Skolkovo Institute of Science and Technology

THE DATA-DRIVEN MODEL OF TECHNOLOGY-BASED NEW VENTURES GROWTH

Doctoral Thesis

by

MAKSIM MALYY

DOCTORAL PROGRAM IN ENGINEERING SYSTEMS

Supervisor
Associate Professor Zeljko Tekic

Moscow - 2022

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgment is made and has not been submitted for any other degree.

Candidate (Maksim Malyy)

Supervisor (Prof. Zeljko Tekic)

Abstract

This doctoral thesis proposes and tests a novel data-driven approach for describing technology-based new ventures' (TBNVs) growth trajectory. Due to its conceptual character with very limited empirical evidence, the actual Organizational Life Cycle (OLC) theory cannot serve as an accurate instrument for operationally identifying TBNVs evolution phases and transition points. However, that is a remarkably important task, directly influencing the process of new venture development.

To solve this problem, a two-phase empirical study was implemented. During the first phase, was validated a new source of high-quality and easy-to-access data for analyzing TBNVs' growth trajectories – Google Trends (GT). Utilizing the diverse sample of 241 US-based TBNVs, I comparatively analyze the relationship between companies' evolution curves represented by search activity on the one hand and by valuations achieved through rounds of venture investments on another. The results suggest that these valuations reflecting TBNV's growth dynamics are positively and strongly correlated with their web search traffic across the sample. This correlation is more robust when a company is a) more successful (in terms of valuation achieved) – especially if it is a "unicorn"; b) consumer-oriented (i.e., b2c); and 3) develops products in the form of a digital platform. Further analysis based on the fuzzy-set Qualitative Comparative Analysis (fsQCA) shows that for the most successful companies ("unicorns") and consumer-oriented digital platforms (i.e., b2c digital platform companies) proposed approach may be extremely reliable, while for other high-growth TBNVs, it is useful for analyzing their growth dynamics, albeit to a more limited degree.

In the second phase, I examined the shapes of TBNVs' growth trajectories built from GT data related to them. In the beginning, the selected growth models were fitted to these data that demonstrated that the S-curve models represent TBNVs' growth more accurately than the other growth-related models. Next, bearing in mind that the common sigmoid equations have limited applicability due to their intrinsic autocatalicity leading to the inability of determining starting point

of the growth, I applied the more advanced S-curve model in the form of probability distribution (i.e., *new products diffusion model*, the Bass distribution) and confirmed that it can accurately describe the TBNVs' growth while providing the mechanism to identify tipping and inflection points of the curves. Since the tipping point precedes accelerated growth and, hence, plays an extremely crucial role for a commercial organization, it can be argued that the proposed data source and growth model will add a significant contribution to the organizational lifecycle theory and practice.

Keywords: new venture, Google Trends, organizational life cycle, new product diffusion, tipping point, S-curve.

Publications

Peer-reviewed publications indexed in Scopus / Web of Science

1. Malyy, M., Tekic, Z., & Podladchikova, T. (2021). The value of big data for analyzing growth dynamics of technology-based new ventures. Technological Forecasting and Social Change, 169, 120794. **[2020: Q1, IF = 8.593]**

2. Malyy, M., Tekic, Z., & Golkar, A. (2019). What Drives Technology Innovation in New Space? A Preliminary Analysis of Venture Capital Investments in Earth Observation Startups. IEEE Geoscience and Remote Sensing Magazine, 7(1), 59-73. **[2020: Q1, IF = 13]**

3. Malyy, M., & Tekic, Z. (2018). Role of Startup's Intellectual Property in VCs Investment Decision-Making: Evidence from Russia. In 2018 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE) (pp. 1-7). IEEE.

Conference presentations

1. Chemikhin A., Malyy, M. & Tekic, Z. (2021). First to patent or first to trademark – what drives unicorns' high growth? Presented at the R&D Management Conference 2021, Innovation in an Era of Disruption, Glasgow, UK, 6-8 July 2021

2. Malyy, M. & Tekic, Z. (2019). Understanding Startups' Dynamics: Big Data Approach, Presented at the R&D Management Conference 2019, The Innovation Challenge: Bridging Research, Industry & Society, Paris, France, 17-21 June 2019

3. Malyy, M., & Tekic, Z. (2017) Technology trends and invisible hand of VCs: evidence from Earth observation start-ups, Poster presented at Gen-Y: Skoltech Young Scientists Cross-Disciplinary Conference, Sochi, Russia, 27 Sep – 1 Oct 2017

Acknowledgments

I would like to express my gratitude to my supervisor, mentor, and friend, Prof. Zeljko Tekic, for continuous guidance and unmeasurable direct support during the whole research period. I believe we have become a good team. Also, I wish to show appreciation to professors who helped me to shape this study into the form it finally got, Prof. Alessandro Golkar and Prof. Tatiana Podladchikova. Your methodological advice was priceless. Additionally, I would like to thank the defense commission for providing invaluable, detailed feedback on my analysis and the final framing, Prof. Clement Fortin, Prof. Alexey Nikolaev, Prof. Dmitry Koroteev, Prof. Fred Phillips, and Prof. Vincent Mangematin. Your reviews also pushed me to think broader and see other derivative possibilities for future research. Finally, I would like to thank my family and especially my wife Tatiana, who was near me all the time, experienced my “backstage” feelings and helped me find the motivation to continue during the most challenging periods.

Table of Contents

List of Symbols, Abbreviations

ACC – adopted cross-correlation

Adj. R^2 – adjusted R-squared

Adj. R^2 cv – adjusted R-squared cross-validated

AIC – Akaike information criterion

aHD – adopted Hellinger distance

b2c – business-to-customer

b2b – business-to-business

DTW – dynamic time warping

fsQCA – fuzzy set Qualitative Comparative Analysis

GT – Google Trends

IPO – initial public offering

M&A – merger and acquisition

MCAP-GR – the market capitalization growth rate

OLC – organizational life cycle

PRI – Proportional Reduction in Inconsistency

RMSE – root mean squared error

RQ – research question

SME – small and medium enterprises

TBNVs – technology-based new ventures

VC – venture capital

List of Figures

List of Tables

Chapter 1. Introduction

Every organization, whether it is commercial or not, due to its definition, has a particular purpose and, in an ideal setting, builds its operations in order to accomplish it[1]. Although this purpose may take various forms, the main goal of commercial ones and, at the same time, their distinguishable feature is to generate value in a market and, thus, profits for their stakeholders as a result of performing operations, i.e., doing a business[2]. Therefore, under the first approximation, reaching a commercial goal may be simplified to a straightforward linear model: while being in point A with zero profits, move to point B with N amount of earnings. However, the world is continuously changing, and organizations have to alter over time, adapting to new factors in order to succeed. These adoptions include adjusting all organizational aspects like leadership, innovativeness, openness, structure, and others, which form the distinguishable patterns and shape the stages of the organizational life cycle (OLC) (Kazanjian, 1988; Miller and Friesen, 1984). Applying the living organisms' analogy, organizations also evolve from birth until death (or revival) while attempting to accomplish their initial purpose of generating profits through delivering value to the market. Thus, the process of moving a commercial entity from zero to success can be divided into particular stages, each of which reflects a specific state of an organization.

In each particular stage of its lifecycle, a commercial organization faces specific problems, sets the corresponding goals, and attempts to reach them to survive. Inconsistency between stages and related managerial decisions may lead to serious negative effects like wasting scarce resources, losing competitive advantage, and closing a company. For instance, one of the common mistakes is to

[1] ORGANIZATION | meaning in the Cambridge English Dictionary [WWW Document], n.d. URL https://dictionary.cambridge.org/dictionary/english/organization (accessed 9.26.20).

[2] What is a business? definition and meaning - BusinessDictionary.com [WWW Document], n.d. URL http://www.businessdictionary.com/definition/business.html (accessed 11.11.18).

manage a high-potential technology-based new venture (TBNV[3]) as a regular mature company during the early stage of its development, i.e., to employ bureaucratic procedures, blur resources to unnecessary operations, or hire high-cost top management[4]. The opposite is also true: to sustain the growth, developed organizations should not, for instance, have an uncertain hierarchy or unclear responsibilities. Skillful managers understand it intuitively (Adizes, 1979; Kazanjian, 1988; Miller and Friesen, 1984), but currently, the majority of startups, from which TBNVs may grow, are launched by the complete "newcomers" who at best had some work experience in the related sector. These nascent entrepreneurs try to apply practices seen in their job places or heard during MBA studies, which will most likely not be suitable for their ventures due to the differences in stages. As a result, this discrepancy often leads to bankruptcy and the close of a startup despite the promising of the initial idea.

In the literature, the question of organizational lifecycle models for TBNVs (Kazanjian, 1988) and regular companies (Greiner, 1972; Miller and Friesen, 1984; Scott and Bruce, 1987) was discussed many times for more than fifty years. Dozens of models and divisions on stages were proposed, each with sustainable theoretical argumentation and some with attempts to find an empirical backing (Churchill and Lewis, 1983; Kazanjian and Drazin, 1989; Miller and Friesen, 1984; Smith et al.,

[3] *High potential* underlines that companies focus on and ability to achieve annual sales of over $10M during the first several years of existence (Roure and Keeley, 1990). In the case of a "startup," which is known as the early stage of a TBNV (Kazanjian, 1988; Lee and Lee, 2006), the *high potential* quality is driven together by scalability and repeatability features (Blank, 2013; Kollmann et al., 2016). Canonically, technology-based new ventures itself represent the firms "recently established by a group of entrepreneurs, based on the exploitation of an invention or technological innovation and which employ a high proportion of qualified employees" (Campos et al., 2011). These companies are understood to be in the focus of venture capitalists and "face unusual time pressures and uncertainty" (Roure and Keeley, 1990). TBNVs are used to be studied for more than thirty years with particular interest on the growth trajectories (Kazanjian, 1988; Lee and Lee, 2006; Strehle et al., 2010) and defining predictors of success (Reymen et al., 2017; Roure and Keeley, 1990; Shrader and Siegel, 2007; Symeonidou et al., 2017). To the purpose of the current research, under *Technology-based new ventures (TBNVs)* are understood entrepreneurial ventures that extensively apply technologies to create and deliver value to customers and stakeholders. They do that by integrating technologies into their products and services in novel ways, by crafting new business models based on tech affordances, or doing all of it together with the aim to achieve and sustain a venture level of growth.

[4] 353 Startup Failure Post-Mortems [WWW Document], 2020. CB Insights. URL https://www.cbinsights.com/research/startup-failure-post-mortem/ (accessed 9.26.20).

1985). However, vague and “loosely defined” (Hanks, 1990) stages in these conceptual models led to the low, at best, practical applicability and to the opinion that “stages of growth modeling has hit a dead end” (Levie and Lichtenstein, 2010).

Empirical studies, in their turn, are relatively scarce, questionable in methodology, and limited in conclusions that make them hardly applicable in real-life managerial practice (Al-Taie and Cater-Steel, 2020; Hanks, 1990; Jirásek and Bílek, 2018; Levie and Lichtenstein, 2010; Phelps et al., 2007). Moreover, none of the models evidences the form of the evolution trajectory, although many assume it to represent an S-curve growth analogous to biological systems (Penrose, 1952; Picken, 2017; Söderling, 1998). Therefore, it may be concluded that existing OLC models are conceptual that is of some value for the theoretical design of a TBNV, but a moderate, at best, applicability for practice. That results in a faulty understanding of TBNV's growth process and in the following imprecise managerial decisions.

From my perspective, the conceptual character of the known OLC models with a lack of empirical confirmation is the result of the data scarcity problem. Although startups attract a lot of interest from researchers, policy-makers, nascent entrepreneurs, and managers alike, analyzing their growth and performance is a challenging task due to the lack of high-quality and directly observable data. Namely, these fragile, early-stage private businesses, which may quickly become high potential TBNVs, do not have time, interest, or obligation to share much data about what they achieved, when, and how. Even when the startup phase ends and a company starts to scale up, the data scarcity issue does not disappear. For an outside observer, it is almost impossible to get enough objective information on a particular new venture’s progress until it becomes public (i.e., carries out IPO); that is, however, also a relatively rare case.

