013). The authors state that a p-value less than the defined alpha value leads to a significant result and the null hypothesis can be rejected. A high p-value signals the opposite and therefore one cannot reject the null hypothesis. Secondly, an F-stat works in par with a P-value and is used to reject or not reject a null hypothesis. An F-stat higher than the F significance shows that there is strong evidence against the null hypothesis (James et al., 2013). The third value of importance is the regression coefficients β_p. This value measures the relationship between two variables. The last value of importance was R^2. James et al. (2013) state that this value shows how much of a change in one variable is attributed to a change in the other. A value close to 1 indicates a very strong relationship and a value close to 0 shows the contrary. An analysis of these tests will be presented in the results and discussion.

The confidence intervals of the regression coefficients were computed to analyse the spread of the possible population mean of each variable. A confidence interval is according to James et al. (2013) a range of values that will contain the population mean with a given percent of certainty. The range of certainty often range between 90 and 99 percent depending on severity of an incorrect prediction. The authors continue to state that the range is defined in both lower and upper terms computed form the data. Therefore, the interval will contain the correct prediction of the regression coefficients with the given percent of probability (James et al. 2013). All tests were done with at least 90 percent confidence.

3.3.2 Hypothesis Testing

The data collected was analysed using three observations per company, a variable for revenue, user base and a dummy variable for industry. This information has been collected from the above given sources and computations made in R. In order to test the hypotheses we had to develop four regression models, regression 1 and 2 to test for the user base to revenue relationship and regression 3 and 4 to test for industry significance.

Regression 1 $R = \alpha + \beta_1 UB + \varepsilon$

Regression 2 $R = \alpha e^{\beta 1 UB} + \varepsilon$

Regression 3 $R = \alpha + \beta_1 UB + \beta_2 D_{SN} + \beta_3 D_{DF} + \varepsilon$

Regression 4 $R = \alpha + \beta_1 UB + \beta_2 D_{SN} + \beta_3 D_{DF} + \beta_4 (UB * D_{SN}) + \beta_5 (UB * D_{DF}) + \varepsilon$

<u>Where:</u>

R = revenue

α = intercept

e = Euler's constant

β_p = coefficient value for variable

UB = user base

D_{SN} = Categorical variable for social network*

D_{DF} = Categorical variable for digital finance*

$UB * D_{SN}$ = Interaction between user base and social network dummy

$UB * D_{DF}$ = Interaction between user base and digital finance dummy

ε = Residual error

*E-commerce lies as reference variable/baseline for dummy coding.

In concurrence with the two hypotheses outlined in chapter 2, the following three tests were designed. The first test investigates weather an exponential association describes the data better than a linear association. Thus, Akaike information criterion (AIC) is a relevant measure. The AIC evaluates models and determines that the model explaining the greatest amount of variation with the fewest possible predictors is the better one (Hu, 2012). AIC calculates this and returns a number based on the number of predictors in the models and the maximum likelihood of the model (Hu, 2012). A model with an AIC score of 2 below the comparable model is generally considered as significantly better. The following hypothesis was thereby formulated:

H₀; $AIC_{regression\ 1} - AIC_{regression\ 2} \geq 2$

Hypothesis 1 **H₁;** $AIC_{regression\ 1} - AIC_{regression\ 2} < 2$

The results from test 1, outlined in the succeeding chapter, did not yield a significant result for either model, however the linear model scored lower than the exponential model. Therefore we chose to test industry significance on the linear model. To test for significance of all three categories in the dummy variable, an F-test between regression 4 and regression 1 was performed and the p-values of the individual dummy variables were investigated. A significant difference between the regression models combined with individually significant dummy variables conclude an industry difference in terms of model intercept. The hypothesis is formulated as follows.

H₀; $\beta_2,\ \beta_3\ and\ \beta_2 - \beta_3\ found\ insignificant$

Hypothesis 2.1 **H₁;** $\beta_2,\ \beta_3\ and\ \beta_2 - \beta_3\ found\ significant$

To check for differences in slope (interaction effect), i.e., user base to revenue behaviour depending on industry, an F-test on the interaction variables was performed. We performed this test between regression 4 and 3 in combination with investigating the p-values for the individual interaction variables. Through rejecting the null hypothesis we conclude that there is a significant difference in terms of added revenue per added user:

H₀; $\beta_4,\ \beta_5\ and\ \beta_4 - \beta_5\ found\ insignificant$

Hypothesis 3.2 **H₁;** $\beta_4,\ \beta_5\ and\ \beta_4 - \beta_5\ found\ significant$

After the regressions and hypotheses were defined, we performed the four tests of normality:

- **Linearity:** Tests for a mean relationship between X and Y and makes sure it is linear.
- **Homoscedasticity:** Looks at if the residual variance is equal for all X.
- **Normality:** Looks at the fixed values of X and Y to see if they are normally distributed.
- **Multicollinearity:** Test to see if the observations are independent form each other.

(Boston University, 2016)

The tests yielded poor results in terms of homoscedasticity and normality (see appendix B). A Cullen and Frey graph was plotted in order to find the residual distribution of the regression (see appendix C). It was found that transforming the continuous variables to log normal products could help the analysis. After transforming revenue and user base with log normal the Cullen and Frey graph indicated normality in the variables (see appendix C). The normality assumptions were tested once again and showed much better results (see appendix B). However, there were still a slight issue with the normality of residuals as indicated in the QQ – plot. To solve this problem the regressions have been bootstrapped with 1000 samples in order to give more reliable confidence interval. Bootstrapping a model means to generate 1000 estimates of the coefficients using slightly different samples (James et al. 2013). The 1000 samples are generated through sampling the dataset with replacement with a sample size equal to the sample size of dataset, in this case 85 (James et al. 2013). This increases the validity of the model output and especially the confidence intervals. Bootstrapping the regressions cross validates them at the same time which increase the probability of a good out-of-sample fit (James et al. 2013). After this, coefficient values, confidence intervals, ANOVA tables and F-tests have been computed and plotted.

3.4 Validity and Reliability

3.4.1 Validity

The validity of a paper can according to Bougie and Sekaran (2020) generally be split up into two components: internal and external validity. Internal validity is according to the authors concerned with the authenticity of the cause-and-effect relationships. Bryman and Bell (2001) state that internal validity mainly relates to the issue of causality. The authors continue to state that internal validity looks at relationships between two or more variables and if the relationship actually holds water. If an association between and X and Y is inferred, one cannot be sure that X is responsible for the change in Y as it may be some other untested parameter that affects the association (Bryman & Bell, 2011). In terms of internal validity this research paper scores low. This is primarily due to the fact that one cannot draw casual relationships in a cross-sectional study, as mentioned in 3.1.1. Even if this study can argue for an association between user base and revenue it cannot argue for any causal relationship between the two. The study has however gone to great lengths to reduce any residual risk of skewed result as argued above. Therefore, the lack of internal validity is not related to the quality of the paper but rather the chosen method to test the hypotheses.

Bougie and Sekaran (2020) state that external validity looks at the generalizability of the cause-and-effect to the external environment. In other words, how well can the study be applied to other situations outside the context of the study. This is a parameter that the study yet again fell short in. The primary focus of the research study was to test the association between user base business and revenue. Therefore, it will be hard to apply the study to any other situation than the one referred. One could argue that the study can be applied to in-store purchases as well by looking at the number of customers. However, more research will have to be conducted into that specific field. Furthermore, as the study solely focused on digital business models it cannot be applied to hybrid once such as Walmart, even though the digital portion could be valued. This is because Walmart both incorporates in-store and online sales into its business model.

3.4.2 Reliability

The reliability of a study according to Bougie and Sekaran refers to the consistency of observations (2013). The author reflects upon the fact that a reliable study should obtain the same results regardless of observer and occasion. In other words, a reliable study should get the same results regardless of when the study is conducted or who is conducting it. In this regard the study presented in this paper is deemed reliable. The data points are collected from sources that would generate the same results regardless of observer or time period. Furthermore, the quantitative information is seldomly contradicting in regard to which secondary source is used as the majority of information comes from the same originating source. This means that even if other secondary sources would be used to gather the data points, the results would remain the same. One should note however that the results could come to differ in the future if more reliable estimates would be implemented from a larger dataset. Moreover, private information could be disclosed in the future making some secondary sources redundant. This in turn means that future studies could get slightly different results if estimated data points would be replaced. However, in term of the regression output the result should stay similar due to the magnitude of data points and low variance of secondary source estimates.

4 Results

Corr. matrix	Revenue	User base	Revenue log	User base log	Social Network	Digital finance
Revenue	1.0000	0.4944	0.7336	0.4676	0.02527	-0.1240
User base	0.4944	1.0000	0.4778	0.6717	0.35075	-0.1257
Revenue log	0.7336	0.4778	1.0000	0.6834	0.15189	-0.3421
User base log	0.4676	0.6717	0.6834	1.0000	0.43326	-0.3440
Social Network	0.0253	0.3507	0.1519	0.4333	1.0000	-0.4782
Digital finance	-0.1240	-0.1257	-0.3421	-0.3440	-0.47820	1.0000

Descriptive statistics	Revenue	User base	Revenue log	User base log	Social Network	Digital finance
Max	20 100 000 000	1 600 000 000	23.7240	21.1933	1	1
Min	2 068 819	47 700	14.5425	10.7727	0	0
Mean	4 662 562 994	275 172 546	20.3958	20.3958	0.4728	0.4683
Standard dev.	2 728 384 093	154 556 030	1.8555	2.0695	0.3294	0.3177

Table 4.1*: Correlation matrix and descriptive statistics for dataset.*

4.1 User Base Impact on Revenue

A highly significant association between revenue and user base can be concluded in both models from the coefficient p-values in table 4.3 and 4.4. However, the results from the Akaike information criterion (AIC) test tells us that we cannot conclude that one model is better than the other at explaining the variance in the dataset since the difference between the two scores is less than two as seen in table 4.2. In figure 4.1 the plotted data points can be observed in relation to the two models with the blue dashed line representing the linear model and the red solid line representing the exponential model. As can be seen, the exponential model displays a close to linear association. Given these facts, we cannot conclude that the association between revenue and user base is exponential and cannot reject the null hypothesis in hypothesis 1.

	Linear model - Regression 1	Exponential model - Regression 2	Difference
AIC	297.8136	298.2623	-0.4487

***Table 4.2**: Akaike Information Criterion for both models.*

The bootstrapped regressions states that the coefficients do not differ greatly depending on sample as seen in table 4.3 and table 4.4. The intercept coefficients are slightly larger when bootstrapped in both models. Conversely, the growth rate displayed by the user base coefficients diminishes slightly in both models when bootstrapped. The bootstrapped confidence intervals confirm the original model confidence intervals.

Linear model – Regression 1				
Coefficients	**Regression 1**		**Bootstrap**	**Signif. codes**
Intercept	9.7379		9.8058	***
ln Users	0.6128		0.6090	***
95% Confidence intervals	**Regression 1 Lower**	**Regression 1 Upper**	**Boot. Lower**	**Boot. Upper**
Intercept	7.2349	12.2408	7.3206	12.5681
ln Users	0.4699	0.7557	0.4531	0.7514
Multiple R-squared	0.467			
F-statistic (on 5 and 79 df)	72.73			
p-value	5.739e-13			

***Table 4.3**: Coefficients linear model (regression 1) including bootstrapped results, confidence intervals and R-squared.*
*(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*

Based on the user base coefficient from regression 1 we can note that on average an increase of log normal 1 user will add log normal $ 0.609 of revenue to the company. We get the actual number by taking the Euler's constant to the power of the regression coefficient and taking the proportion of that number towards Euler's constant. Doing this stipulate that an increase of 1 user will add $ 0.6764 of revenue. The bootstrapped confidence intervals tell us that the coefficient mean for the population with 95 percent certainty lies between 0.4531 and 0.7514. We can therefore conclude with 99 percent certainty that the population coefficient will not vary with more than 25 percent.