To solve this issue, scholars use various methods and data sources. One of the most accurate approaches is to apply a company valuation, achieved through various funding rounds, as a proxy performance measure (Chang, 2004; Gornall and Strebulaev, 2017; Malyy et al., 2019). In this approach, the growth trajectories of new ventures are reflected by valuation data (DeTienne, 2010). Although it is generally accepted as a good representation of a new venture evolution (Davila et al., 2003; Gornall and Strebulaev, 2017; Malyy et al., 2019; Ratzinger et al., 2018), this approach has (at least) two significant limitations. First, valuation data frequently are non-disclosed (Malyy et al., 2019), and, second, there are few companies that secure venture capital or angel investments annually, even in the most developed markets (Gornall and Strebulaev, 2017). Due to these limitations, there are three other proxies most frequently used in the literature to measure the growth of new ventures (Colombo and Grilli, 2010; Davila et al., 2003; Gilbert et al., 2006): sales growth (variation in sales expressed as value), employee growth (variation in the number of employees), and market share growth (variation in the controlled share of a market). These proxies are subjected to particular limitations as well. Market share is the indirect measure, which may be affected by the industry dynamics itself (Davila et al., 2003), while variations in sales and employees are rarely available for an outside observer in a statistically significant number of cases. Finally, different measurement practices across companies may affect interpretation of these metrics significantly.

The frequent way to solve the information availability problem is to communicate with top managers and venture founders directly through two typical channels: online questionnaires (Audretsch et al., 2012; Lee et al., 2001; Zhou et al., 2016) or in-person interviews with cofounders and decision-making management (Bocken, 2015; Carter et al., 1996; Delmar and Davidsson, 2000; Malyy and Tekic, 2018). Without a doubt, these data collection methods have the power to provide valuable first-hand insights, but, at the same time, they have significant limitations, which are hard to overcome. Online questionnaires are subjected to low response rates (10-15% in the best cases) and selection bias (Audretsch et al., 2012; Lee et al., 2001; Zhou et al., 2016) that may raise questions about the generalizability of the conclusions (Blair and Zinkhan, 2006; King and He, 2005). Similarly, while requiring significant time and resources to implement, in-person interviews with ventures' founders may provide inherently biased information due to the inability to remove the founders' subjective perception of events and measures (Al-Taie and Cater-Steel, 2020; Kazanjian and Drazin, 1989; Lester et al., 2008, 2003). These studies are commonly employed for qualitative research design while aiming to identify research directions for a subsequent quantitative study (Jarratt, 1996; Qu and Dumay, 2011).

Considering these data and methods-related drawbacks, in this thesis, I first ask *(1) which data of a TBNV can demonstrate its growth*[5]*?* Through answering it, I propose and examine the objective, widely available, and rich source of new ventures information. After that, this source is employed for analyzing TBNVs evolution trajectories, and thus, provides an answer to the second supporting research question: *(2) how does the trajectory of TBNVs growth change over time?* To solve it, I comparatively analyze the existing frequently applied growth models and develop several research propositions, which, in their turn, are intended to reply to the key research question of the thesis: *how do technology-based new ventures grow?* By finding an answer to it, it will be possible to enhance the practical applicability of the existing OLC theory and propose new precise practice-oriented instruments for TBNVs growth analysis. The graphical overview of the thesis structure is presented in Fig.1.1.

The first study within this doctoral research aims to answer the first supporting research question and contribute to overcoming the data scarcity problem in studying startups and high

[5] It should be noted that under the term *growth* is understood a period of a TBNV's evolution from its birth to the last available valuation point when a company stayed private, even if at this point a company demonstrated a decline in valuation. This understanding is based on the common knowledge of TBNVs development process and implies that its one of the main tasks is to reach an exit, i.e., to sell the company shares (Cumming, 2008; DeTienne, 2010; DeTienne et al., 2015). Since the best possible result of this process is exiting with the *maximum* possible valuation, the pre-exit period is understood here as *growth*, although for some companies, it might have ended with decline. In other words, in the first study, I aim to determine if TBNVs' valuation growth (or pre-exit) period correlates with their dynamics in the analyzed data source and, if yes, apply the growth (or pre-maximum) part of these data for analysis of their growth trajectory.

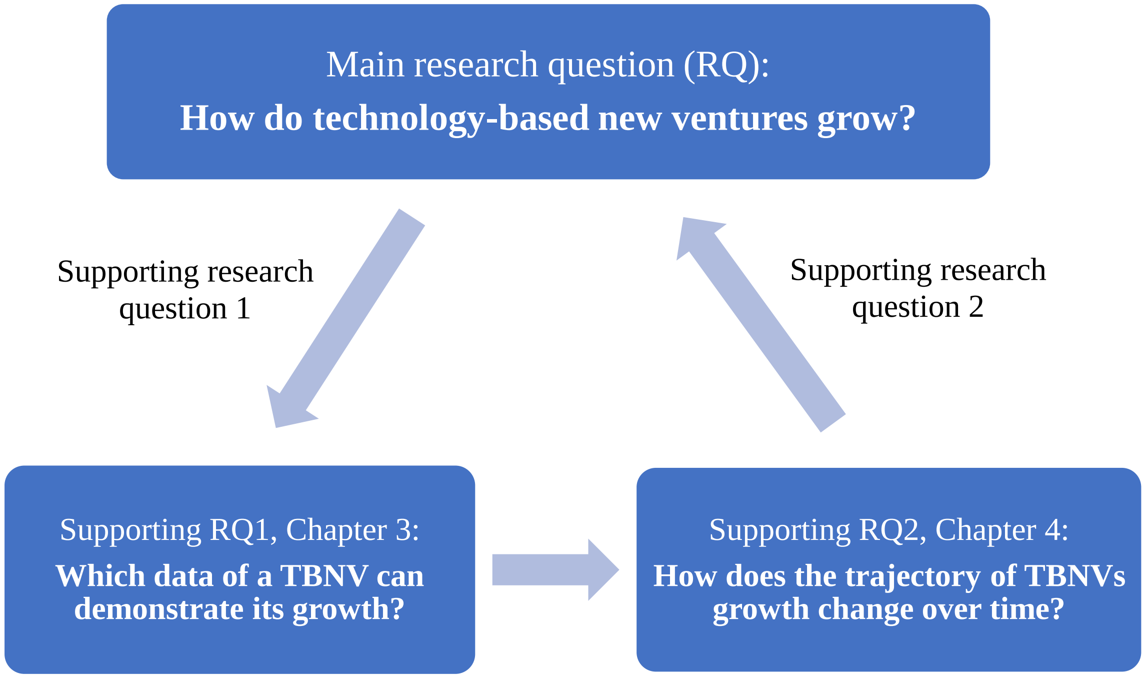


Figure 1.1. The structure and logic of the thesis.

potential technology-based new ventures, which evolve from them. Bearing in mind that TBNVs' valuations achieved through the series of venture funding are often used as a growth proxy (Chang, 2004; Davila et al., 2003; Gornall and Strebulaev, 2017; Malyy et al., 2019), I analyze the correlation between companies' valuation history and the web-search traffic time series connected to them. Search query statistics were evidenced to increase the accuracy of forecasting sales of various products and services (Choi and Varian, 2012) and especially the novel ones (Jun et al., 2014a, 2014b). That makes it possible to hypothesize the existence of the positive connection between TBNVs search traffic time series and their valuations. For the source of search query data, Google Trends[6] (GT) is applied – an instrument that proved itself to be a credible source of scientific information in more than 1,800 studies[7] TBNVs' valuations, in their turn, are collected from two dominant databases on startups

---

[6] Google Trends [WWW Document], n.d. URL https://trends.google.com/trends/?geo=US (accessed 11.26.18).

[7] Scopus - Document search results [WWW Document], n.d. URL https://www.scopus.com/results/results.uri?sort=plf-f&src=s&st1=%22Google+Trends%22&sid=8ccf315f70e783499658ae655faa66ac&sot=b&sdt=b&sl=30&s=TITLE-ABS-

and TBNVs – Crunchbase[8] and CB Insights[9]. The consequentially built research sample contains 241 US-based TBNVs from a variety of industries and consists of business-to-business (i.e., b2b) and business-to-customer (i.e., b2c) companies, "unicorns" and "non-unicorns," digital platforms, and traditional products (Appendix A). Based on the Fisher z-transform criteria for the correlational research design (May and Looney, 2020), this size of the sample is sufficient for getting the results with 90% power and 0.05 level of statistical significance.

The results suggest that for most TBNVs (83% of the sample), growth dynamics reflected by their valuations are positively and strongly correlated with the associated GT search traffic. In particular, it was found that this link is stronger for TBNVs, which are (a) "unicorns," (b) business-to-consumer oriented, (c) delivering value as digital platforms. Additionally, to find out if any feature or combination of features significantly influences this link, I implement the fuzzy-set Qualitative Comparative Analysis (fsQCA). The results of this step demonstrate that two equifinal combinations exist: to provide a stronger correlation, a TBNV should be either a "unicorn" or a b2c digital platform. However, a similar combination for the low correlation was not observed, meaning that TBNVs with any other configuration of taken features can also demonstrate the strong relationship between valuation history and GT search query statistics. Altogether, according to these results, it can be concluded that the related to a TBNV web search query data are a credible source of rich public information reflecting their growth dynamics.

The second study within the thesis, in its turn, proposes a fresh view on the OLC theory while applying previously mentioned Google Trends data as the source of information about TBNVs'

KEY%28%22Google+Trends%22%29&origin=searchbasic&editSaveSearch=&yearFrom=Before+1960&yearTo=Present (accessed 4.03.22)

[8] Crunchbase: Discover innovative companies and the people behind them [WWW Document], n.d. URL https://www.crunchbase.com/#/home/index (accessed 9.4.17).

[9] CB Insights: Machine Intelligence Platform [WWW Document], n.d. URL https://www.cbinsights.com/ (accessed 10.5.18).

growth dynamics. Since it was shown that TBNVs valuations and GT dynamics are positively and strongly correlated (in the majority of cases), in phase one, I utilize the latter to mathematically analyze the companies' growth trajectories, compare various growth-related mathematical models, and conclude on the general shape of TBNVs' growth trajectory. To avoid the previous limitation on the number of available valuation points, I use the different sample of 246 US-based TBNVs (Appendix B), fit the selected analytical models to TBNVs' GT data, and compare the outputs with various quality-of-fit measures. The results demonstrate that compared to other growth models, S-curves (in particular, the *logistic model*) provide a more accurate description of TBNVs growth dynamics by five out of six taken measures. In addition, 78% of cases show a stronger link to the S-shape by the results of cross-validation analysis, which is known to be frequently used to examine the model's forecasting power (Kuhn and Johnson, 2013). In other words, these results suggest that S-curve models have the higher power not just to describe but also to forecast TBNVs growth dynamics. Finally, the configurational analysis was applied, which showed that 17 out of 21 configurations lead to the S-curve models. From this fact, it can be inferred that the S-shape of the growth trajectory is not conditioned by the particular TBNV quality or combination of qualities and can be generalized to the out-of-sample companies.

Despite the fact that S-curves outperform other growth models, they have limited applicability if used to identify the trend-changing points of the curve, i.e., beginnings of the growth, stagnation, or decline. Due to their intrinsic autocatalyticity, it is not possible to identify the tipping point preceding the accelerated growth, which is known to play an extremely crucial role for a commercial organization or a particular product (Gladwell, 2000; Phelps et al., 2007; Phillips, 2007). Therefore, in the second phase of the second study, I employ the advanced model with an S-curve growth part, which has the theoretical (but so far, not empirical) potential to identify this tipping point under the *innovation/imitation* paradigm (Bass, 1969; Phillips, 2007): the Bass new product diffusion model. I fitted this model to the previous step TBNVs GT data and calculated coordinates of the tipping point, as well as the coordinates of the other meaningful points (Brdulak et al., 2021; Orbach, 2016). Visual observation of the obtained results made us conclude that in all given cases, the Bass model provided an accurate fitting and trustworthy tipping points preceding the accelerated growth start of the TBNVs GT curves.

According to these findings, it is proposed to use the taken new product diffusion Bass model as the model describing TBNVs evolution. It relatively accurately explains the growth dynamics of new ventures and is able to provide analytical division on three phases: (1) absence of growth, (2) accelerated growth, (3) de-accelerated growth ending with saturation. These three phases can be aligned to the existing OLC concepts. For example, according to some of the previous OLC concepts, the first phase may relate to the “startup” or “inception,” the second to the “scaleup” or “growth,” and the third to the “exit” phase or “maturity” (Miller and Friesen, 1984; Picken, 2017; Scott and Bruce, 1987). Since the proposed model is driven by the empirical data of TBNVs and, thus, directly linked to real life, each phase of the curve can be undoubtedly related to the particular period in company evolution with its specific state and combination of characteristics (i.e., leadership, innovativeness, openness, structure, etc.). Consequently, states of the companies can be precisely analyzed that will make it possible to develop better and growth-dependent managerial practices.

The achieved results have four key contributions. First, they bring value for retrospective (i.e., *post-hoc*) analysis of TBNVs. Since the dates of the transition points can be precisely identified, they can be directly related to preceding events in companies’ histories. This analysis can help to understand better the premises of particular decisions and assess their quality by measuring the change in the rate of growth. Second, curve parameters and their variation can be used to carry out between companies’ comparisons in order to identify potentially more successful practices and under various contexts. Third, by proposing and describing this model, I introduce a solution to the existing ambiguity, which occurred in the field of the organizational lifecycle theory. Since the model is mathematically straightforward in its description and application, it has the potential to become a single “frame” that can be applied to the existing OLC concepts and, thus, to align the proposed by them stages of growth. Fourth, the proposed source of data (i.e., Google Trends) eliminates the limitations, specific for the questionnaire-based (in other words, almost for all) OLC empirical studies (Garnsey et al., 2006; Levie and Lichtenstein, 2010). Due to this quality, GT data can be called a source of *objective* TBNVs information in terms of avoiding the in-person bias and, therefore, the more reliable scientific instrument for the organizational lifecycle field of research. Finally, the proposed tipping points may provide a solid theoretical backing for an ambiguous practical term of the *product/market fit* (Andreesen, 2007; Göthensten and Hellström, 2017). Since its invention in the mid-2000s by a number of US VCs[10], to the best of my knowledge, no study, article, or book has provided a straightforward and methodologically valid way of its identification. In other words, it is not possible to make any firm conclusion on whether a company reached the product/market fit and, if yes when they did it. At the same time, this event is known to be a well-known indicator of a company's quality during the VC decision-making process. The model proposed and tested in the current study is able to solve this issue by providing a precise theoretically-backed instrument for identification of this event that may be further employed to develop concrete practices, methodologies, and frameworks.

The remainder of the thesis is organized as follows. In the Literature review, the actual state of the art of the empirical sector of the OLC theory is observed. Next, I present Study 1 and Study 2, which provide the empirical outcomes to the two claimed research sub-questions. I proceed with the

[10] According to Andy Rachleff, the co-founder of one Silicon Valley VC fund, the term was originally invented by Don Valentine, the founder of Sequoia Capital VC fund (Rachleff, 2019). Later, it was widely used and popularized by Rachleff himself and other US VCs like Marc Andreessen (Andreesen, 2007).

discussion of integral results of the thesis in the Discussion part, following with Theoretical and Methodological contributions, Practical implications, and Limitations. Finally, I propose the ways for Future research and summarize the outcomes in the Conclusion part.

The thesis also includes four appendixes. In Appendixes A and B, I provide the list of TBNVs employed for analysis in Studies 1 and 2, respectively. Appendix C describes an algorithm for assessing the quality of Google Trends data for a search term related to a particular TBNV from the sample. Appendix D, in its turn, presents the first practical application of TBNVs related Google Trends data with a goal to understand the power of this instrument for estimating the TBNVs' valuations under the industry and segment-specific contexts.

Chapter 2. Literature review

The discussion on entrepreneurial companies' evolution process is almost as old as the origin of business. British economist Alfred Marshall was the first to add a notion of "organization" to the common agents of production: land, labor, and capital (Marshall, 1920). He understood new organizations as the young trees growing in the forest under the shade of older ones. The resources available to them are scarce, and only a few will survive, "reaching the light," i.e., the market top positions. This biological metaphor reflected the first appreciation of organizations as separate structures, which are born, evolve, and die. During the next eighty years, the discussion on entrepreneurial organizations' lifecycle gained a continuously increasing interest in the area of management research (Levie and Lichtenstein, 2010; Phelps et al., 2007). However, the first notable OLC concept of the current understanding was developed almost fifty years later after the work of Marshall, albeit it followed a similar biological metaphor (Levie and Lichtenstein, 2010; Phelps et al., 2007). In 1967, Lippitt and Schmidt proposed their view on the organizational evolution process based on the link between the phases of development and crises met (Lippitt and Schmidt, 1967). The authors claimed that during three stages of organizations' advancement (birth, youth, and maturity), they meet six particular challenges, solving of which will sustain transitions to the next phases. It was also insisted that the crises organizations face do not depend on their size in its traditional monetary meaning (Lippitt and Schmidt, 1967): "family business" SMEs[11] may be mature and industrial giants

[11] For the purpose of the current study, I reviewed all available and relatively popular works discussing conceptual and empirical OLC models, either focusing on TBNVs, high technology organizations, SMEs, or other growing commercial enterprises. Analysis of the existing literature led to the conclusion that the majority of OLC concepts are not being further discussed exclusively within the type of context for which they were originally developed. For instance, the influential work of Gaibraith (1982) employed the notion of "start-up ventures," which did not limit the study's utility only to this particular type of commercial entities but was further successfully applied for discussions related to "SMEs" (Rutherford et al., 2003) and "technology-based ventures" (Cavallo et al., 2019). In a similar vein, in OLC literature reviews, scholars often use generalizations of various organizational terms, calling their subjects "growing organizations" (Phelps et al., 2007), "firms (large or small)" (Tam and Gray, 2016), "enterprises" (Gupta et al., 2013), "business" (Muhos et al., 2010), etc.

can be at the youth stage. The idea of crises as a change predecessor was further developed by Greiner in his highly influential work "Evolution and revolution as organizations grow" (Greiner, 1972), having more than 8,000 citations of all versions[12]. Greiner asserted that each commercial organization develops through the five periods of relatively stable *evolutions* followed by turbulent *revolutions*, which, in their turn, lead to new evolutions. While during evolutions, companies steadily grow while employing the same managerial patterns, revolutions lead to the shakeouts, changing all meaningful routines. Thus, the critical task in revolution periods, which managers have to solve to sustain organizational growth, is to find and apply a new set of practices that will form the grounds for managing the next period of evolutionary growth (Greiner, 1972). Many more OLC concepts were developed in this early period of organization lifecycle area of knowledge (e.g., Adizes, 1979; Katz and Kahn, 1978; Lyden, 1975; Scott, 1970; Torbert, 1975); however, the majority of them reflect the theoretical side of knowledge and only a few provide empirical support (e.g., Churchill and Lewis, 1983; Hanks et al., 1994; Kazanjian and Drazin, 1990, 1989; Lester et al., 2003; Miller and Friesen, 1984; Smith et al., 1985).

One of the first attempts to prove the proposed OLC concept empirically was made by Churchill and Lewis in 1983 (Churchill and Lewis, 1983). The authors based their model on Greiner's curve (Greiner, 1972) and tested its phase division on the sample of 110 small companies ("with $1 million to $35 million sales range" (Churchill and Lewis, 1983)). Bearing in mind that companies' founders were familiar with Greiner's work, they asked them to identify the stages their companies had passed through to the best of their knowledge. After that, accounting for the questionnaire responses, the authors proposed the first empirical OLC model, which contained the five, thought to

---

[12] Evolution and revolution as organizations grow - Google Scholar [WWW Document], n.d. URL https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Evolution+and+revolution+as+organizations+grow&btnG= (accessed 8.4.21).

be proven, evolution stages (Churchill and Lewis, 1983): existence, survival, success, take-off, resource-mature. Although this framework goes in line with previous concepts, its methodological validity raises some concerns. First, the founders' level of knowledge, beliefs, and experience could significantly bias their understanding of the initial model and, thus, the attribution to the particular stage. Second, the sample was rather convenient than random since the questionnaires were distributed among the participants of a "small company management program" (Churchill and Lewis, 1983). Finally, the distribution of results raises some questions since some stages with a high number of responses were not included in the final concept, while some with lower numbers were.

The next notable empirical research on the OLC topic was performed by Miller and Friesen in 1984 (Miller and Friesen, 1984). The authors developed fifty-four variables describing a company's state during a particular stage and performed a longitudinal study of thirty-six companies and their 161 periods of history, trying to attribute each period to a particular conceptual lifecycle phase. The results demonstrated that, in general, existing OLC models were correct in their phases' description, but they were wrong in putting lifecycle phases into sequential order (Miller and Friesen, 1984). Despite the fact that this research was highly influential with more than 2,000 citations[13], it has particular limitations that decrease its generalizability and applicability. One of these limitations is connected with the fact that for some organizations, the full lifecycle was not observed. Based on the results, it can be seen that the majority of companies do not start with the birth phase that should have happened due to common sense (Miller and Friesen, 1984). One possible explanation for this drawback is the data scarcity issue because some companies from the sample were more than one hundred years old, and getting information about early ages would have become an impossible task.

---

[13] A Longitudinal Study of the Corporate Life Cycle - Google Scholar [WWW Document], n.d. URL https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=A+Longitudinal+Study+of+the+Corporate+Life+Cycle&btnG= (accessed 9.2.20).

Therefore, it cannot be undoubtedly claimed that some phases do not go sequentially: they simply might not have been observed in full detail to make rigid conclusions.

An empirical examination of organizational lifecycle concepts gained significant attention from academia and resulted in a number of doctoral dissertations. In this paragraph, the focus will be on three Ph.D. theses that laid the ground for the development of OLC theories and models. The first researcher to dive deeply into the empirical examination of the OLC problem was Robert Kazanjian, who proposed the case-study-based four-stage model laid in the basis of his future empirical analyses (Kazanjian, 1983; Kazanjian and Drazin, 1989). In his doctoral research, Kazanjian applied the cross-sectional survey study of 105 VC-backed high-technology ventures and demonstrated that the hypothesized model was supported by the empirical data: the structural specialization of the five primary new venture's functions showed the stage-related variability in four out of five cases (Kazanjian, 1983). Next, the author employed this model in the subsequent studies and enforced the initial conclusions with the results of longitudinal research (Kazanjian and Drazin, 1989). At this time, another scholar Steven H. Hanks came up with a doctoral dissertation, which was motivated by an opinion that "life cycle stage definitions remain vague and general, making application of the theory in specific cases difficult" (Hanks, 1990). To solve this issue, he proposed to step back and redefine life cycle stages according to the configurational approach claiming that each stage is a unique configuration of organization context, strategy, and structure. With a field study of 166 high technology organizations, Hanks supported the configurational approach, according to which at least four evolution stages were identified and their characteristics discussed (Hanks, 1990). Almost ten years later, another researcher, Matthew W. Rutherford, returned to the question of validating the OLC theory with empirical data adding value in terms of recommending between-stages thresholds and in the industrial context (Rutherford, 2001). The author employed the massive sample of 4600 SMEs from seven industries and demonstrated that in five of these industries, the OLC theory is a valid indicator of firm progression measured by the chosen contextual variables (size, age, and growth), however, with a remark that “exactly what organizational phenomenon that these are linked to remains elusive” (Rutherford, 2001, p. V). This work has also supported Kazanjian’s model of problem-driven transition between phases (Addison and Beazley, 2007).