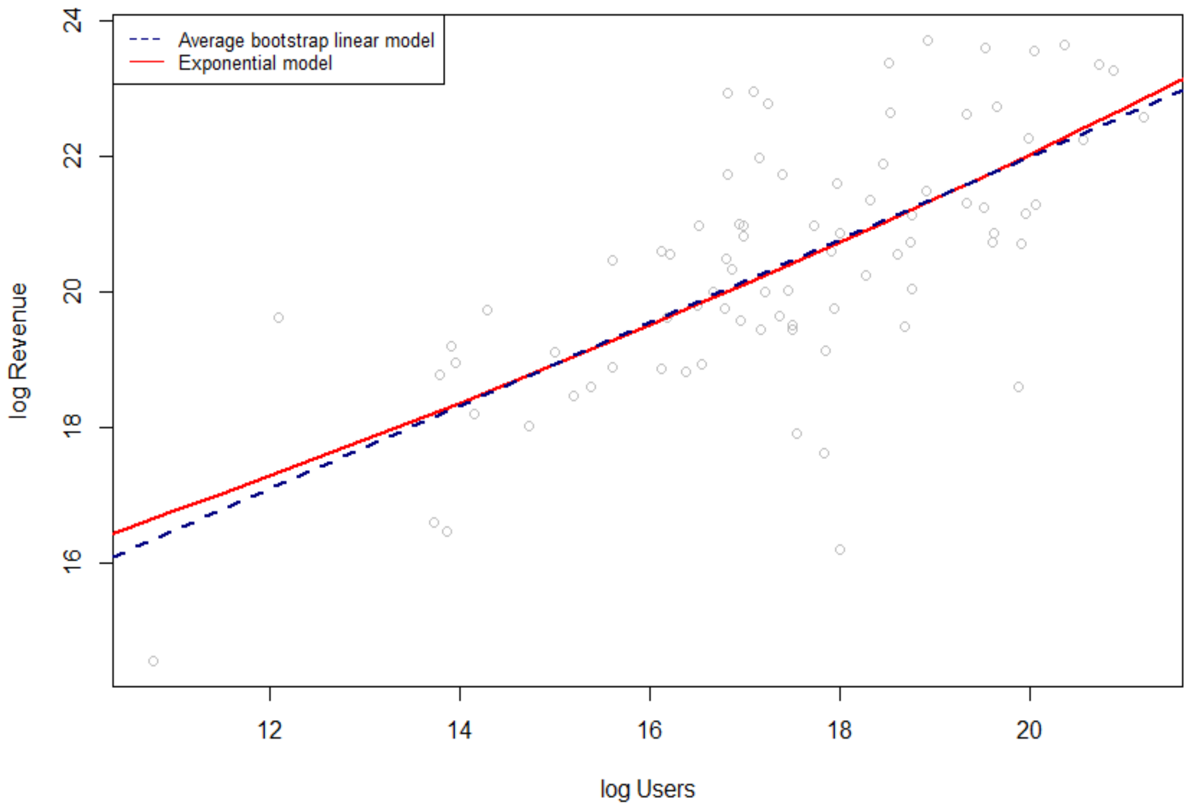


Figure 4.1: *The linear and the exponential model fitted to the log transformed dataset of revenue and user base.*

The coefficient for user base in the exponential model stipulates the growth rate of the exponential association and is set to 0.0303 in the bootstrapped model. If inserted into the equation from regression 2 and user base size is assumed to be log normal 1 we get $e^{0.0303*1} = 1.0308$, or approximately 3.08 percent. In other words, the percentage point increase in revenue growth rate that one natural number increase in log normal user base generates is estimated to 3.08. Like the user base coefficient of the linear model, the bootstrapped confidence interval in the exponential model allows the population coefficient to vary with approximately 20 percent above or below the estimated value.

The intercepts are highly significant in both models and are predicted to coefficient values of 9.8058 in the linear model and 12.0481 in the exponential model. In this case, the intercept can be interpreted as the value of a platform without any users, i.e. the fully developed platform pre-launch to the public. In this case log normal 9.8058 translates to $19 139 log normal 12.0481 to $170 775. However, any conclusions drawn on platforms with user bases smaller than 47 700 lacks validity since that is the smallest observation in the dataset and therefore the interpretation of the intercept lacks credibility. Interpolating the results restricts conclusions to be drawn on platforms with less than 47 700 users and more than 1.6 billion users. Likewise, conclusions are restricted to platforms with revenue between $2 million and $20.1 billion. Distribution of the data points in regard to revenue and user base is visualized in appendix D.

Exponential model – Regression 2				
Coefficients	**Regression 2**		**Bootstrap**	**Signif. codes**
Intercept	11.99943		12.0481	***
ln Users	0.03039		0.0303	***
95% Confidence intervals	**Regression 2 Lower**	**Regression 2 Upper**	**Boot. Lower**	**Boot. Upper**
Intercept	10.5389	13.6382	10.6866	13.6308
ln Users	0.0232	0.0377	0.0231	0.0369
RSE	1366			
Df	83			

Table 4.4*: Coefficients exponential model (regression 2) including bootstrapped results, confidence intervals and residual standard error on degrees of freedom*
*(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*

4.2 Industry Impact

	Res.df	RSS	Df	SST	F stat	p-value	Signif. codes
Regr. 1	83	154.14					
Regr. 4	79	125.96	4	28.179	4.4182	0.002829	**

Table 4.5*: ANOVA F-test for determining significance of industry interaction variables.*
*(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*

	Res.df	RSS	Df	SST	F stat	p-value	Signif. codes
Regr. 3	81	136.16					
Regr. 4	79	125.96	2	10.202	3.1992	0.04613	*

Table 4.6*: ANOVA F-test to determine significance of industry variable.*
*(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*

The ANOVA table 4.5 determines that adding the categorical variables including the interaction variables (regression 4) to the model without categorical or interaction variables (regression 1) makes the model more accurate with 99 percent certainty by presenting a p-value of approximately 0.0028. Additionally, table 4.7 determines that the social network and digital finance dummies included in regression 4 are significant on a 99 and 95 percent level respectively after bootstrapping the results. Thus, the null hypothesis of hypothesis 2.1 can be rejected.

The ANOVA table 4.6 determines that adding the interaction variables (regression 4) to a model including the continuous and dummy variables (regression 3) makes the model more accurate with 95 percent certainty. Thus, the model that concludes to the best in terms of describing the variance in the data is regression 4. Furthermore, table 4.7 determines the interaction variable of social media to be significant on a 90 percent level and the interaction variable for digital finance to be significant on a 95 percent level. As the difference between all three levels of interaction are significant on at least a 90 percent level, the null hypothesis of hypothesis 2.2 can be rejected as well.

	Regression 4	Bootstrap	Signif. codes
Intercept	14.5185	14.4814	***
ln Users	0.3697	0.3721	**
Social network	-7.2853	-7.1802	*
Digital finance	-8.0289	-7.8642	**
Social network interaction	0.3569	0.3503	.
Digital finance interaction	0.4240	0.4142	*

95 % Confidence intervals	Regression 4 Lower	Regression 4 Upper	Boot. Lower	Boot. Upper
Intercept	9.922	19.115	11.2347	18.3054
ln Users	0.1029	0.6365	0.1321	0.5712
Social network	-15.1948	0.6242	-13.5351	-0.8576
Digital finance	-13.8011	-2.2568	-12.3608	-3.1349
Social network interaction	-0.0784	0.7922	-0.0118	0.7124
Digital finance interaction	0.0836	0.7644	0.1194	0.7008
Multiple R-squared	0.5645			
Adjusted R-squared	0.5369			
F-statistic (on 5 and 79 df)	20.48			
p-value	4.707e-13			

***Table* 4.*7*:** *Coefficients regression 4 including bootstrapped results, confidence intervals and R-squared.*

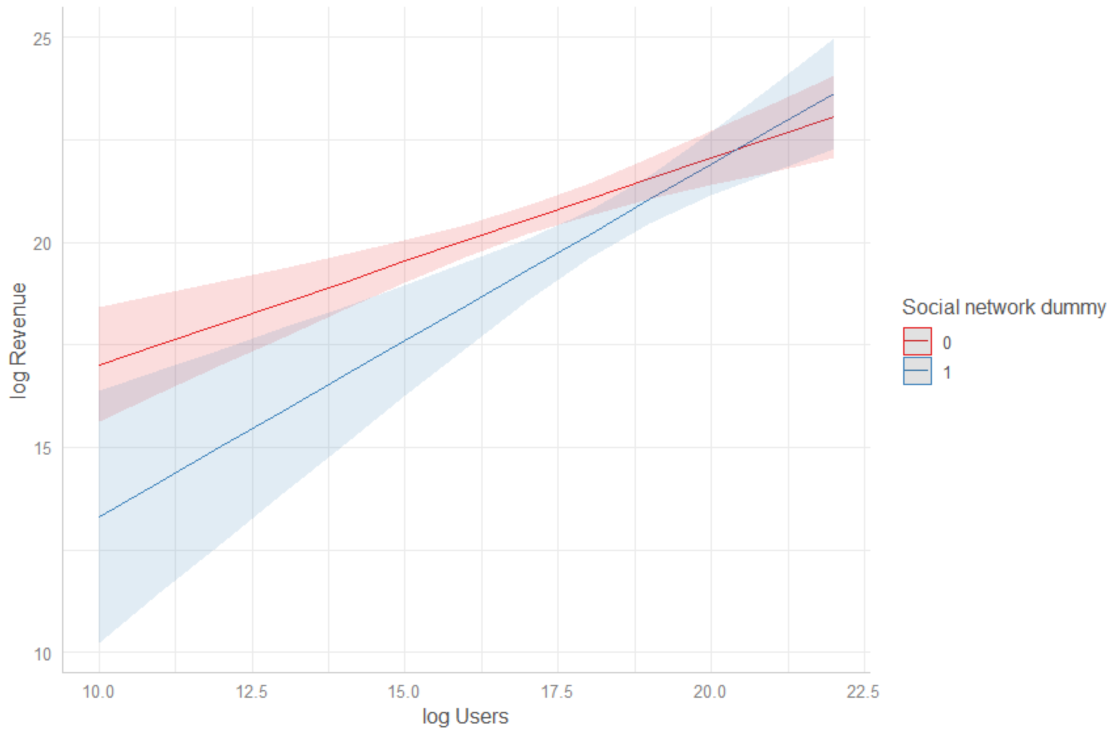


Figure 4.2: *Social network dummy and interaction effect.*
Small e-commerce platforms are more valuable than small social networks even though the marginal added revenue for social networks is higher. Red represents e-commerce and blue social networks, shaded area represents confidence interval.

Bootstrapped results indicate well predicted industry coefficient values as demonstrated in table 4.7 (see appendix E for visualization of bootstrapping regressions on dummy variables and the user base coefficient). The relationship between digital finance and revenue exhibits a stronger relationship than between social networks and revenue. The dummy variables regulating the intercept and the interaction variables regulating the slope for each industry are all significant. The coefficient values of user base and the interactions in table 4.7 indicates that the marginal added revenue of one user is higher for social networks and digital finance than for e-commerce. The marginal revenue increase with log natural 0.7224 for social network companies and with log natural 0.7863 for digital finance companies adding up to an actual marginal revenue added per customer of $0.7576 and $0.8076 respectively.

However, the dummy variables tell us that smaller e-commerce networks exhibit a higher revenue than small social networks and small digital finance platforms. As can be seen in figure 4.2 and figure 4.3 the marginal added revenue per user is higher in social media networks and digital finance however it is only when companies in these two categories become really large that the total value exceeds that of an e-commerce platform with the same size user base.

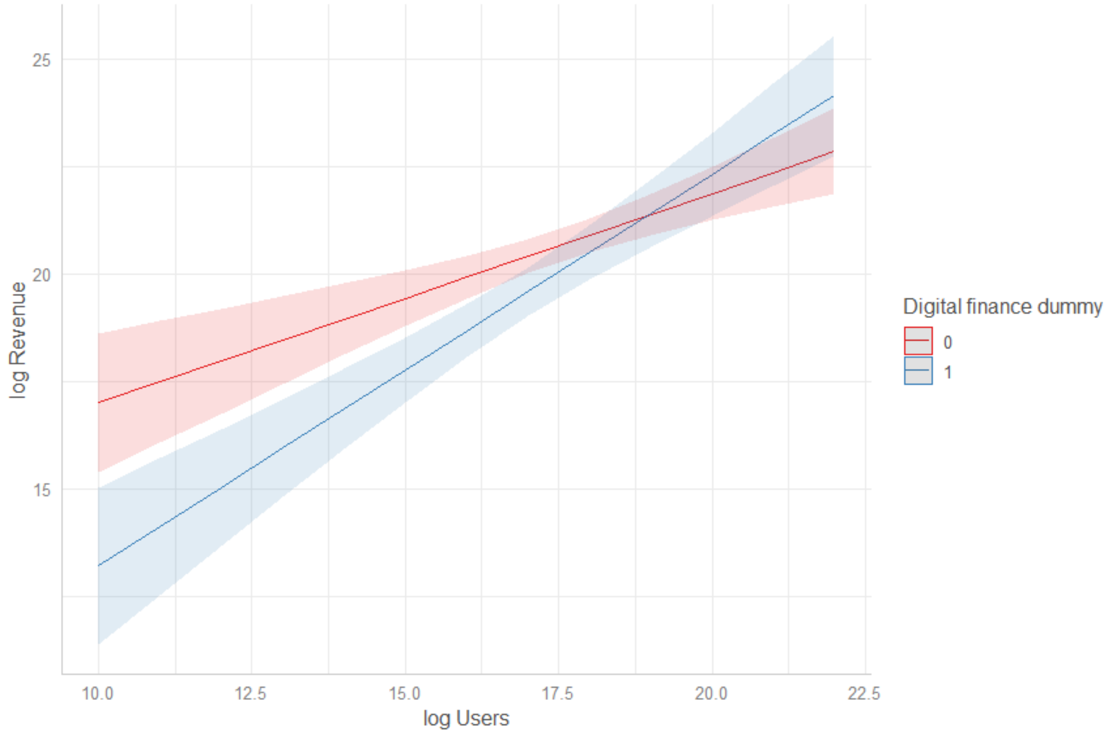


Figure 4.3*: Digital finance dummy and interaction effect.*

Small e-commerce platforms are more valuable than small digital finance platforms even though the marginal added revenue for digital finance is higher. Red represents e-commerce and blue digital finance, shaded are represents confidence interval.

Thus, e-commerce platforms are intrinsically more valuable even though the value per customer is lower. Figure 4.2 shows the revenue increase of social networks and e-commerce platforms. Putting the coefficient values of each industry regression equal to each other gives us their intersect. Thereby, a social network's revenue exceeds that of an e-commerce platform at log natural 20.5 or 797.3 million users. The same threshold can be found for digital finance, which exhibits the strongest marginal revenue added per customer, exceeding the revenue of an e-commerce platform at log natural 18.98 or approximately 174.3 million users as seen in figure 4.3. Digital finance platforms exceed the revenue of social network platforms at log natural 10.625 or approximately 41 150 users. However, the exact intercept of the different platforms is to be interpreted as an approximate value as the data for each variable is thin at large values. The nature of network effects often restricts more than a few large players in each industry competing on the same standard and thus this reflects in the data. Distribution of data points for each industry can be found in appendix D.

Since regression 4 make use of five variables the adjusted R^2 to describe the goodness-of-fit in this model as it introduces a penalty for added variables. As determined earlier, analysis of table 4.5 and 4.6 determined that adding the dummy variables and the interaction variables improved the model's ability to describe the variance significantly. This is reflected in the adjusted R^2 which show a higher figure than R^2 for regression 1. Adding the industry dummy and interaction variables improved the model's ability to predict revenue depending on change in user base with 6.99 percent, derived from taking the difference between the R^2 values of regression 1 and 4 found in table 4.1 and 4.7. Hence, the bulk of the variation explained is attributed to the user base variable.

5 Analysis and Discussion

5.1 Revenue and User Base

Previous literature on network effects, standards, increasing returns and their association with revenue concur in the view that companies operating a platform-mediated business model that profits from network effects should generate more positive externalities as the network grows larger (Gallaugher & Wang, 2002; Katz & Shapiro, 1986; Kolasky, 1999; Shapiro & Varian, 1999). The positive feedback loops and virtuous cycles that help platform owners develop their offering in accordance with the needs of their user base should further enable companies to capture the value of the user base generated by the aforementioned effects (Bresnahan, 1998; Eisenmann, Parker & Van Alstyne, 2006). According to Metcalfe, Reed and Odlyzko, this should in turn generate an exponential association between user base and revenue that shows that as user base increase linearly, revenue should grow exponentially with different growth rates depending on the author (Metcalfe, 2013; Odlyzko & Tilly, 2005; Reed, 1999). As opposed to customers and revenue in a non-platform-mediated business model, whose association is linear (Collier, 2015), the space between the linear growth of the user base and the exponential growth of the revenue would theoretically be explained by the capture and monetization of network effects.

The results of this study regarding the association between user base and revenue are in clear opposition to that of the literature. In the preceding chapter it was shown that the linear model and exponential model explained the data to the same degree with an insignificant difference. The exponential model was fit with a small growth coefficient close to 1 which signifies a close to linear relationship inside the extreme values of the dataset. With the insignificant difference between models, the reasoning which resulted in the formulation of hypothesis 1 cannot be supported.

A number of reasons for why the data did not exhibit the expected relationship can be explored. Firstly, revenue might not be the proper proxy for value in this aspect. Since many platform-mediated companies are still relatively new in comparison to companies with other business models, there might be a chance that the industry has not matured in a way that would imply that all types of monetization strategies are exploited and the value of network effects would thereby be fully captured and reflected in the revenue. The prevailing monetization models in platform-mediated companies are advertisement, transaction or subscription models. These models might not yet enable extracting revenue from the number of additional connections made in a network by connecting a new network node. Rather, they could be less efficient in a manner that only allows them to extract revenue from the one added node. This could explain the seemingly linear relationship in the results. As a consequence of this, future research employing the same research design might find that companies have explored ways of monetizing the added connections instead of the added node. This could cause the observed revenue in the future to be significantly better explained by an exponential association in comparison to a linear one. However, this value could be captured in the market value of a company already as determined by the shareholder value approach (Rappaport, 1986). Therefore, market value might act as a better proxy for the value of network effects.

Secondly, there could also be other explanations to the fast development of digital platform-mediated companies than network effects. Perhaps, the competitive advantage that the literature describes as stemming from network effects is actually located in other factors. These factors could include uniquely low variable costs, unrestricted scalability and platform-mediated companies being unrestrained by geographical impediments. Additional factors could include that the market has not yet been regulated as a consequence of its novelty or that recruiters experience limited talent availability as an effect of the same cause.

Thirdly, the three types of network effects described by Katz and Shapiro (1985) might generate different kinds of cash flow and in turn revenue. Reasonably, direct effects and two-sided mutually reinforcing effects should have a more noticeable effect on revenue in contrast to indirect effects. This is because the externalities are kept within the network as opposed to indirect effects that are generated in collaboration with other players in the same and adjacent markets. Additionally, as described by Katz and Shapiro (1985), companies that share a network or operate on the same standard could enjoy different degrees of revenue as an effect of user base in comparison to other collaborating firms in the network. Asymmetry in network effect value capture between participating firms could generate skewed data points. Companies that operate more than one network might also experience another user base to revenue association. As an example, there might be a difference in effect on revenue between a company operating two networks with 500 thousand users in both and a company operating one network with 1 million users. According to the literature, the revenue increase should differ in the two cases and is something that the data in this paper fail to consider.

5.2 Not all Industries are Created Equal

Valuation literature states that industry analysis is important when valuing a company as ROIC and future prospects etc. can vary between industries (Koller, Goedhart & Wessels, 2015). Previous literature that examines digital platform industries, business models and revenue models point to the fact that digital platform-mediated companies differ depending on what industry they operate in. Since the digital industries examined in this paper have a physical counterpart, they have all taken a concept of a physical business model and digitized it. In digital finance, the focus lies on reducing transaction costs and information asymmetry that physical banks and brokers have a hard time accomplishing because of antiquated systems and operations (Assocham, 2016). Additionally, focus lies on innovative solutions to attract a user base which generate a large flow of information available to other platform users. In this way, the network effects generated are of the direct kind. Additionally, transactions being brokered on the platform create two-sided mutually reinforcing network effects. In the social network industry the value creation process lies in enabling long tail connections and enabling connections with other users which an individual might not have made in the real world. Revenue stem mostly from advertising but some networks also employ the transaction or subscription method to capture value. In the e-commerce industry, the value offer is based in selling niche products from the long tail that physical retail stores cannot carry in their assortment because of storage impediments. In this case, the e-commerce platform steps in as a market intermediary by connecting buyers and sellers which means that e-commerce platforms can circumvent the issue of storage by carrying less inventories (Patil & Diverkar, 2014).

Our study confirms the literature in that industries differ and show that revenue’s dependency on user base vary between industries. The slope coefficients and Y intercept are different depending on each industry as shown in the results. This means that some industries are more reliant on its user base to obtain revenue growth in the early stages of their business life cycle than others. E-commerce is the industry that has the highest Y intercept at 14.48 which could mean that e-commerce platforms are inherently more valuable than social networks and digital finance platforms. The social network industry has the second largest Y intercept at 7.3 while digital finance follows closely at 6.62. However, what is interesting to note is that digital finance, having the lowest intercept, has the highest slope coefficient at 0.7864. The slope coefficient of social networks also shows a steep slope at 0.7224. This means that e-commerce achieves comparably high revenue at a small size user base while digital finance and social networks achieve a comparably high revenue when the user base becomes large. E-commerce has the highest revenue until the platforms reach a total of 174.3 million MAU, when it is surpassed by digital finance. Social networks surpass e-commerce revenue at 797.3 MAU however do not reach above digital finance. Therefore, we can conclude that below 174.3 million MAU e-commerce is the industry exhibiting the highest revenue and over 174.3 million MAU digital finance generate the largest revenue out of the three industries.

The fact that the results showcase an industry difference can depend on a number of factors. Firstly, the industries operate where different types of network effects are prevalent. The nature of the offering that digital finance firms present makes the network effects generated to fall in the direct and two-sided category. The information flows between participants on the platform generate the direct effects and the connection between investors, clearinghouses and markets generate the two-sided effects. Social networks profit from direct that are generated by connections on the platform. E-commerce generates two-sided effects by connecting buyers and merchants. Secondly, different types of revenue models are predominant in the different industries. Digital finance and e-commerce mainly employ the transaction model while social network platforms primarily use the advertising model. These factors combined could be the reason as to why there is an observed industry variation in the dataset. Given the above two factors, the argument that direct effects generate a higher marginal revenue added per customer than two-sided effects could be made. As observed in both social networks and digital finance, the slope coefficients are significantly higher than the one of e-commerce. Additionally, an observation of low revenue is made for social networks and digital finance that generates direct effects when they exhibit a small user base. Thus, the conclusion can be drawn that direct effects are hard to monetise in the beginning but as the user base grows the revenue appreciates quickly.