In the most recent empirical work on OLC concepts, Al-Taie and Cater-Steel critically examined the psychometric properties of the five-stage model proposed by Lester, Parnell, and Carraher (Al-Taie and Cater-Steel, 2020; Lester et al., 2003). The authors employed a sample of 174 Australian IT executives, who were asked about various statements believed to reflect an attribution of the firm to a particular stage of growth. Next, they applied several statistical tests to assess various reliability and validity measures of the five-stage OLC model and concluded that this scale demonstrated “reasonable psychometric properties” except minor weaknesses, e.g., four constructs related to the stages of Success, Survival, Renewal, and Decline were claimed to be less precise and, thus, requiring the change in conceptual wording (Al-Taie and Cater-Steel, 2020, pp. 306–307). Despite the fact of rather positive outcomes, the study shares one significant limitation: the research is based on the perception of a single company manager who may be biased in his assessments for various reasons. Also, identification of the invalid model-related conceptual constructs raises questions on the process, which led to this result since it is obvious that in the initial empirical work of Lester, Parnell, and Caraher, these constructs were claimed to be trustworthy.

Altogether, the majority of empirical OLC studies can be separated into two big clusters according to the methods they apply: works, which utilize longitudinal data (i.e., data from the same companies but taken in various time periods) and those related to the cross-section analysis (i.e., taking information from a diverse companies’ sample at a single time point). The first method is known to provide more beneficial results since it makes it possible to describe a causal link between the taken variables (Gupta and Chin, 1993; Rutherford et al., 2003). In particular, based on the configuration theory, scholars hypothesize on the change of various parameters configurations (or *gestalts* as they called in early studies (Kazanjian and Drazin, 1989; Miller, 1987) relatively to the company stage and attempt to evidence and explain this change by empirical data taken with some delay. However, this method did not get high popularity, most probably due to its time-consuming character. I found only two studies utilizing an exact longitudinal approach (Kazanjian and Drazin, 1989; Miller and Friesen, 1984), two additional studies that use quasi-longitudinal data where scholars asked respondents to comment on the first (or past) year of the company existence while being years away from this point (McCann, 1991; Terpstra and Olson, 1993), and one longitudinal case study (Quinn and Cameron, 1983). In order to detect the change of examined parameters caused by the probable shift in lifecycle stages, it is needed to have at least an eighteen months delay between the data collection events (Kazanjian and Drazin, 1989) or from three to fifteen (!) years (Abetti, 2001; Greiner, 1997)**.** Despite the fact that this method looks more beneficial in the eyes of OLC empiricists (Bowman et al., 2012; Dufour et al., 2018; Gupta and Chin, 1993; Rutherford et al., 2003), it has another critical drawback: an examiner cannot be completely sure that the shift in stages has happened during the taken lag period, and if yes, how many times. This problem can be eliminated by more data-collection events, which, however, may result in a lower response rate (Kazanjian and Drazin, 1989), leading to the decline of external validity and increasing the probability of the non-response bias (Weaver et al., 2019).

The cross-section method gained significantly more attention from the OLC researchers, with more than 10 OLC empirical studies implemented during the last 30 years (Dufour et al., 2018). In this approach, the core idea is similar: search for the evidence on the stage-dependent change in configurations but implemented for the cross-section data reflecting the companies' states at a particular single moment (Lewis-Beck et al., 2004). Despite its inability to provide a causal link between the configurations of the taken variables, it is still possible to demonstrate the dependency of these configurations from a company stage and under some selected perspectives, like psychometric (Al-Taie and Cater-Steel, 2020), human resources (Rutherford et al., 2003), and top-level management priorities (Smith et al., 1985). The most common result of the cross-section studies is the proof that features configuration change in a particular way depending on the specific stage of OLC (Hanks et al., 1994; Lester et al., 2008; Rutherford et al., 2003; Shim et al., 2000) but without an answer on whether these configurations *drive* an organization phase transition.

Due to their nature, both clusters of methods share some limitations. First, since both of them are typically based on the surveys commonly distributed among single representatives of each sample company, the one-person bias cannot be excluded. This bias may be caused by various reasons, like denial (Lester et al., 2008), self-reporting (Al-Taie and Cater-Steel, 2020; Kazanjian and Drazin, 1989; Lester et al., 2003), or personal philosophy and personality (Lester et al., 2003). Second, due to the random principle of data collection, there is no guarantee that the sample will contain companies from all stages, so the results may not evidence on the particular stage existence simply because of the absence of the companies on this stage in the sample (Churchill and Lewis, 1983; Kazanjian and Drazin, 1989; Lester et al., 2008; Smith et al., 1985). Another shared limitation is also related to the sample collection and clustering process. In the studies utilizing the cross-sectional method, the need to control industry-specific effects is often raised in the limitations section and proposed to be solved in a "further" research (Gupta and Chin, 1993; Hanks et al., 1994; Lester et al., 2003; Rutherford et al., 2003). In works based on longitudinal approach, this problem is not so explicit but still possible: in one study, the industrial factor was well controlled by limiting the sample to the one particular industry (Kazanjian and Drazin, 1989), while in another this limitation wasn't raised and the control mechanism wasn't provided (Miller and Friesen, 1984). Finally, the configurational approach applied in the vast majority of empirical OLC studies leaves room for an error related to the wording variability. Of course, each scholar bases his concepts on previous studies, tries to follow (or at least to reference) the previous logic, and controls the results by statistical measures, but, at the same time, immediately creates new definitions and formulations. This leads to the “vague” definitions of growth stages (Hanks et al., 1994), inconsistency of models (Levie and Lichtenstein, 2010; Muhos et al., 2010), and, as a result, to an extremely low practical applicability (Al-Taie and Cater-Steel, 2020; Levie and Lichtenstein, 2010; Phelps et al., 2007).

Altogether, these method-specific and general issues were not solved in the previous forty years of empirical research and led to a quite unpleasant state of OLC knowledge succinctly described as the “dead end” (Levie and Lichtenstein, 2010, p. 318). In the related literature reviews, it was demonstrated that “stages theory is not appropriate for understanding business growth” (Levie and Lichtenstein, 2010, p. 329) due to the lack of consensus on the number of stages and their order (i.e., sequential or not) and the empirical studies weakness (Gilbert et al., 2006; Gupta et al., 2013; Levie and Lichtenstein, 2010; Muhos et al., 2010; Phelps et al., 2007; Tam and Gray, 2016). In general, I can agree with this assessment, however, not with reaching the “dead end.” Sharing the point of view of Garnsey, Stam, and Heffernan (Garnsey et al., 2006), I believe that the mentioned problems are mainly related to the absence of a solid methodological basis for conceptual and empirical alignment, having which will bring the majority of OLC models under a common denominator and significantly increase their practical applicability. Aiming to provide this denominator, I propose a new source of data and a mathematical apparatus, which together have the potential to overcome the mentioned drawbacks and open a new era of organizational lifecycle research.

Chapter 3. The analysis of TBNVs growth-related information source[14]

3.1 Introduction

Startups and high potential technology-based new ventures (TBNVs) that emerge from them are considered to be key drivers of economic development, innovation, and job creation on the national and global levels (Acs and Armington, 2006; Henrekson and Johansson, 2010; Kane, 2010; Mason and Brown, 2014). Bringing novel solutions to existing and emerging problems, startups create new value for their customers, at the same time increasing competition within the economy. More than that – startups are the main agents of disruption. Over the last 15 years, digital-native startups have been scaling globally from zero to billion dollars in value in just a couple of years, changing the logic of entire industries and setting standards for the next generation of products and companies (Tekic and Koroteev, 2019).

However, although startups are very important for the economy, and although they attract a lot of interest from researchers, policy-makers, nascent entrepreneurs, and managers alike, analyzing their growth and performance is a challenging task. Despite high and continuous research efforts (for in-depth reviews see, for example, Gilbert et al., 2006; Shepherd and Wiklund, 2009; Weinzimmer et al., 1998), our theoretical understanding of how (new) firms grow is limited and develops slowly (Coad, 2007; Gilbert et al., 2006; McKelvie and Wiklund, 2010). One of the major reasons for this is the lack of appropriate indicator(s) that will effectively capture growth (Weinzimmer et al., 1998) and help in adequately answering the question "how companies grow?" (McKelvie and Wiklund, 2010). Different methods and data sources have been used in the literature to measure the growth of new

[14] Based on the recent publication: Malyy, M., Tekic, Z., & Podladchikova, T. (2021). The value of big data for analyzing growth dynamics of technology-based new ventures. Technological Forecasting and Social Change, 169 (August 2021), 120794. https://doi.org/10.1016/j.techfore.2021.120794, JIF = 8.593 The author of this thesis was first and corresponding author. He completed all formal investigation. – if needed.

ventures. Three the most frequent are sales growth, employee growth, and market share growth (Colombo and Grilli, 2010; Davila et al., 2003; Gilbert et al., 2006). However, these proxies have particular limitations (Shepherd and Wiklund, 2009), which are amplified in the case of startups, temporary organizations (Blank, 2013) that develop new products under conditions of extreme uncertainty (Ries, 2011), and have no or very short operating and performance history. To resolve these issues, scholars recently started to use a company valuation, achieved through funding rounds, as a proxy growth measure (Chang, 2004; Gornall and Strebulaev, 2017; Malyy et al., 2019). However, this approach has the same significant limitation as earlier mentioned approaches – data frequently are non-disclosed. Even when the startup phase ends and a company starts to scale up, the data scarcity issue does not disappear.

The information availability problem is solved through direct communication with venture founders and top managers. Data are collected either through online surveys (Audretsch et al., 2012; Lee et al., 2001; Zhou et al., 2016) or face-to-face interviews with co-founders and higher management (Bocken, 2015; Carter et al., 1996; Delmar and Davidsson, 2000; Malyy and Tekic, 2018). Online surveys offer access to valuable data but suffer from sample bias due to the low response rates (10-15% in the best cases) and other selection criteria (Audretsch et al., 2012; Lee et al., 2001; Zhou et al., 2016). This particular drawback signals the questionable generalizability of conclusions achieved through this approach (Blair and Zinkhan, 2006; King and He, 2005). Face-to-face interviews with founders promise particularly insightful results but require substantial time, effort, and resources for implementation. Additionally, such studies are inherently biased as founders have a subjective perception of events, causes, and results.

Although researchers came up with a number of viable theories and explanations using the abovementioned methodologies, the lack of high-quality data and a number of issues impose significant limitations on researchers’ ability to discover more substantial patterns and connections between observable phenomena and advance theoretical understanding of firm growth (Coad, 2007; McKelvie and Wiklund, 2010; Shepherd and Wiklund, 2009). For example, during the last sixty years, scholars have been trying to explain and model the process of new venture development and proposing various organizational lifecycle models. According to the recent literature reviews, more than a hundred different models exist, of which all are conceptual (Levie and Lichtenstein, 2010; Salamzadeh, 2015). In other words, our understanding of new venture development is based on ideal constructs, lacking empirical validation and data verification. Of course, conceptual models are valuable and useful in a number of cases, but at the same time, they have numerous limitations (or can even mislead!), especially when it comes to practical usage (Coad, 2007).

This research aims to tackle these issues and to contribute to overcoming the data scarcity problem in studying startups and high potential technology-based new ventures. I do so by demonstrating the credibility of web-search traffic information as a novel source of high-quality data in analyzing growth trajectories of high potential technology-based new ventures (TBNVs) that emerged from them. Relying on the growing evidence that aggregated Internet search query data can be very useful in predicting underlying social and economic trends (Choi and Varian, 2012; Duwe et al., 2018; Jun et al., 2018; Wu and Brynjolfsson, 2009), I analyze the relationship between companies' growth trajectories represented by search activity, on the one hand, and by valuations achieved through rounds of investment, on another. I use a diverse and transparently selected sample of 241 US-based TBNVs from a variety of industries. The sample includes b2b and b2c companies, "unicorns" and "non-unicorns," digital platforms, and traditional products. Their valuation data are collected from two leading databases on startups and TBNVs – Crunchbase[8] and CB Insights[9]. The search activity is measured using Google Trends[6] (GT), a widely applied big data instrument (Jun et al., 2018).