6 Conclusion

6.1 Research Aims, Objectives and Findings

The aim of the paper was threefold. Firstly, we wanted to test the association between user base and revenue for companies who operate a platform-mediated business model. Moreover, as these types of platforms operate in different settings that incorporate a variety of different business models and generate revenue from different types of network effects, the value of each user should theoretically vary depending on the industry. Therefore, the second aim became to test the association for an industry variable. More specifically, the idea was to test which one of the main industries, i.e. e-commerce, social networks and digital finance, that is most reliant on its user base for revenue generation. The two aims together would supply managers and analysts with helpful information about network effects impact on business as well as industry specific insights.

The study was also conducted in order to add the perspective of network effects to traditional valuation theory. This is because the shareholder value approach values a company based on its ability to generate cash flow. However, in order to implement the shareholder value approach, future cash flow needs to be estimated by analysts drawing assumptions about the future of the company. This means that the estimated value of a company can vary between analysts. Therefore, a researched pattern between revenue and user base could be applied to more accurately predict future cash flows and the spread of analyst's estimations effectively become reduced. An industry specific variance would further increase the validity of the estimations.

The findings in this study cannot observe a significant difference between an exponential or a linear association between revenue and user base as determined when testing hypothesis 1. The test cannot provide any evidence of an exponential association. However, as shown in chapter 4.2 one can draw the conclusion that user base significantly affects revenue as R^2 amounts to 46.7 percent when simple linear regression is applied to user base and revenue. This means that the variance in the change of revenue can be predicted to 46.7 percent.

Moreover, as not all users are created equal, chapter 4.2 shows us that different industries exhibit varying marginal added revenue per customer as well as different opportunities to generate revenue with a small network size. While e-commerce platforms generate more revenue already early on, digital finance platforms and social networks demand a larger user base to generate the same amount of revenue as an e-commerce platform. The aim of supplying managers and analysts with useful information to gain insights about their business was accomplished. The findings of this study show that the variation of revenue can be to 46.7 percent predicted by change in user base. Additionally, different industries display different behaviour with altering early-stage revenue generation and changing marginal revenue added per users.

6.2 Theoretical and Practical Implications

The theoretical contributions of the study are mainly found in the revenue to user base growth laws put forth by Metcalfe, Odlyzko and Reed as well as showing that the industry differences outlined by Koller, Goedhart and Wessels (2015), Velu (2015), Amit and Zott (2000) and Enders et al. (2008) reflects in revenue created by the user base. Our findings weaken the notion of an exponential association between user base and revenue as put forth by Metcalfe (2913), Reed (1999) and Odlyzko (2005). They have developed their ideas on the exponentially increasing number of possible connections that a network creates as it grows (see Figure 2.3). Reed and Metcalfe's law formed an exponential equation while Odlyzko's is comparatively linear. The results in this study cannot determine which of the laws most accurately describe the association between revenue and user base as they have not been tested directly. However, as an exponential association cannot be confirmed to describe the data better than a linear, it weakens the laws of Metcalfe and Reed while slightly strengthens Odlyzko's as he hypothesized an association comparatively linear to the others. Network effects as a theory is unaffected by the results of this study as network effects does not explicitly draw the connection between user base and exponentially increasing revenue and as the demonstrated result can depend on a number of other factors as well. Regarding the industry differences, this study contributes to theory through showing that industry specific factors have an effect on the amount of revenue that can be expected from an investment in different industries. It also contributes through stating at what rate revenue is expected to increase depending on industry and size of the user base.

In terms of practical implications, this study provides analysts and investors with certain applicable points of high relevance. The results and discussion of this paper can work as guidelines for analysts at financial institutions when valuing companies that operate a platform-mediated business model in the e-commerce, social networks or digital finance industry. Investors can use the results of this paper to evaluate a possible or past investment. Business managers and entrepreneurs can likewise use the results and discussion of this paper to set expectations with investors and to back up strategies focusing on user base growth. The result of this paper points to the fact that user generation are imperative to platform success. Furthermore, the discussion opens up the possibility of business processes not being developed to capture all of the positive externalities generated by network effects. Future innovative business models, revenue models or user applications might enable companies to capture the great value of network effects, increasing returns and standards as determined by the literature. Therefore, in these types of companies, emphasis should be put on trying to capture the aforementioned values. A player successful in such a venture could generate an unprecedented market position.

6.3 Limitations and Future Research

The study formulated in this paper is limited in certain aspects and could therefore be improved in future research. The data imposes limitations on the applicability of the results of the study. As mentioned in chapter 4, the results from this analysis should not be extrapolated on companies outside the dataset's limits in terms of revenue, user base size and industries. The revenue variable ranges from 2 million and 20.1 billion USD and the user base size range from 47 700 and 1.6 billion MAU (see appendix D). This means that the results from this study should not be applied to companies outside of these ranges. More importantly, the study should not be applied to companies in the early stages of development. Companies in an early stage of development seldomly publishes the variables used in this study to the public. Thus, the study has been limited to companies in more mature states. In addition, because of the dynamics of platform industries outlined in chapter 2, few companies survive the tough competition long enough to become sustainable business entities that publish reliable data. Moreover, as the study only test international companies the results should not be applied to solely domestic platforms. Therefore, future research should try to mitigate the problem of applicability that lack of data has caused this study.

This study uses horizontal data to test the association between revenue and user base. Future research could find it interesting to look at this research question with panel data that would describe the revenue to user base development over time for each company. Due to data unavailability this method could not be used in this study. By implementing panel data each firm could be tested over a period of time and therefore significantly strengthen any association between revenue and user base in a firm specific manner. This is because the association between revenue and user base can be drawn not just different companies but also in terms of the companies themselves. A panel data study would also open up the possibility of analysing the churn and flows of users. Furthermore, it establishes the average time it takes for a company to generate a critical mass and in turn the impact of successfully implementing a standard as described by Shapiro and Varian (1999). However, as the data at the moment is limited a horizontal study became the best alternative.

Another limitation with the study is the use of variables. This study uses user base and industry category to estimate the revenue of platform-mediated companies. However, as the literature and results hint about, there might be significant differences between revenue models and what type of network effects that are generated by the platform. This is something future research could look further into. The same applies to the independent variable. Using market capitalisation as proxy for value instead of revenue could yield a more interesting result. Market capitalisation is more closely associated with value as it incorporates more aspects of the financial statements. When more future data becomes available these limits could be mitigated. This is because more companies will have gone public leading to more reliable data spanning across a time period making that makes panel data analysis possible. Therefore, future research could find it interesting to add firm specific network effects and revenue models to the set of predictors and switch revenue for market capitalisation.

7 References

Abramson, J. (2021). Fitting Exponential Models to Data, Mathematics LibreTexts, Available online: https://math.libretexts.org/Bookshelves/Precalculus/Book%3A_Precalculus_(OpenStax)/04%3A_Exponential_and_Logarithmic_Functions/4.08%3A_Fitting_Exponential_Models_to_Data [Accessed 26 May 2021]

Amit, R., & Zott, C. (2000). Value Drivers of E-commerce Business Models, INSEAD, Available online: https://flora.insead.edu/fichiersti_wp/inseadwp2000/2000-54.pdf [Accessed 23 May 2021]

Arthur, W. B. (1996). Increasing Returns in the World of Business, *Harvard Business Review*, July-August 1996, pp. 100 – 109. Available online: Increasing Returns and The New World of Business.pdf (fing.edu.uy) [Accessed 23 April 2021]

Assocham. (2016). Indian banks can reduce costs by 50% on per-transaction basis by going digital: ASSOCHAM-PwC study, Available online: https://www.assocham.org/newsdetail.php?id=5600 [Accessed 26 May 2021]

Bayón, T., Gutsche, J., & Bauer, H. (2002). Customer Equity Marketing: Touching the Intangible, *European Management Journal,* Vol. 20, No. 3, pp. 213-222, Available online: https://www.sciencedirect.com/science/article/pii/S0263237302000373 [Accessed 10 May 2021]

Berenson, M., Levine, D., & Szabat, Ka. (2014). Basic Business Statistics Concepts and Applications, Thirteenth edition, Pearson Education Limited, Edinburgh England

Bonardi, J., & Durand, R. (2003). Managing Network Effects in High Tech Industries, *Academy of Management Executive* Vol. 17, No. 4, pp. 40–52, Available online: https://journals.aom.org/doi/abs/10.5465/ame.2003.11851827 [Accessed 11 May 2021]

Boston University. (2016). Simple Linear Regression, Correlation and Regression with R, Available online: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R5_Correlation-Regression/R5_Correlation-Regression4.html [Accessed 11 May 2021]

Bougie, R., & Sekaran, U. (2020). Research Methods for Business a Skill Building Approach, Great Britain, Glasgow, John Wiley & Sons

Bresnahan, F. (1998). New Modes of Competition: Implications for the Future Structure of the Computer Industry, Stanford Institute for Economic Policy Research, SIERP Discussion Paper No. 500 (June 1998) Stanford University, Available online: https://siepr.stanford.edu/sites/default/files/publications/500_0.pdf [Accessed 6 May 2021]

Brue, S. (1993). Retrospectives, the Law of Diminishing Returns. *Journal of Economic Perspectives*, Vol. 7, No. 3 (Summer 1993) pp. 185-192, Available online: https://www.aeaweb.org/articles?id=10.1257/jep.7.3.185 [Accessed 12 May 2021]

Bryman, A., & Bell, E. (2011). Business Research Methods, 3rd edition, Oxford University Press

Brynjolfsson, E., Hu, Y., & Smith, M. (2006). From Niches to Riches: Anatomy of the Long Tail, *MIT Solan Management Review,* Vol. 47, No. 4 (Summer 2006) pp.67 – 71, Available online: https://sloanreview.mit.edu/article/from-niches-to-riches-anatomy-of-the-long-tail/ [Accessed 23 May 2021]

Busse, S. (2012). Marketing Nerd: Why Your Business Should Care About Networks, When 4 plus 1 equals 10, Kinesis, Available online: https://www.kinesisinc.com/using-networks-to-grow-your-business/ [Accessed 12 May 2021]

Cambridge Dictionary. (2021a). E-Commerce, Available online: https://dictionary.cambridge.org/dictionary/english/e-commerce [Accessed 5 May 2021]

Cambridge Dictionary. (2021b). Social Network, Available online: https://dictionary.cambridge.org/dictionary/english/social-network [Accessed 5 May 2021]

Collier, P. (2015). Accounting for Managers: Interpreting accounting information for decision making, Fifth Edition, John Wiley & Sons Ltd, West Sussex, United Kingdom

David, A. (1985). Clio and the Economics of QWERTY*, The American Economic Review*, Vol. 75, No. 2, (May 1985) pp. 332 – 337, Available online: https://www.jstor.org/stable/1805621?seq=1#metadata_info_tab_contents [Accessed 5 May 2021]

De Boer, S. (2021) Intangible ironies: investor mispricing of company assets on and off its balance sheet, SSRN, Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3789714 [Accessed 26 May 2021]

Delevingne, L. (2020). U.S. Big Tech Dominates Stock Market after Monster Rally, Leaving investors on edge, Reuters, Available online: https://www.reuters.com/article/us-usa-markets-faangs-analysis-idUSKBN25O0FV [Accessed 15 May 2021]

Dhar, V., & Stein, R. (2017). Economic and Business Dimensions, Fintech platforms and strategy, *Communications of the ACM,* Vol. 60, No. 10, pp. 32 – 35, Available online: https://dl.acm.org/doi/pdf/10.1145/3132726 [Accessed 23 May 2021]

Eisenmann, T. (2007). Managing Network Businesses: course overview for educators, Harvard Business School note no. 806103

Eisenmann, T., Parker, G., & Van Alstyne, W. (2006). Strategies for Two-Sided Markets, *Harvard Business Review*, 84, 10, p.92, Available online: https://edisciplinas.usp.br/pluginfile.php/1704705/mod_resource/content/1/Eisenmann%2

0-%20Estrat%E2%80%9Agias%20para%20mercados%20multilaterais.pdf [Accessed 6 May 2021]

Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform Envelopment, *Strategic Management Journal*. Vol. 32, No. 12 (December 2011) pp. 1270 – 1285, Wiley. Available online: https://www.jstor.org/stable/41261793 [Accessed 6 May 2021]

Enders, A., Hungenberg, H., Denkel, H., & Mauch, S. (2008). The Long Tail of Social Networking: Revenue models of social networking sites, *European Management Journal*, Vol. 26 No. 3, pp. 199 – 211, Available online: https://www.sciencedirect.com/science/article/pii/S0263237308000200#! [Accessed 23 May 2021]

European Commission. (n.d.). What is digital finance?, Available online: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/digital-finance_en [Accessed 5 May 2021]

Gallaugher, M., & Wang, Y. (2002). Understanding Network Effects in Software Markets: Evidence from web server pricing, *Management Information Systems Quarterly*, December 2002, Vol. 26, No. 4, pp. 303-327, Available online: https://www.jstor.org/stable/4132311?seq=1#metadata_info_tab_contents [Accessed 5 May 2021]

Google. (n.d.). Frequently asked questions about the Google+ shutdown, Available online: https://support.google.com/googlecurrents/answer/9217723?hl=en [Accessed 5 May 2021]

Greco, M., Crielli, L., & Grimaldi, M. (2013). A Strategic management framework of tangible and intangible assets, *European Management Journal,* Vol. 31, No. 1, February 2013, pp. 55-66, Available online: A strategic management framework of tangible and intangible assets - Science Direct [Accessed 20 May 2021]

Granovetter, M. (1973). The Strength of Weak Ties, American Journal of Sociology, Vol. 78, No. 6, pp. 1360-1380, Available online: https://www.journals.uchicago.edu/doi/abs/10.1086 /225469 [Accessed 23 May 2021]

Hu, S. (2012). Akaike Information Criterion, Centre for Research in Scientific Computation, North Carolina University, Available online: https://www.researchgate.net/profile/Shuhua-Hu/publication/267201163_Akaike_Information_Criterion/links/599f662aa6fdccf5941f894b/Akaike-Information-Criterion.pdf [Accessed 25 May 2021]

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to Statistical Learning with Applications in R, Springer, Springer Texts in Statistics, Springer Science + Business media, Ney York, Corrected at 8th Printing 2017

Katz, M., & Shapiro, C. (1985). Network Externalities, Competition, and Compatibility, *The American Economic Review*, Vol. 75, No. 3 (Jun., 1985), pp. 424- 440, Available online:

https://www.jstor.org/stable/pdf/1814809.pdf?refreqid=excelsior%3A20bd25e98f1b870766877c93eb83e86f [Accessed 19 April 2021]

Kolasky, J. (1999). Network effects: A contrarian view, *George Mason Law Review*, Vol. 7, No. 3, p. 577 – 616, Available online: https://heinonline.org/HOL/Page?handle=hein.journals/gmlr7&div=27&g_sent=1&casa_token=&collection=journals [Accessed 19 May 2021]

Koller, T., Goedhart, M., & Wessels, D. (2015). Valuation Measuring and Managing the Value of Companies, Mckinsey & Company, Sixth edition, university edition

Metcalfe, B. (2013). Metcalfe's Law after 40 Years of Ethernet, Published by the IEEE Computer Society, *Journal of Computer*, Vol. 46, No. 12, pp. 26-31 Available online: https://ieeexplore.ieee.org/abstract/document/6636305 [Accessed 14 April 2021]

Francis, G., & Minchington, C. (2002). Regulating Shareholder Value: A Case Study of the Introduction of Value-based Measures in a Water Company, *British Journal of Management,* Vol. 13, No. 3, pp. 233-47, Available online: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8551.00240 [Accessed 10 May 2021]

Odlyzko, A., & Tilly, B (2005). A refutation of Metcalfe's Law and a better estimate for the value of networks and network interconnection, University of Minnesota, Available online: http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf [Accessed 14 April 2021]

Ojala, A. (2013). Software-as-a-Service Revenue Models, *IEEE IT Professional*, Vol. 15, No. 3, pp. 54-59, (May-June 2013), Available online: https://ieeexplore.ieee.org/abstract/document/6243129 [Accessed 23 May 2021]

Patil, H., Diverkar, R. (2014). Inventory Management Challenges for B2C E-commerce Retailers, Procedia Economics and Finance, Available online: https://www.researchgate.net/publication/275541054_Inventory_Management_Challenges_for_B2C_E-commerce_Retailers [Accessed 25 May 2021]

Peters, L. (2021). Top online stores in the United States in 2019, by e-commerce net sales, Statista, Available online: https://www.statista.com/forecasts/646030/united-states-top-online-stores-united-states-ecommercedb [Accessed 5 May 2021]

Princeton University Library. (2007). Working with Dummy Variables, The trustees of Princeton University, Available online: https://dss.princeton.edu/online_help/analysis/dummy_variables.htm [Accessed 7 May 2021]

Rappaport, A. (1986). Creating Shareholder Value: The New Standard for Business Performance, New York: The Free Press, a division of Macmillan Inc.

Reed, P. (1999). Weapon of math destruction: A simple formula explains why the Internet is wreaking havoc on business models, *Context Magazine* (spring 1999).

Shapiro, C., & Varian, R. (1999). The Art of Standards War, *California Management Review*, Vol. 41, No. 2, (Winter 1999). Available online: https://journals.sagepub.com/doi/pdf/10.2307/41165984 [Accessed 6 May 2021]

Stephen, T., & Toubia O. (2010). Deriving Value from Social Media Commerce Networks, *Journal of Marketing Research*, Vol. XLVII (April 2010) pp. 215 – 228, ISSN: 1547-7193, Available online https://journals.sagepub.com/doi/pdf/10.1509/jmkr.47.2.215 [Accessed 8 May 2021]

Schweiger, A., Nagel, J., Böhm, M., & Krcmar, H. (2016). Platform Business Models, TUM Living Lab Connected Mobility State of The Art Report, Available online: https://mediatum.ub.tum.de/doc/1324021/1324021.pdf#page=98 [Accessed 23 May 2021]

Tankovska. H. (2021). Projected revenue of Instagram from 2017 to 2019, Statista, Available online: https://www.statista.com/statistics/271633/annual-revenue-of-instagram/ [Accessed 5 May 2021]

Thomala, L. (2021). Number of monthly active users (MAU) of the leading messaging apps in China as of December 2020, Statista, Available online: https://www.statista.com/statistics/1062449/china-leading-messaging-apps-monthly-active-users/ [Accessed 5 May 2021]

Velu, C. (2015). Business Model Innovation and Third-Party Alliance on the Survival of New Firms, *Technovation*, Vol. 35 (January 2015) pp. 1-11, Available online: https://www.sciencedirect.com/science/article/pii/S0166497214001291 [Accessed 23 May 2021]

Wilson, J. (2010). Essentials of Business Research: A Guide to Doing Your Research Project, SAGE Publications

Zhang, X., Liu, J., & Xu, Z. (2015). Tencent and Facebook Data Validate Metcalfe's Law, *Journal of Computer Science and Technology*, 30(2): 246–251, Available online: https://link.springer.com/article/10.1007/s11390-015-1518-1 [Accessed 14 April 2021]

Appendix A: Equations

Equation A.1: Metcalfe's Law

$$V = \alpha \times\ n(n-1) \quad or\ approximately \quad V\ \propto n^2$$

Where:

V = value, often using revenue as a proxy

α = an unknown constant describing the proportionality

n = number of users at any given time

Equation A.2: Odlyzko's Law

$$V = \alpha \times n\ (\log n) \quad or \quad V\ \propto n(\log n)$$

Where:

V = value, often using revenue as a proxy

α = an unknown constant describing the proportionality

n = number of users at any given time

Equation A.3 Reed's Law

$$V = \alpha \times 2^n \quad or \quad V\ \propto 2^n$$

Where:

V = value, often using revenue as a proxy

α = an unknown constant describing the proportionality

n = number of users at any given time

Equation A.4: Multiple Linear Regression, (James et al. 2013, Figure 3.19, page 71)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Where:

Y = Independent variable

X_p= Dependent variable

β_0 = Intercept

β_p = Slope coefficient for variable X_p

ϵ = Residual error

Equation A.5: Exponential Regression (Abramson, 2021)

$$Y = \alpha e^{\beta_p X_p} + \epsilon$$

Where:

Y = Independent variable

X_p= Dependent variable

α = Intercept

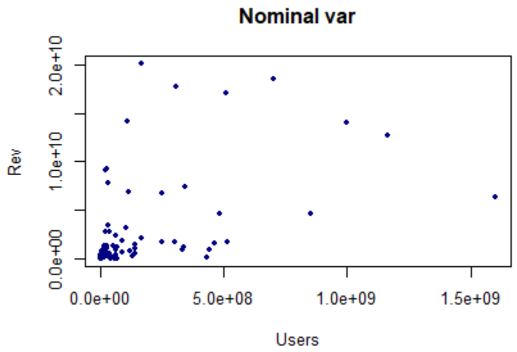
β_p = Growth or decay rate of variable X_p

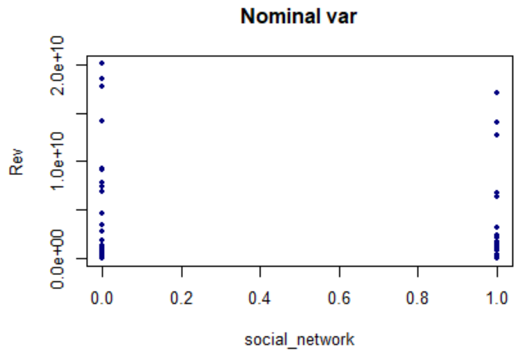
ϵ = Residual error

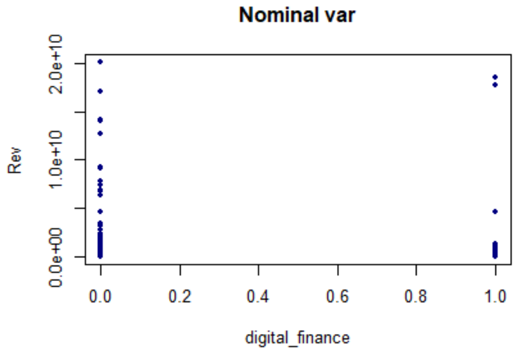
Appendix B: Normality Assumptions

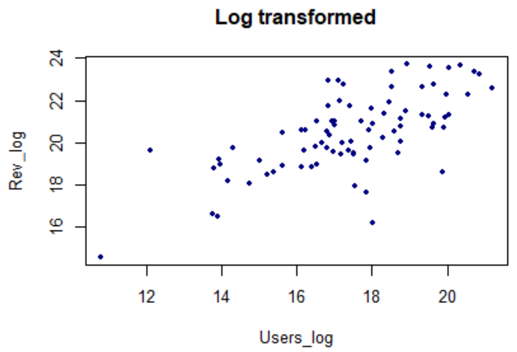
The four tests of normal distribution. Plots before and after variables Users and Revenue were log normal transformed as suggested by the Cullen and Frey plot.