The results suggest that TBNV's VC valuations reflecting their growth dynamics are positively correlated with their web search traffic across the sample. This correlation is stronger when a company is a) more successful (in terms of valuation achieved) – especially if it is a "unicorn"; b) consumer-oriented (i.e., b2c), and 3) a digital platform. In the second step, to understand better which TBNV's feature or combination of features (i.e., b2b vs. b2c, "unicorn" vs. "non-unicorn," and digital platform vs. traditional products) leads to achieving a high positive correlation between the TBNV's growth dynamics and its web search traffic, fuzzy-set Qualitative Comparative Analysis (fsQCA) is employed on the data. The results suggest that being a "unicorn" is a sufficient condition for the high positive correlation between the TBNV's growth dynamics and its web search traffic. However, it is not a necessary condition. A combination of consumer and digital platform orientation (i.e., b2c digital platform companies) is leading to the same result.

3.2 Theoretical background and hypotheses

Growth is crucial for startups – while established firms grow to sustain viability, startups grow to obtain it (Gilbert et al., 2006). Startups grow in a non-linear fashion (Garnsey et al., 2006) and almost always organically (McKelvie et al., 2006). However, the variance of their growth rates is considerably greater than in the case of established firms (Gilbert et al., 2006).

The new venture growth process is dominantly described by stage models of growth (Greiner, 1972; Kazanjian, 1988). However, the researchers are far from agreement about what stages exist, how many of them, and what are the relationships between them (Aidin, 2015; Levie and Lichtenstein, 2010; Zupic and Giudici, 2017). The alternative models, including those with tipping points (Phelps et al., 2007) and dynamic states (Levie and Lichtenstein, 2010), are also far from (wide) acceptance (Zupic and Giudici, 2017). Thus, the question "how startups grow?" stays a fundamental question that needs to be better understood (McKelvie and Wiklund, 2010).

One of the major limitations in answering the "how" question is the lack of the appropriate growth indicator to apply (Achtenhagen et al., 2010; Shepherd and Wiklund, 2009; Weinzimmer et al., 1998), especially in the context of growth-oriented new ventures. Different indicators have been used to measure growth, including sales growth, employment growth, asset growth, and equity growth. Followed by employment growth (variation in the number of employees) and market share growth (variation in the controlling share of a market), sales growth (variation in sales expressed as value) is the proxy measure most frequently used in the literature to measure the growth of new ventures (Colombo and Grilli, 2010; Davila et al., 2003; Gilbert et al., 2006). The logic behind this is that when sales grow, a venture's revenues grow as well as the venture's ability to reinvest into resource expansion or capability development (Gilbert et al., 2006). Also, sales growth is easy to translate across countries and industry contexts (Delmar et al., 2003) and closer reflects entrepreneurs' point of view (Zupic and Giudici, 2017). However, all these proxies, including sales growth, are subjected to particular limitations as well (Shepherd and Wiklund, 2009). For example, market share growth is the indirect measure, which is dependent on the industry dynamics (Davila et al., 2003); while sales happen, a new firm has to have a product or service available to sell (Gilbert et al., 2006). Also, different measurement practices across companies may affect their interpretation significantly.

To resolve these issues and better reflect entrepreneurs' point of view (Achtenhagen et al., 2010), scholars recently started using a company valuation, achieved through funding rounds, as a proxy growth measure (Chang, 2004; DeTienne, 2010; Gornall and Strebulaev, 2017; Malyy et al., 2019). In this approach, the growth trajectories of venture capital-backed new ventures are reflected by valuation data (DeTienne, 2010). Although it is generally accepted as a good representation of a new venture evolution (Davila et al., 2003; Gornall and Strebulaev, 2017; Malyy et al., 2019; Ratzinger et al., 2018), this approach has the same significant limitation as earlier mentioned sales, and employee growth (and many other approaches) – data are rarely disclosed and available, and for an outside observer, it is almost impossible to get enough objective information on a particular new venture's progress until it becomes public (i.e., carries out IPO) that is, however, also a relatively rare case.

3.2.1 Valuation and growth dynamics

Two basic premises of this research are that changes in market valuations of venture capital (VC) backed TBNVs reflect their growth dynamics and that statistics on web search activities is a good predictor of product acceptance by individuals or society. Based on them, it can be argued that search traffic information, in particular, Google Trends data, will reflect well the growth dynamics of TBNVs. Building on this argument, I develop hypotheses relating the growth dynamics of startups and TBNVs that emerge from them with web search traffic.

High potential technology-based new ventures are understood to have various stages of growth (Kazanjian and Drazin, 1990; Lee and Lee, 2006; Tzabbar and Margolis, 2017), which condition the multiple series of VC investment – Pre-seed, Seed, Series A/B/C/etc., (Dahiya and Ray, 2012; Gompers, 1995) – and subsequent change in their market valuation (Davila et al., 2003). Financing through investment rounds is used to reduce information asymmetry and agency problems (Wang and Zhou, 2004). Each valuation event reflects investors' (re)assessment of a venture's ability to grow (i.e., generate future cash flows) and risks associated with that growth. To secure the first (and every other) funding round, TBNVs have to pass a systematic, disciplined and selective assessment. They have to provide VCs with evidence about their high-growth potential, and progress in realizing that potential, by indicating elements like the attractiveness of the market, soundness of strategy, the feasibility of the technology, the existence of product-market fit, customer adoption, and the quality and experience of the management team (Davila et al., 2003). There is an evidence that venture capitalists are highly selective while making a decision to invest (Gompers and Lerner, 2001; Zider,

1998) and are known to be focused on the fast growth of the company's valuation (Zider, 1998) with simultaneous diminishing the potential risk of investment (Davila et al., 2003). Thus, during the early stages, the more convincing the evidence about the venture's progress is (e.g., the market is proved, and its size is significant, technology is feasible, schemes for value creation and value appropriation are verified, first customers are acquired, or a number of customers and sales are growing exponentially), the lower the risks and the higher valuation the venture will achieve. With each next funding and valuation round, new ventures are expected to show more tangible results reflecting their dynamics.

In this paper, the logic of Davila et al. (2003) is followed. Namely to examine the relevance of big data (i.e., Google Trends) for understanding the growth of new ventures, I study its relationship with changes in the value of equity. Equity valuation data for successive rounds of funding available from the CB Insights database allow me to estimate the growth of ventures over successive rounds (I selected only those with at least six investment rounds per venture to be sure sufficient data is present).

3.2.2 Web search and consumers' behavior

The big data taken from various sources have proven their value in management theory and practice. The studies discussing big data applications in terms of organization performance started to emerge in the early 2000s and gained strong momentum in the 2010s (Batistič and der Laken, 2019). Considering the research subject, the related literature can be divided into two major clusters: focusing on the impact of big data on firms' performance and discussing new sources of big data for decision making. The studies from the first cluster, for instance, cover the application of big data for supply chain orientation (Gunasekaran et al., 2017; Wolfert et al., 2017; Yu et al., 2019), the role of big data in companies' capabilities context (Akter et al., 2016; Mikalef et al., 2019; Singh and El-Kassar, 2019), and predicting firm's manufacturing performance (Dubey et al., 2019; Ren et al., 2019). The works related to the second cluster, in their turn, discuss such topics as the use of customers' reviews in understanding sales performance (Lee and Bradlow, 2011; Mudambi and Schuff, 2010; Sheng et al., 2019), the influence of social networks users' information on business decisions (Antretter et al., 2019; Bradbury, 2011; Tambe, 2014), and application of web-search statistics for performance prediction and "nowcasting" (Choi and Varian, 2012; Jun et al., 2018; Shim et al., 2001; To et al., 2007). Since in the current research, web search statistics are used as a source of big data, I will focus on reviewing this topic in more detail.

The search feature of the internet was identified by scholars as one of the most important for users (Maignan and Lukas, 1997). Klein (1998) demonstrated that the process of search is specifically useful for obtaining information about goods due to the low costs of receiving objective data. This point was supported by another study (Liang and Huang, 1998), which claimed that effective minimization of transaction costs for the consumer is the key driver for the successful online selling of any product. Additional motivation for extensive search is the need to reduce the uncertainty about the product and decrease the level of the potential risk of purchasing something inappropriate (Dowling, 1986; Mitchell and Boustani, 1994). Thus, search over the internet during the pre-purchase process fits the basic economic theories and behavioral motives. Further studies demonstrated that the search over the internet and subsequent purchase under particular circumstances should be treated as the dependent processes due to the intention of consumers to use the same medium for getting data and obtaining goods (Shim et al., 2001; To et al., 2007).

Google Trends (originally known as Google Insights for Search[6]) is a big data tool. It was launched by Google in 2006. The first evidence of its usage for analyzing social trends came in 2009 when Google scientists Choi and Varian presented how Google Trends (GT) data could be applied for predicting automotive, retail, and home sales as well as for traveling (Choi and Varian, 2012). In the same year, GT data were used in a study published in Nature - Ginsberg et al. (2009) developed a model for detecting influenza epidemics using GT search query data that was consistently 1-2 weeks ahead of predictions of Centers for Disease Control and Prevention (CDC). After these pioneering studies, GT data were used in more than 1,800 studies across scientific fields, including business and management[7].

The willingness of consumers to decrease the pre-purchasing risk increases the importance of searching for information about it, especially for novel products (Assael, 1992). This phenomenon was demonstrated by Goel et al. (2010), who used GT search query volume to analyze and forecast the opening weekend box-office revenue for recent films, first-month sales for new video games, and the rank of songs in the Billboard Hot 100 chart. He showed that search statistics are generally predictive of consumer activities like attending movies, purchasing music, or video games that will happen in days or even weeks in the future. Considering the innovative products, web search queries can be extremely useful to analyze new technology adoption and able to provide higher explanatory power than indices used in the past, such as the GDP growth rate, patent applications, and news coverage (Jun et al., 2014b). It was also shown that brand search statistics explained the purchase of new technology adopters better than solely the number of search queries about the technology itself (Jun et al., 2014b). This fact gives us the ability to assume that a web search of a TBNV brand name is a potentially strong predictor of future sales of this brand. In another study, Jun et al. (2014a) demonstrated that GT search traffic information related to a particularly innovative product serves as an accurate indicator of consumer attitudes towards it. In short, Google Trends data have a positive connection with actual consumers' interests in the novel products and new-technology brands and, by analyzing web search queries, it is possible to predict forthcoming sales, understand new technology levels of acceptance, and reveal apparent and hidden attitudes towards it. Thus, I hypothesize:

H1: Expressed in VC valuations, TBNVs' growth dynamics are positively correlated with web search traffic associated with them.

3.2.3 TBNVs' features as the dimensions for analysis

3.2.3.1 "Success" dimension: "unicorn" vs. "non-unicorn"

A new venture becomes a "unicorn" when it reaches a valuation of $1 billion or more during at least one private series of venture funding (Gornall and Strebulaev, 2017). The value of "unicorns" is generally driven by VCs' expectation that these TBNVs will significantly grow in the near future, becoming even more attractive for investment and, thus, generating above-average returns for the owners. These expectations are typically the consequence of investors' assessment (and perception) that to-be-unicorns are solving an important problem for a very large market in a way that is significantly faster, cheaper, safer, or more comfortable (rarely a combination of two or more of these features) than competing solutions, and can design a business model that will secure value appropriation and scale. Or that they are solving an important problem for a very large market for the first time.

Valuing new ventures in different ways – some as "unicorns," some as "non-unicorns" – VCs, as informed agents (Baum and Silverman, 2004) offer us rare insight about growth predictions of these companies. Although the "unicorn" status does not guarantee future success (even survival), these ventures are considered as the most successful in their class. Reflecting this, for example, Fan (2016) proposes that "unicorns" should be regulated differently than "non-unicorns" due to the greater potential influence on the market of the former and increase of their investment risks. Thus, two poles in the success dimension are "non-unicorns" (successful companies) and "unicorns" (the most successful companies).

Many TBNVs generate only modest profits when they become "unicorns" but indicate significant future user interests and value for potential customers and, thus, the potential for growth. This interest could be first observed among early adopters and tech enthusiasts, then other users, who search the web for gaining more information about a new venture and its product (which typically is known under the same name). A growing number of studies show (Jun and Park, 2016; Shim et al., 2001; To et al., 2007) that the intent of web searches plays an important role in the intent to purchase a product, that is, to realize growth potential. This seems to be especially true when products are new (Assael, 1992; Jun and Park, 2016). Based on this argumentation, I expect the difference in how search traffic information will reflect (i.e., correlate with) the growth dynamics in the case of “unicorns” and “non-unicorns,” and propose:

H2: Expressed in VC valuations, growth dynamics of TBNVs with the “unicorn” status are better correlated with web search traffic associated with them than growth dynamics of TBNVs with the “non-unicorn” status.