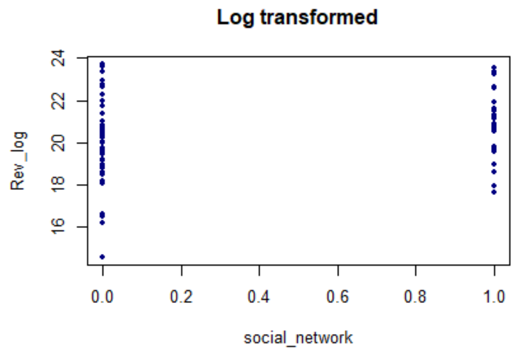
Linearity

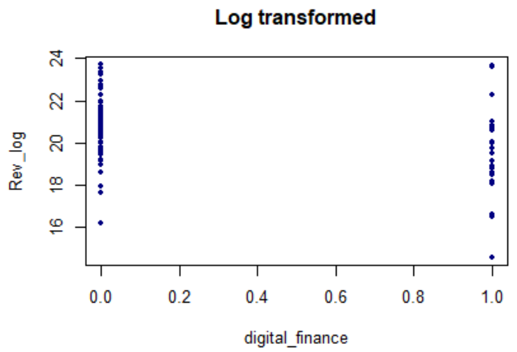












Homoscedasticity

Equal variance (homoscedasticity) assumption passed after transforming the variables. The variance remain approximately consistent over the range of values. LM1 and LM1_log signifies model with revenue and user base variables. LM2 and LM2_log signifies model LM1 and LM1_log with added dummy and interaction variables.

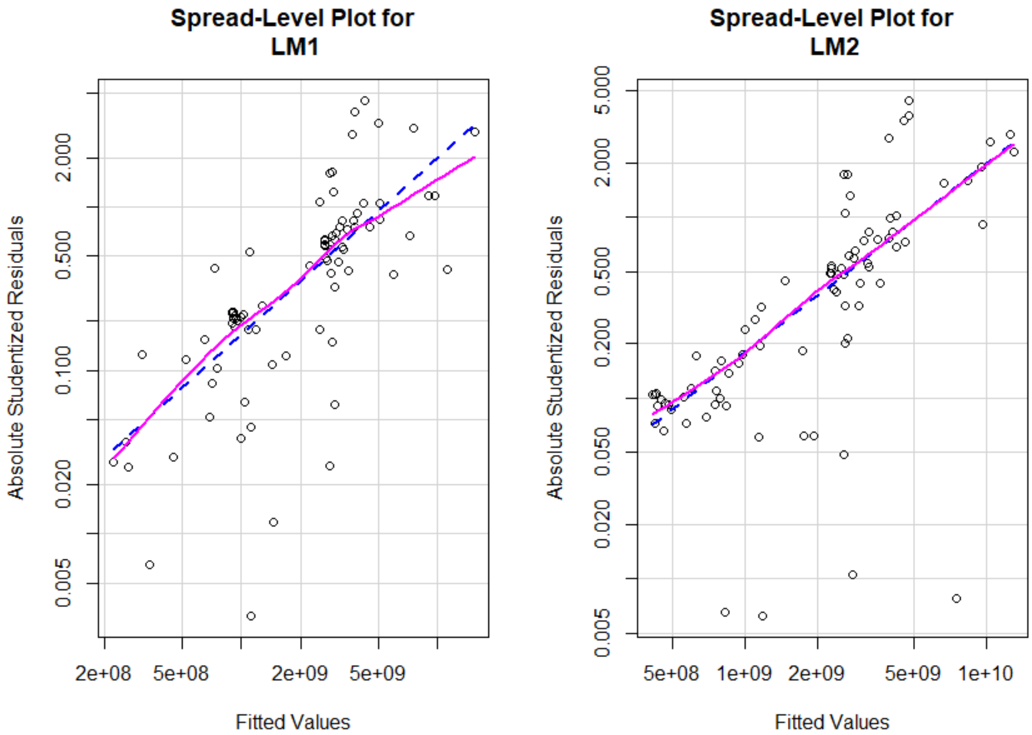


Figure 8: Homoscedasticity plots of nominal values.

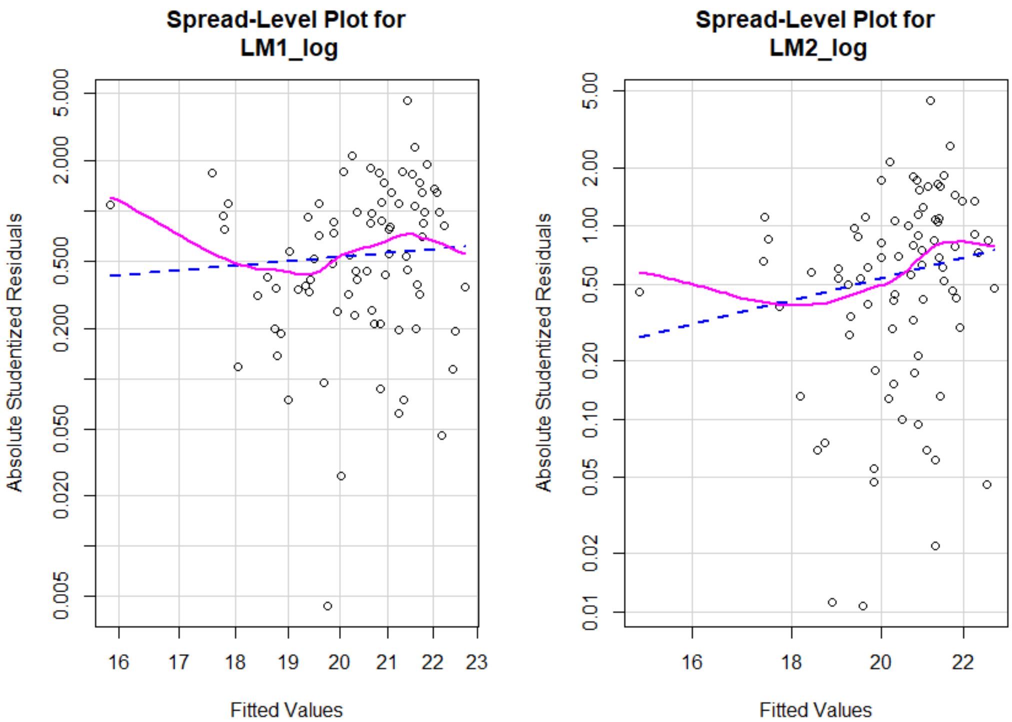
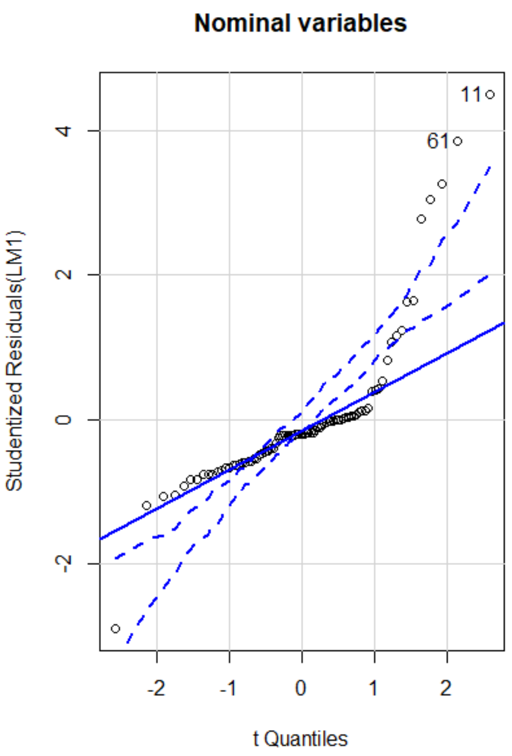
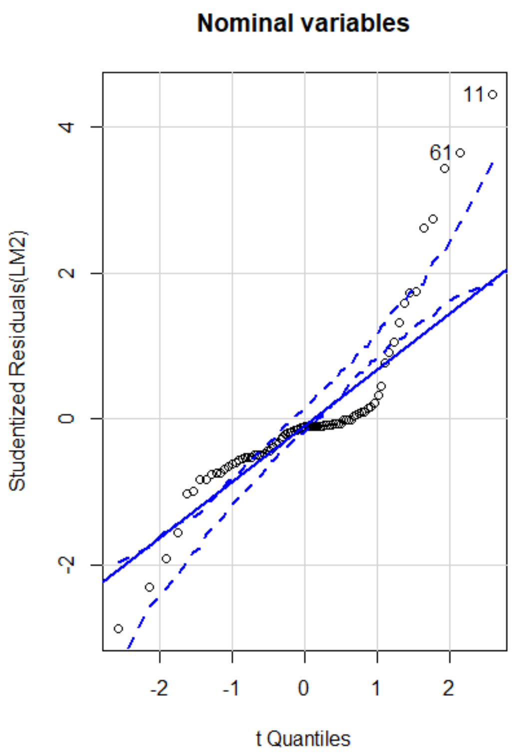


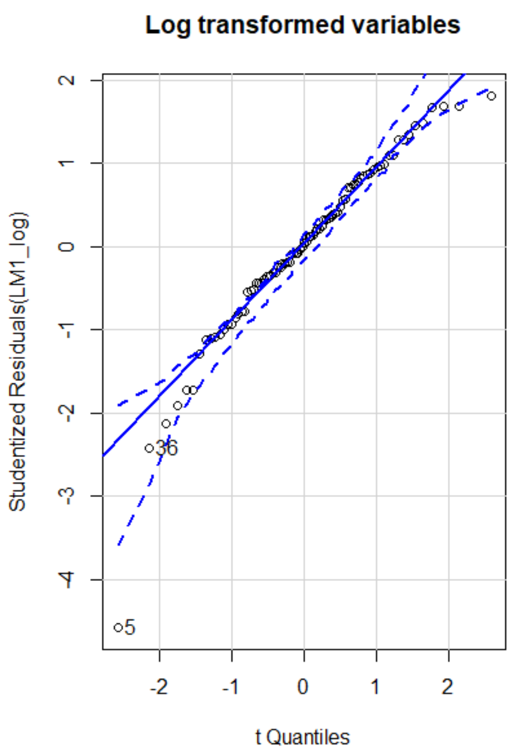
Figure 9: Homoscedasticity plots of log transformed values.

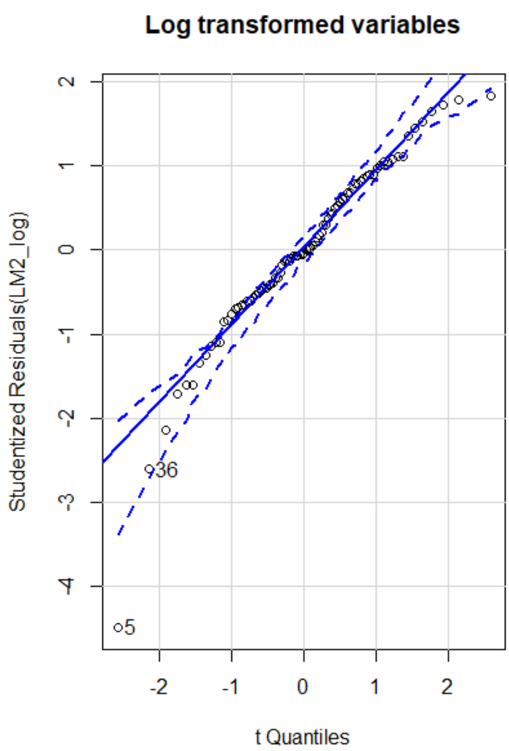
Normality of Residuals

Checked with QQ-plots. After variables were transformed the QQ-plot showed much better fit. However, since the relationship still is not completely linear, the regression s going to be bootstrapped in order to ensure reliable values in the confidence intervals.









Multicollinearity

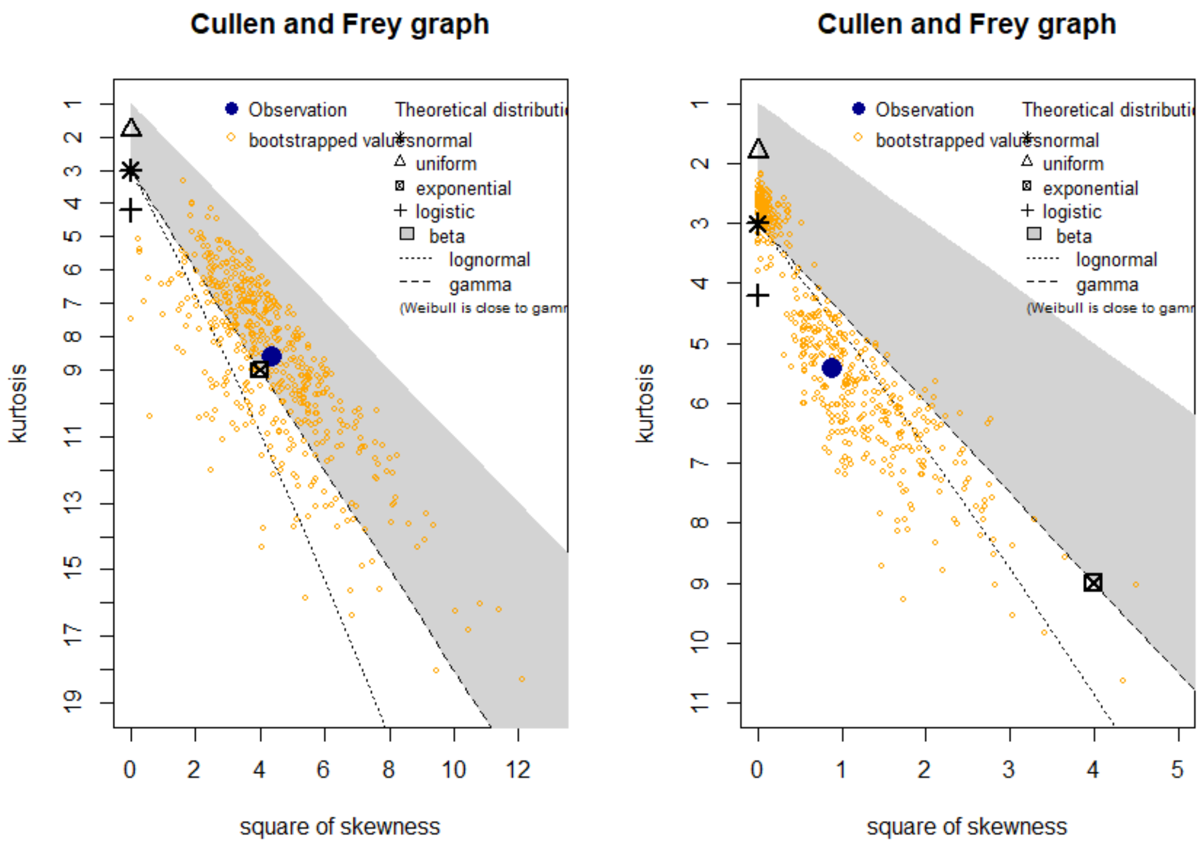
Checked through the variance inflation factor. After transforming the variables, the VIF increase slightly. However, as all variables have a VIF close to one we conclude that there is not a problem of multicollinearity (James et al. 2013). Interaction variables excluded since they inherently cause multicollinearity.

Multicollinearity analysed via variance inflation factor.

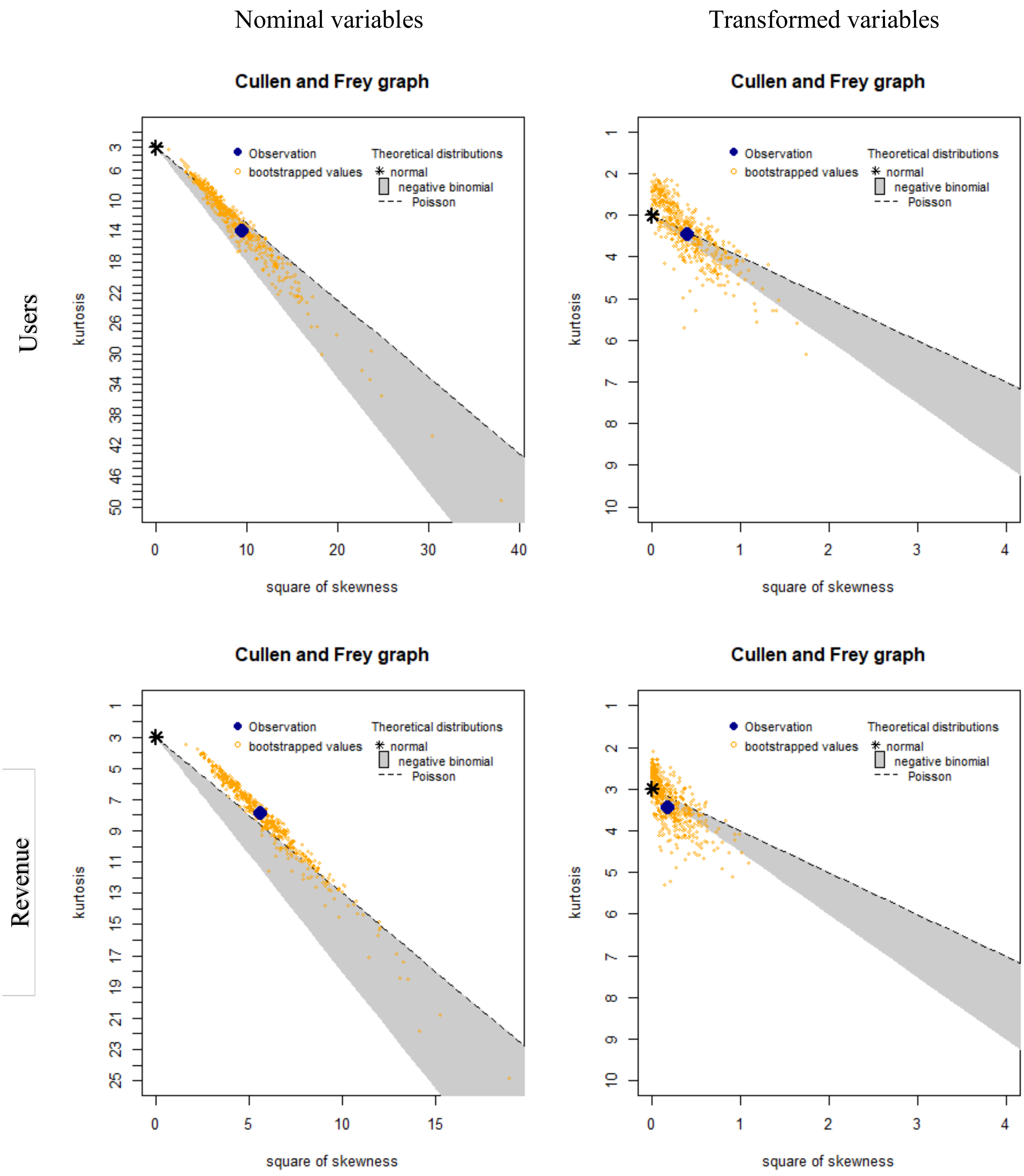
	VIF
Nominal variables	
Users	*1.143269*
Social network	*1.458807*
Digital finance	*1.299870*
Transformed variables	
Users	*1.269015*
Social network	*1.450542*
Digital finance	*1.336400*

Appendix C: Transformation of Variables

Nominal variables Transformed variables



Cullen and Frey graph of residuals before and after log normal transformation of variables Revenue and Users. Plotting the residuals from the regression performed with the nominal values suggests that a log normal distribution might show better results than a normal distribution. After natural log normal transformation of the variables, the result shows a better fit to the normality assumption.



After transforming the variables, the Cullen and Frey graphs show evidence of normal distribution in the data.

Appendix D: Distribution of Data

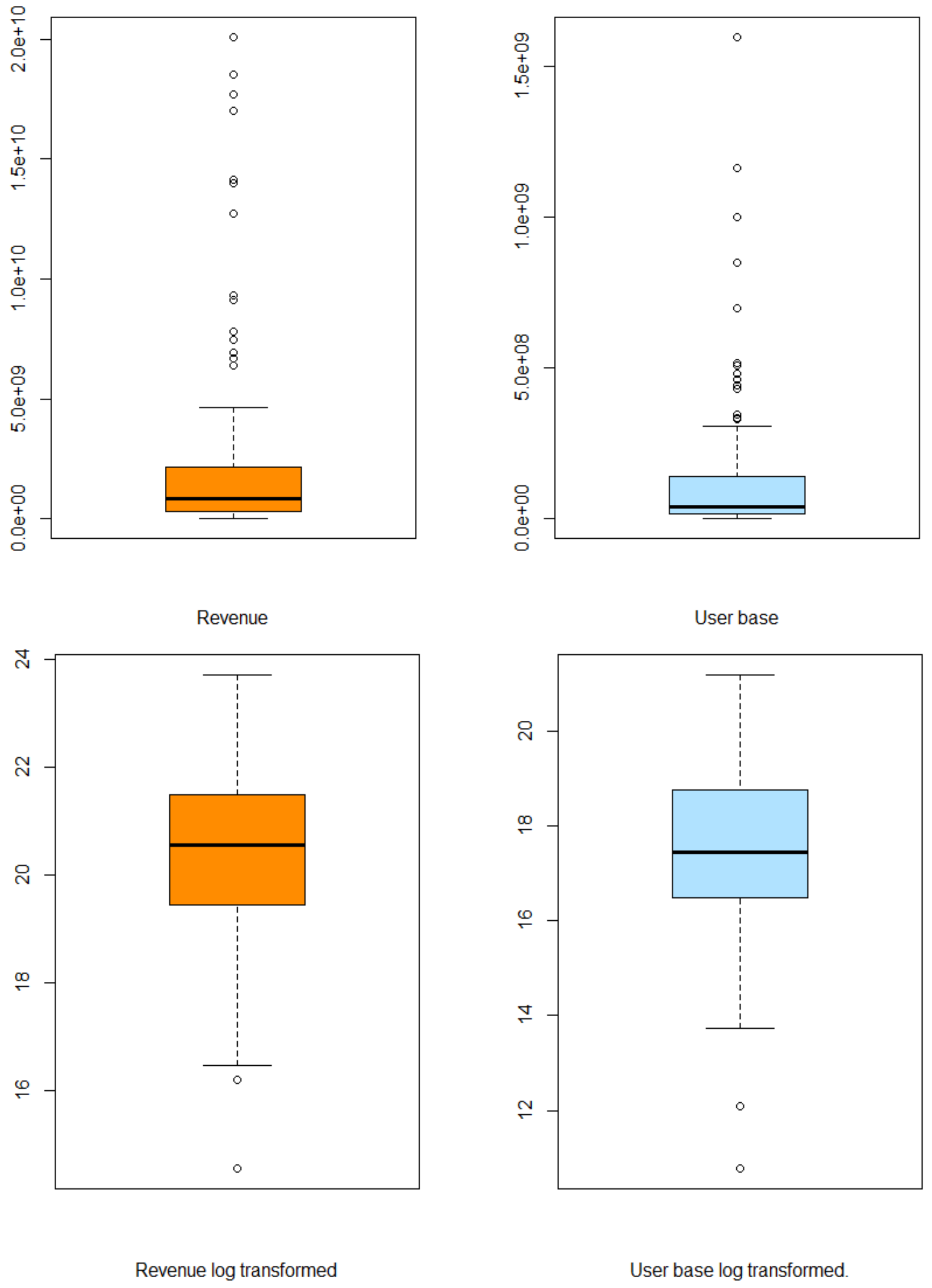


Figure D.1 Revenue and user base variable ranges plotted to depict the range in which the results can be applied. Nominal and log transformed values.