Achieving the “unicorn” status is not the only measure of a company’s success. It is useful but binary, potentially hiding much useful information. For example, it is not possible to distinguish between the effects of superior success (a “unicorn”) and success (high-growing TBNV, but not a “unicorn”). A more granulated view on this can offer measures of relative success that show the speed of valuation growth. For that purpose, I adopt the approach of Ramadan et al. (2015) and use Market Capitalization Growth Rate (MCAP-GR), a measure that shows the annual growth rate of a company’s market valuation. In the spirit of H2, I posit that the more successful TBNVs are, the better reflected their growth by web search traffic will be. However, this time the company’s success is not considered as a binary outcome, but through the prism of the speed of achieving success (i.e., maximum valuation), and hypothesize:

H2a: The more successful a TBNV is, in terms of speed of valuation achieved, the better correlation will be between its growth dynamics expressed in VC valuations and web search traffic associated with it.

3.2.3.2 Customer type dimension: b2c vs. b2b

Further, I am interested in how the type of customers served by a TBNV influences the correlation between web search traffic and its growth dynamics. There are two main customer types – individual and business customers. They define two poles of the customer type dimension. Companies that make commercial transactions primarily with other companies are called business-to-business (b2b) oriented. Those who primarily serve individual customers are called business-to-customers (b2c) oriented. There are significant differences between these two types of companies (Ellis, 2010). B2b sales process is usually subjected to long negotiations (Järvinen et al., 2012), which are often focused on facts, i.e., terms of trade, reliability of delivery, etc. (Rėklaitis and Pilelienė, 2019). Decisions to buy from b2c companies are made faster, more emotionally, and more frequently (Saha et al., 2014). Further, b2b businesses are typically specialized and have a smaller number of customers (with bigger bills) than b2c businesses with the same cash flow. That means that in the case of b2c companies, much more independent purchase decisions happen, driving up information search activities of potential customers. In line with this and previous research that showed that search traffic correlates more strongly with sales of consumer goods than that of industrial goods (Jun and Park, 2016), I hypothesize:

H3: Expressed in VC valuations, growth dynamics of b2c-oriented TBNVs are better correlated with web search traffic associated with them than the expressed in VC valuations growth dynamics of b2b-oriented TBNVs and the related to them web search traffic.

3.2.3.3 Product type dimension: digital platforms vs. traditional products

In this research, I distinguish between two types of digital products – individual (i.e., traditional) products and platforms[15] and use this differentiation to define two poles of this dimension. Platformization is a recent trend in developing products for a digitally connected world. It aims not at building stand-alone products but ecosystems that will profit from synergetic coexistence and collaboration of many complementary modules that are building their businesses on the platform (Nambisan et al., 2018). A number of sectors accepted digital platforms as an attractive business model and strategy (Asadullah et al., 2018). Typically, a platform provides the technological foundation (e.g., application programming interfaces and software development kits), legal protection (e.g., IP protection), and access to an established market (e.g., through the platform's existing user base and reputation). All this with the goal to incentivize partnerships and development of complementary products and services, leverage economies of scale, and scope in innovation (Gawer, 2014). Further, digital platforms increase efficiency by significantly reducing costs of distribution, search, contracting, and monitoring (Asadullah et al., 2018). Successful platforms, like those in tourism (e.g., TripAdvisor[16] and Expedia[17]) or software development (e.g., Apple iOS, Google Android), are those that attract numerous complementors, which create value for platform users (Nambisan et al., 2018). Due to the synergetic effects of the created ecosystem, digital platforms lead to an increased number of platform users, with subsequent enhancement of the platform value (Asadullah et al., 2018). At the same time, the offers of multiple platform modules (complementors)

[15] In this study, digital platforms are defined as "two-sided networks that facilitate interactions between distinct but interdependent groups of users" (Asadullah et al., 2018), mediated by digital technology (Hein et al., 2019), not as the products, which consist of the functional core and the amount of possible third-party modules.

[16] Tripadvisor: Read Reviews, Compare Prices & Book [WWW Document], n.d. URL https://www.tripadvisor.com/ (accessed 2.1.20).

[17] Expedia Travel: Search Hotels, Cheap Flights, Car Rentals & Vacations [WWW Document], n.d. URL https://www.expedia.com/ (accessed 2.1.20).

and the platform itself are expected to generate high information search activities of potential customers. Based on this, I hypothesize:

H4: Expressed in VC valuations, growth dynamics of digital platform-oriented TBNVs are better correlated with web search traffic associated with them than the expressed in VC valuations growth dynamics of traditional product-oriented TBNVs and the related to them web search traffic.

3.2.3.4 Configurational perspective

However, the three dimensions mentioned above do not exclude each other. New ventures are not only “unicorns” or b2b companies or platform owners. They may be all that at the same time – they can be “unicorn” b2b platforms or “non-unicorn” b2c traditional products or any of the other six possible combinations of these features. That means that there are eight different sub-groups of TBNVs, defined by a combination of the three features-dimensions ($2^3 = 8$). Although I predict (and aim to prove in this paper) that web search traffic is positively correlated with the growth dynamics of all high potential TBNVs, I also want to check if there are differences between sub-groups regarding this quality. I would like particularly to know if some TBNVs “always” lead to a high correlation with web search traffic while others do not. Or if there is a TBNV sub-group that has “always” low correlation.

I thus rely on configuration theory (Ketchen et al., 1997; Miller, 1986) to help us understand how the three dimensions combine to generate an outcome rather than how they individually compete to explain it. The fundamental proposition of the configuration theory is that outcome of interest can best be understood if patterns of causes are analyzed (Fiss, 2007; Misangyi et al., 2016). The specific causal patterns are called “configurations” (in my case, there are eight configurations, defined by different values of three features: company valuation, type of customer, and type of product). While emphasizing the importance of causal complexity, the configurational theory has three core assumptions (Misangyi et al., 2016): 1) there is rarely a single independent cause of an outcome and causes rarely operate in isolation from each other (multiple conjunctions); 2) different configurations can lead to the same outcome (equifinality), and 3) the configurations that lead to the presence and absence of an outcome are not symmetrically opposite to each other (causal asymmetry);. Based on configurational theory and previous discussion, I hypothesize:

H5: There is different configurations of TBNVs' features, i.e., the b2c vs. b2b, "unicorn" status vs. "non-unicorn," and digital platform vs. traditional product, that are equifinal in achieving a high positive correlation between the expressed in VC valuations TBNV's growth dynamics and web search traffic associated to it.

3.2.4 Analysis and results

3.2.4.1 Description of the sample and data collection

In order to test the developed hypotheses, the popular startup databases were used: Crunchbase[8] and CB Insights[9]. The former provides a convenient search tool with the ability to use various filters and get the sample of companies meeting the general boundaries. The study is focused on TBNVs, which were founded in the US not earlier than on January 1, 2004, and not later than on August 31, 2019. These companies also should have reported on at least six rounds of venture financing, thus making it possible for us to obtain enough valuation points for analysis. The initial sample consisted of 2774 companies. For each company in the sample, I manually checked if it had enough (i.e., six and more) valuation points in the CB Insights database and collected valuation data for positive cases. I also excluded spinoffs of existing companies as search query data may be biased by a "mother" company noise and, thus, be unpredictably distorted. This stage of data collection resulted in 269 cases, for which their industry tags were also aggregated according to the CB insights classifications.

The last applied filter accounted for the quality of the companies' GT data. Like any big data instrument, GT provides huge portions of information, which requires an assessment before usage. To assess the quality, I used a number of rules and developed a quantitative index (Appendix C). By applying them, I excluded 28 companies whose search query statistics were not good enough according to the developed index. For the rest of 241 companies, I automatically collected their GT data starting from the date of the company foundation and ending on August 31, 2019. For companies that I could not identify the exact date of foundation, I took January 1 of the corresponding year as the day of their foundation. Google Trends provides normalized values of search queries or, in other words, in each period of interest, there will be a point equal to 100 and other points related to it. In addition, to reach the weekly dimension of time series, only 200 points could be provided by GT, what turns to the need to divide companies' lifecycles into 200-week periods (approx. 3.8 years). Taking into account the normalization of data built in the GT, after collection of the framed time series, I need to "sew" them together with the simultaneous repeating of normalization in order to have only one global maximum of 100 points.

After that, to eliminate the level of the fast noise, I also applied double exponential filtering (Huang et al., 2012; LaViola, 2003) for the valuation data with a small filtering coefficient: $\alpha = 0.99$. The coefficient α lies between zero and one, and values closer to the upper bound represent the lower level of filtration (Huang et al., 2012). During the last preprocessing step, in order to reach a larger number of analysis points, linear interpolation to the valuation data was applied, thus, making the number of examined points in GT and valuation data equal.

Finally, I obtained 241 research cases (Fig. 3.1; Appendix A) whose data were good enough to perform further correlation analysis. Applying the method specified for the hypothesis testing in this research (Cohen, 2013; Fieller et al., 1957; May and Looney, 2020), it was identified that to reach the 90% level of test power with the 0.05 significance and under the one-sided test (i.e., correlation is higher than the selected threshold) it is needed to have at least 17 subjects in the sample (more precisely – 16.41). Therefore, it can be concluded that the sample is sufficient for testing the taken hypotheses. I have also analyzed the distribution of companies among industrial fields. According to the CB Insights principle of aggregation, there are three sequential levels of field identification from wider to more narrow[18]: industrial sector, industry, and sub-industry. The descriptive statistics of the sample are presented (Table 3.1). A more holistic representation of the distribution of the sample, according to the three selected dimensions, is offered in Table 3.2.

[18] Industry Analytics - CB Insights [WWW Document], n.d. URL https://www.cbinsights.com/industries (accessed 3.2.20).

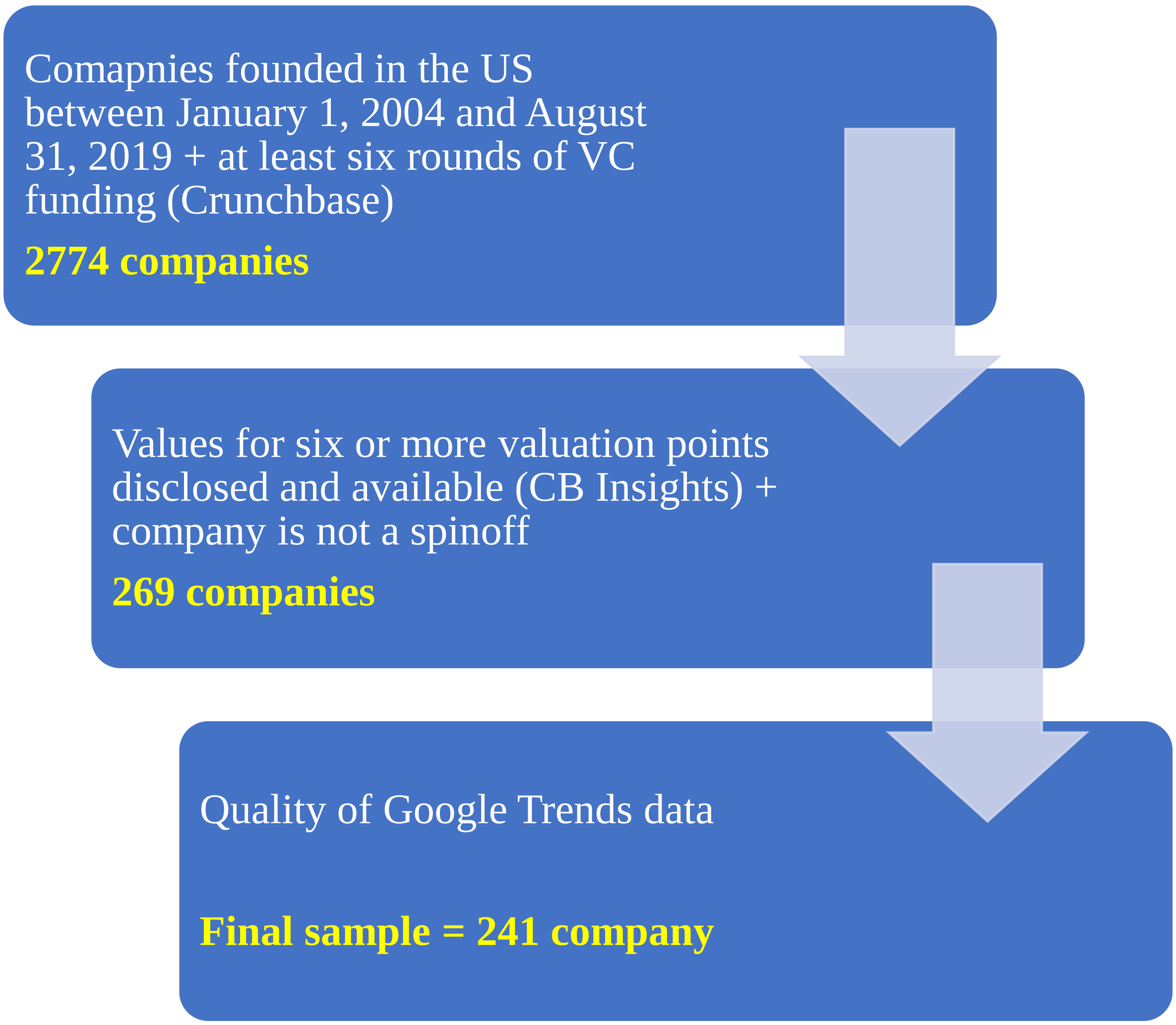


Figure 3.1. The sample selection and filtration process.

Name of the feature	Number	in % of the sample
Total	241	100%
Success feature		
Unicorns / Non-unicorns	106 / 135	44 / 56
Customer type feature		
b2b / b2c	150 / 91	62 / 38
Product type feature		
Digital platform / Traditional product	52 / 189	22 / 78
Top industrial sectors (1st level)	198	82%
Internet	141	59%
Mobile & Telecommunications	33	14%
Healthcare	24	10%
Top industries (2nd level)	168	70%
Internet Software & Services	110	46%
eCommerce	31	13%
Mobile Software & Services	27	11%
Top sub-industries (3rd level)	30	12%
Business Intelligence, Analytics & Performance Mgmt	15	6%
Advertising, Sales & Marketing	15	6%
MCAP-GR, $M/year		
Mean	373.60	-
25th pct	33.94	-
Median	92.11	-
75th pct	200.37	-

Table 3.1. Descriptive statistics of the sample.

	"Unicorn"		"Non-unicorn"	
	Digital platform	Traditional product	Digital platform	Traditional product
b2c	28	23	14	26
b2b	6	49	5	90

Table 3.2. Holistic representation of the distribution of the sample.

3.2.4.2 Correlation analysis

3.2.4.3 Methodology

As a result of the data collection process, for each TBNV, two series of data were obtained: search query statistics on the related term (i.e., company name) from Google Trends and the company's valuation data with interpolated points between the initial ones. Next, to decrease the level of noise while maintaining the weekly resolution of data, I applied the additional filtering to the GT time series with a stronger filtering coefficient α. After a number of tests, I empirically tuned the value of α in such a way as to provide the most similar curve to the original one while keeping a low amount of noise at the same time. To decrease the possible random errors, I have also applied low double exponential filtering to valuation data points (Fig. 3.2). Thus, for GT data, the filtering coefficient α equals 0.2, while for the valuation time series, it equals 0.9. After this step, I normalized the obtained search query and valuation data for the interval between 0 and 1 in order to receive the same scale of data series.

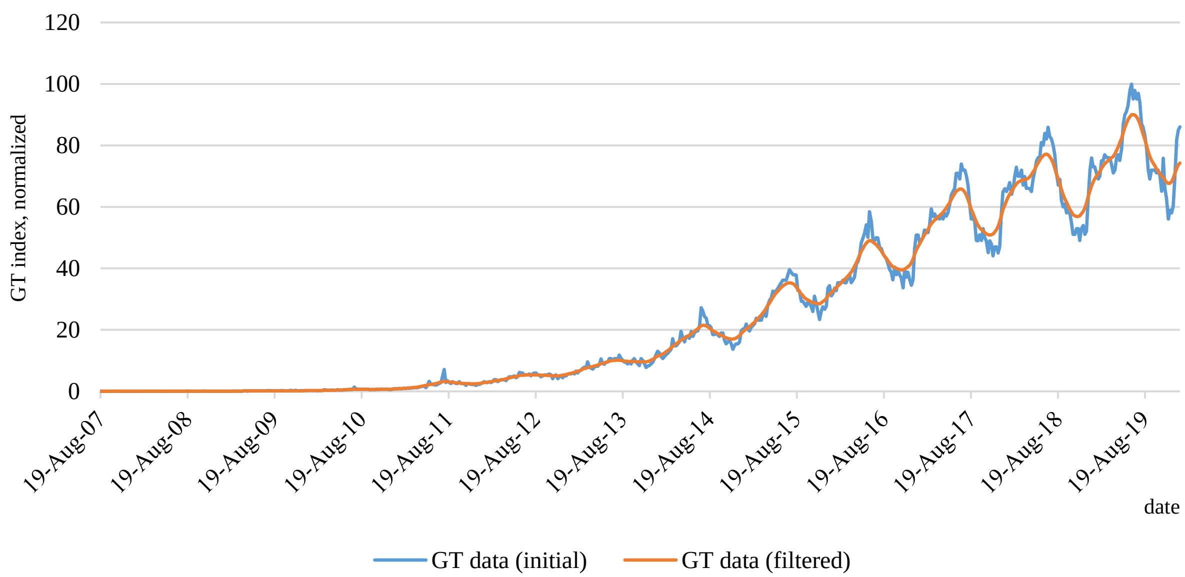


Figure 3.2. Example of the initial and filtered GT data (*α=0.2*), case of *Airbnb.*

After finishing the data preparation step, the rank correlation coefficient *Kendall's tau* (Kendall, 1948) was employed. Three correlation coefficients are typically employed for analyzing the relationship between two variables (Allen, 2017): Pearson's rho, Spearman's rho, and Kendall's tau. Due to its mathematical nature, Pearson's product-moment correlation coefficient rho is a measure of a linear relationship and, thus, can be correctly used for describing linear relationships between examined variables (Puka, 2011). In addition, Pearson's rho assumes that the distribution of the examined variables is normal (Allen, 2017). Since the relation between TBNVs GT data and valuation has not been studied previously, we cannot reliably assume the linear link between them or the normal distribution of each. Pearson's rho is also sensitive to outliers, the presence of which cannot be avoided in the current setup. Therefore, I concluded that the non-parametric tests would be more accurate and reliable for the current research than the parametric one (i.e., Pearson's rho).

Considering the non-parametric tests, there are two measures commonly applied (Allen, 2017): Spearman's rho and Kendall's tau. As it is mentioned in Puka (2011, p. 714), "Kendall's Tau is equivalent to Spearman's Rho, with regard to the underlying assumptions" and "(i)n most of the situations, the interpretations of Kendall's Tau and Spearman's rank correlation coefficient are very similar and thus invariably lead to the same inferences." However, despite these similarities, Puka also recommends using a lower value of these two measures to stay on the "safe" side (Puka, 2011, p. 714). In another study, the values of all mentioned correlation measures taken for the same set of data were compared and resulted in the fact that Kendall's tau yields lower values than Spearman's rho (Gilpin, 1993, p. 89). Taking this into account and adding that Kendall's tau is also evidenced to be beneficially used for studying various GT data related phenomena (Le Nghiem et al., 2016; Preis et al., 2013), I decided to take Kendall's tau as a correlation measure coefficient for the current research.

As for the threshold level, I selected 0.5, thus, treating the equal to and higher values as evidence of a strong link. According to the established practice (Akoglu, 2018; Moore et al., 2012), a strong level of Pearson's rho is typically considered from 0.7, while Kendall's tau is connected with Pearson's rho by the formula (Kendall, 1948): $R = sin(0.5^* \ Pi \ ^* \ tau)$. This formula was mathematically derived in the original work of Kendall (1948) and further discussed in detail in (Waerden, 1969) as well as the other relations between the mentioned correlation measures. This relationship is derived from the formulation of probability density integrals for the relation between Spearman's rho and Pearson's rho (Waerden, 1969, pp. 329–330, Eqs. 24-28), the mathematical expectation of studied pairs of values (Waerden, 1969, p. 330, Eq. 29), and further generalizations for the case of the relationship between Kendall's tau and Pearson's rho (Waerden, 1969, pp. 333–334, Eqs. 1-6). Thereby, the Equation 6 (Waerden, 1969, p. 334) can be applied to calculate the estimation of Pearson's rho. For the purpose of the current research, this formula was applied backward to calculate the threshold of Kendall's tau, which corresponds to the "strong" level of Pearson's rho = 0.7 accepted by academia. The calculations resulted in a "strong" Kendall's tau = 0.5, which is also supported by the outcomes of Gilpin (1993). Analysis of the level of Kendall's tau provides the ability to make conclusions whether the character of search query data corresponds to dynamics demonstrated by a valuation history and in which manner.

Due to the reason that GT data of each case cover the whole lifecycle from the company foundation until the chosen date and series of investments are happening during the particular timespan of this lifecycle, the amount of search query data points is significantly larger than the number of valuation points. To deal with this fact, the approach from Kirk et al. (2013) is adopted – I presume that in some cases, a particular time lag between GT data and valuation dynamics may be present. Therefore, to check if the analyzed curves have similar dynamics but spaced in time, I employed the adopted cross-correlation method (ACC). The obtained shift in weeks, as well as the correlation coefficients, were recorded for each case, which did not provide a high correlation level otherwise. However, the method demonstrated high sensitivity that resulted in "false positives." For instance, in some cases, the increase in correlation after applying the ACC method was minimal with, at the same time, relatively significant shifts between the series of data. In order to avoid such results, I measured the percentage of increase in Kendall's tau after applying the adopted cross-correlation and, conservatively, treated enhancements lower than 50% as not significant. Next, the obtained results are presented.

3.2.4.4 Results of the correlational analysis

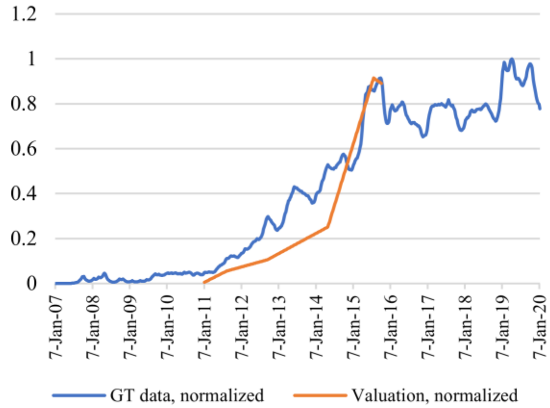
For each TBNV in the sample, I obtained a number of outputs, which included the highest Kendall's tau correlation level and the size of a weekly shift for the TBNVs when ACC was applied. The minus sign of the shift means that valuation data should be moved "back in time," i.e., growth in the GT data started earlier than in valuation. The opposite is correct as well: the plus sign tells that the valuation of the company started to grow before the GT search queries and should be shifted "forward in time" in order to reflect the highest correlation level. Considering the results, all sample TBNVs can be separated into three groups (Table 3.3):

1. G1 *"Strong link, without a shift"*: cases that showed a strong correlation and no shift was applied;
2. G2 *"Strong link, with a shift"*: cases that showed a strong correlation but a time shift exists;
3. G3 *"Weak link"*: cases that did not show a strong correlation in any option.