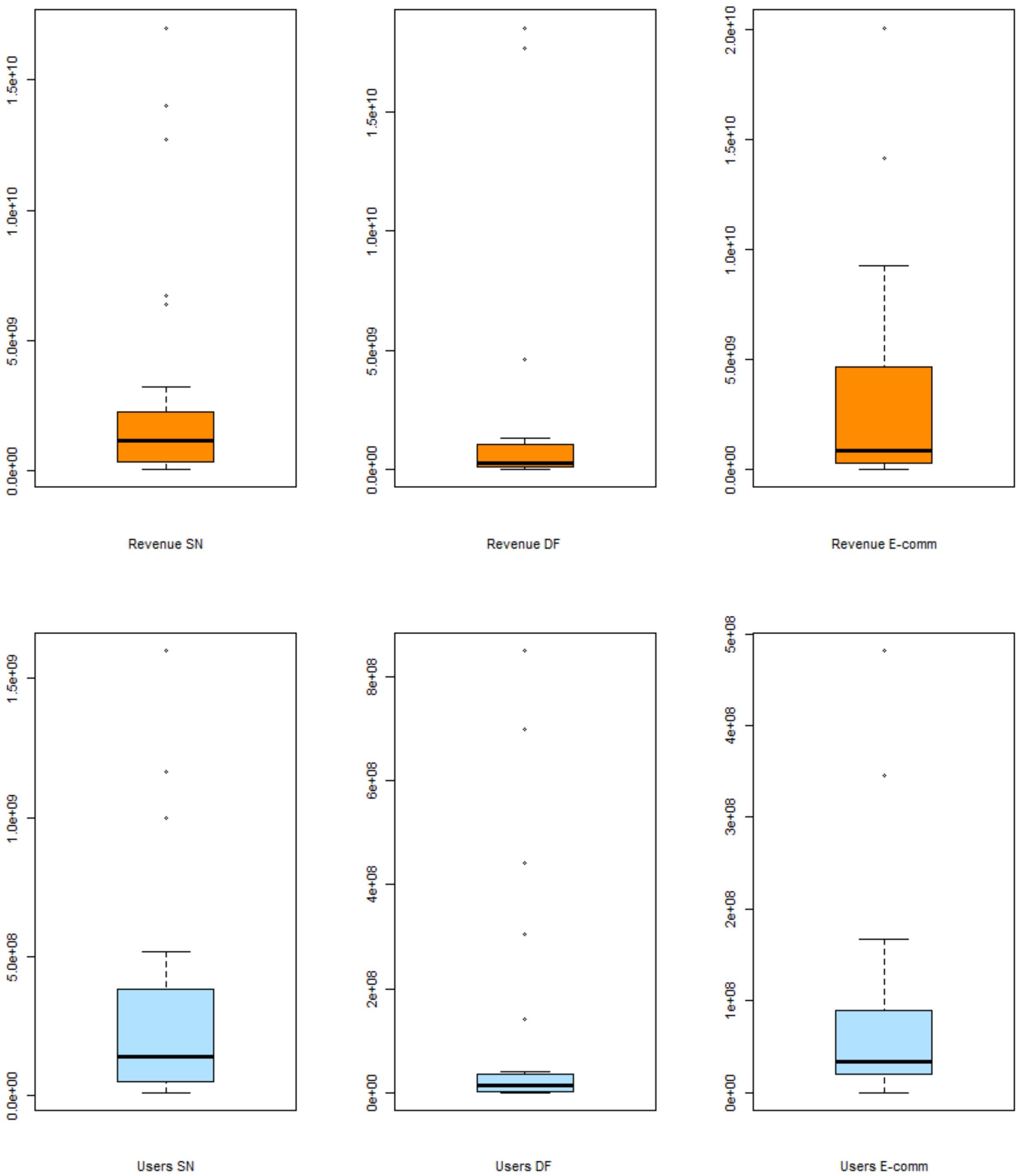


Figure D.2: Distribution of revenue and user base according to industry segmentation. Nominal values. SN = social networks, DF = digital finance, E-comm = e-commerce.

Appendix E: Bootstrapped Variables

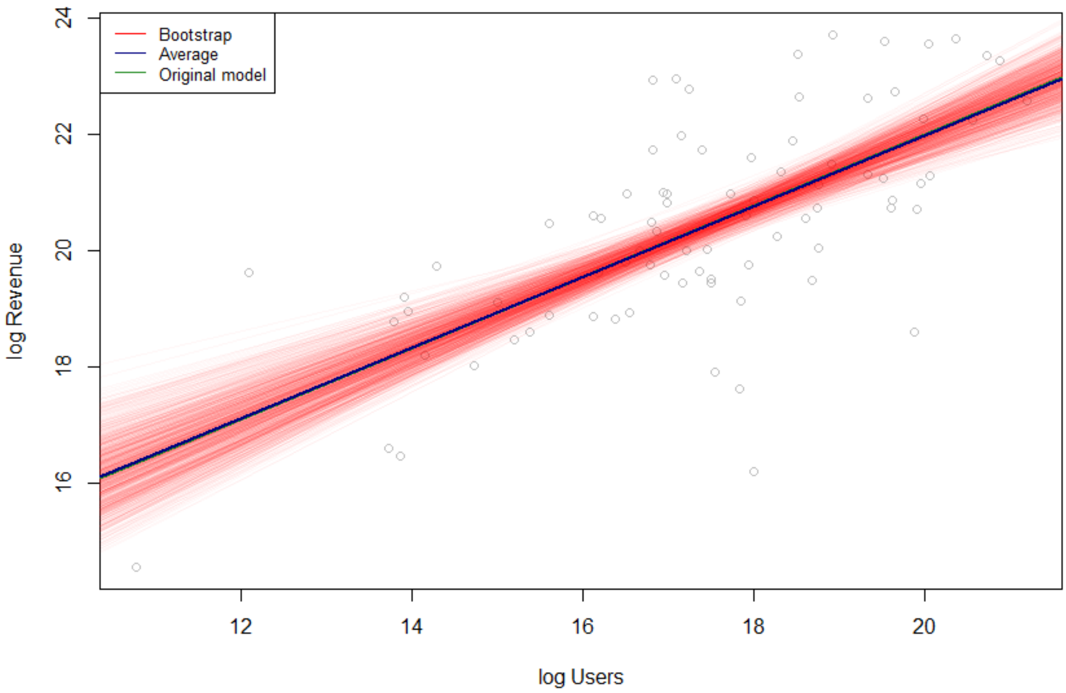


Figure E.1: Regression model 1 with 1000 regressions on 1000 bootstrapped samples indicating a strong relationship between user base and revenue. The average bootstrapped model and the original model plotted on the original dataset is so similar that the difference is not visible in the figure.

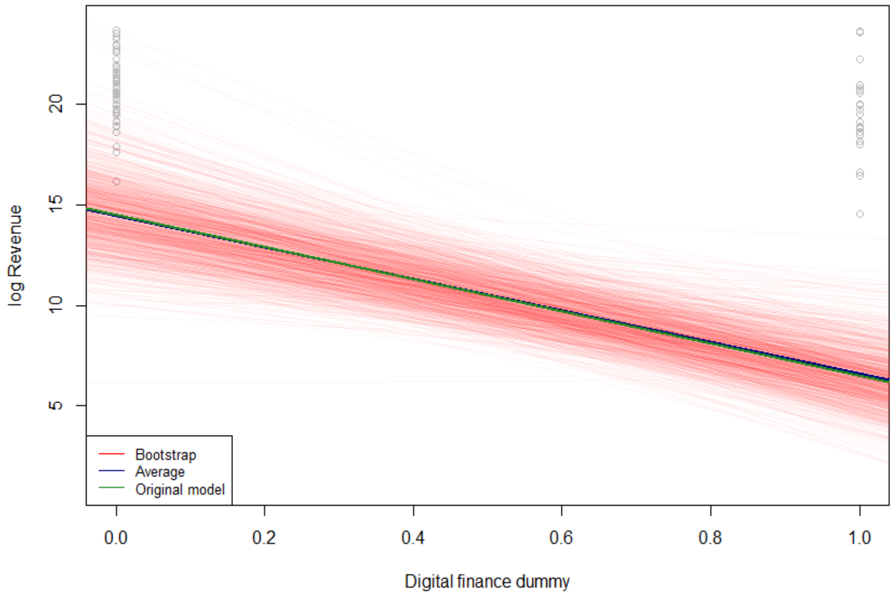


Figure E.2: Bootstrapped model with 1000 regressions on 1000 samples demonstrating a lower intercept for social network companies. Average bootstrap model and original model are hard to see since they are similar in slope and intercept.

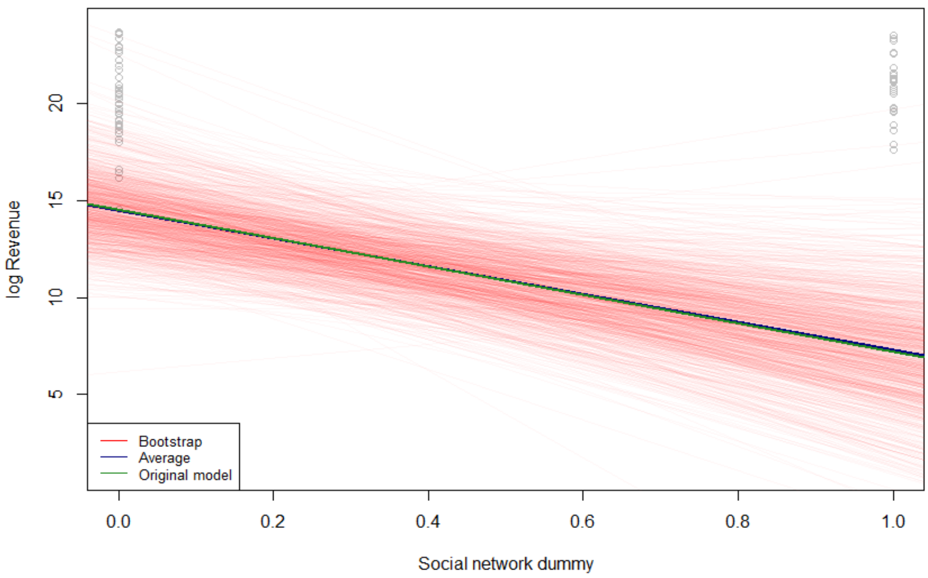


Figure E.3: Bootstrapped model with 1000 regressions on 1000 samples demonstrating a lower intercept for digital finance companies. Average bootstrap model and original model are hard to see since they are similar in slope and intercept.