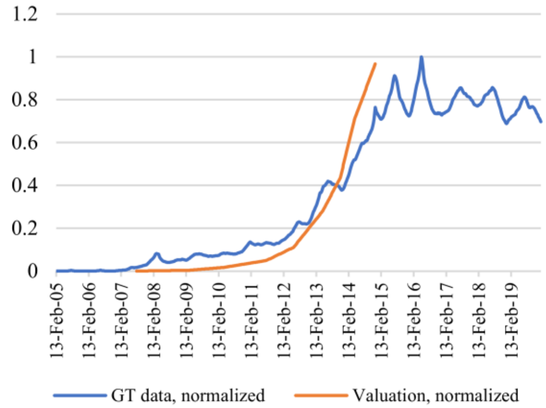
Examples of the plots with analyzed time series for all groups are presented (Fig. 3.3-3.11).

			Kendall's tau			
	Count	Percentage	Mean	25th pct	Median	75th pct
Total sample	241	100%	0.66	0.56	0.70	0.82
G1 “Strong link, without a shift”	161	66.8%	0.77	0.68	0.78	0.88
G2 “Strong link, with a shift”	39	16.2%	0.63	0.56	0.63	0.70
G3 “Weak link”	41	17.0%	0.25	0.15	0.38	0.45
Shift direction (from G2)						
Positive shift	7	17.95%	116.71	94	101	160
Negative shift	32	82.05%	-105.03	-131	-97	-39

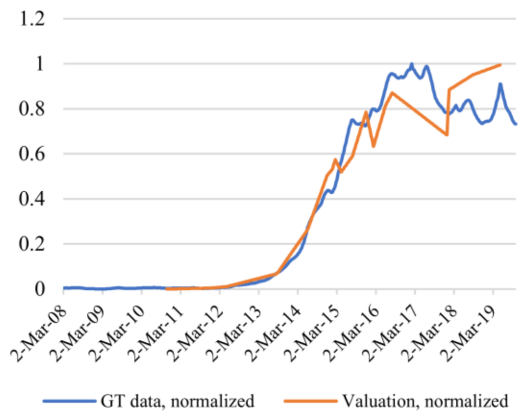
Table 3.3. Descriptive statistics of the obtained results.



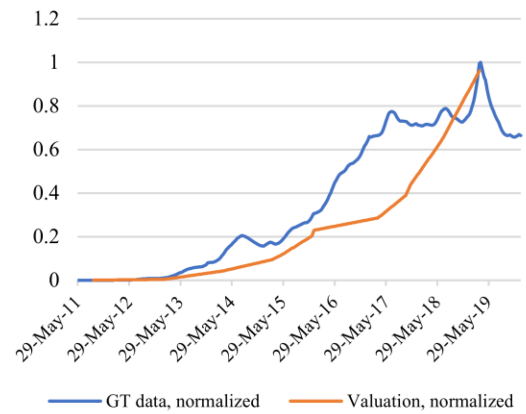
(a)



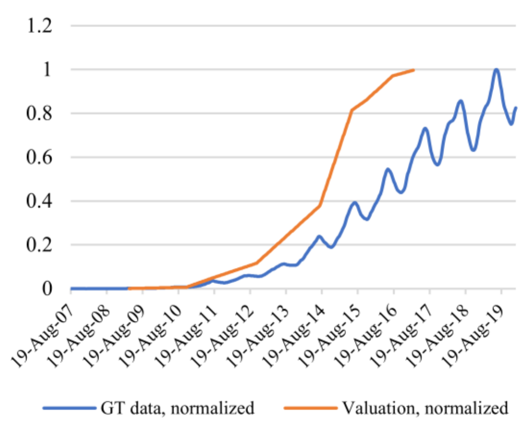
(b)



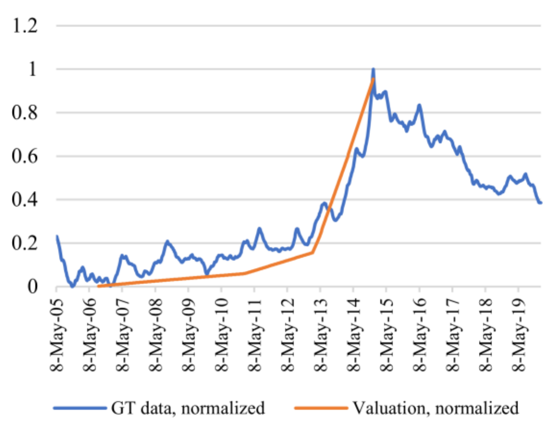
(c)



(d)

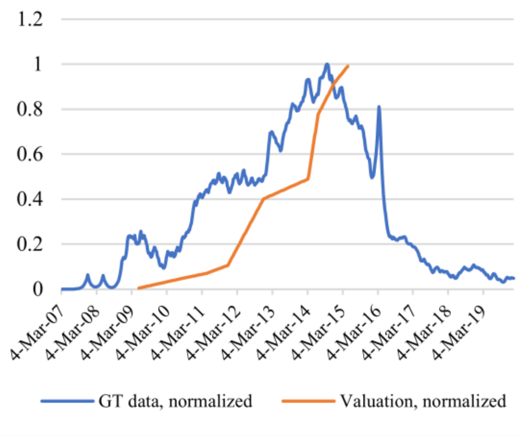


(e)

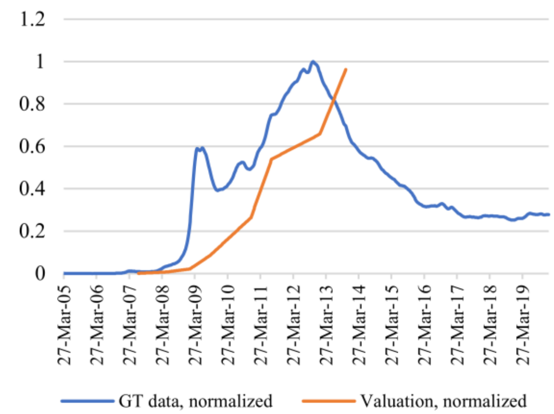


(f)

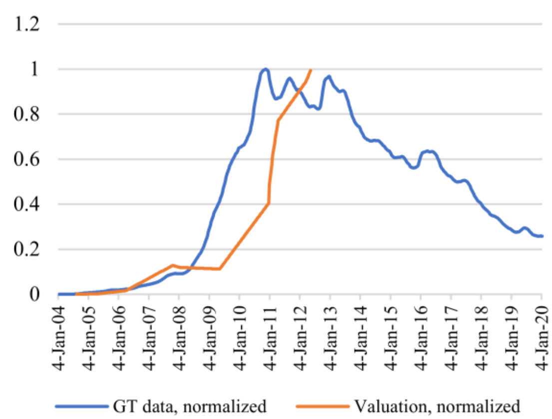
Figure 3.3. GT data and valuation curve of G1 “Strong link, without a shift”: Kabbage (a), Lending Club (b), Uber (c), Lyft (d), Airbnb (e), OnDeck (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.



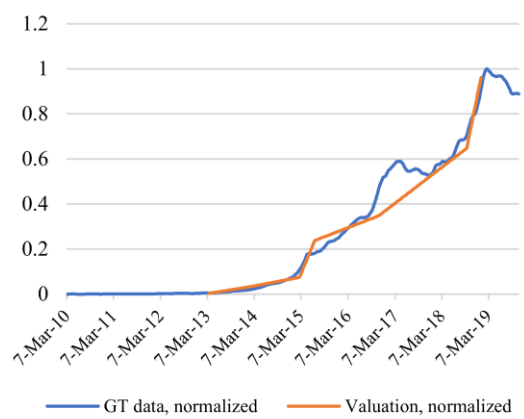
(a)



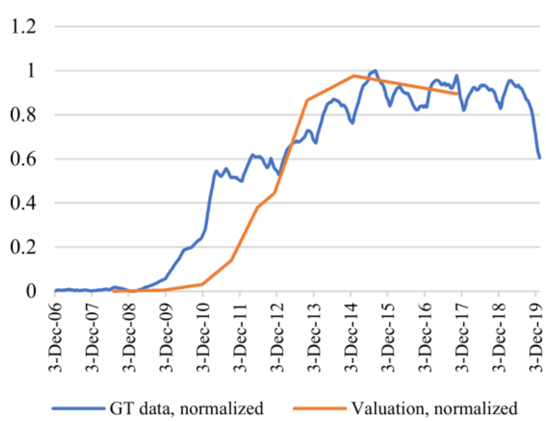
(b)



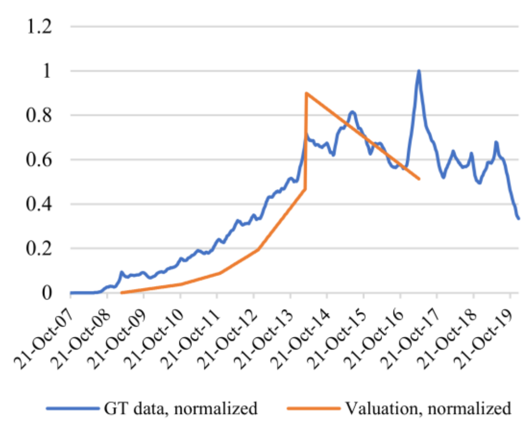
(c)



(d)

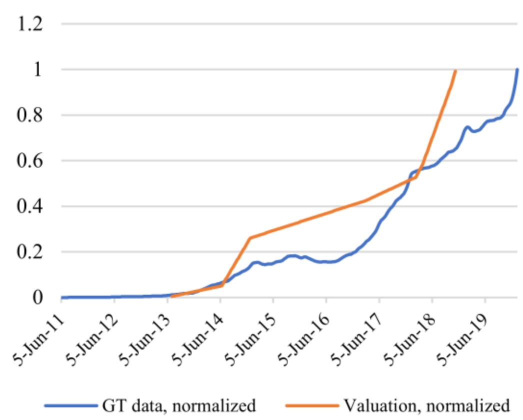


(e)

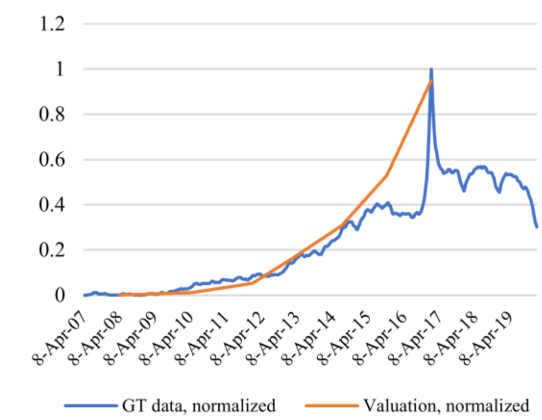


(f)

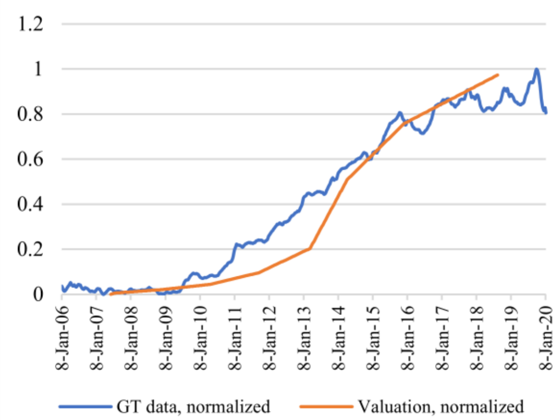
Figure 3.4. GT data and valuation curve of G1 “Strong link, without a shift”: Next Step Living (a), Twitter (b), Facebook (c), Postmates (d), MongoDB (e), Cloudera (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.



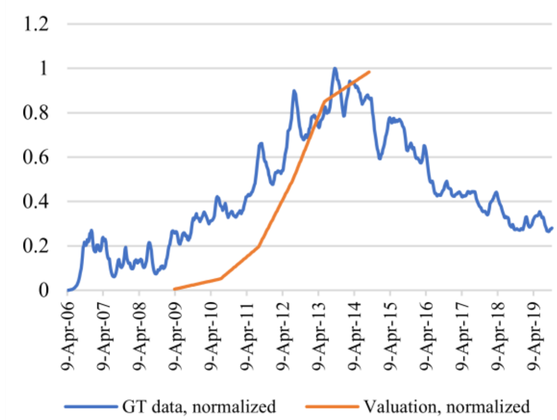
(a)



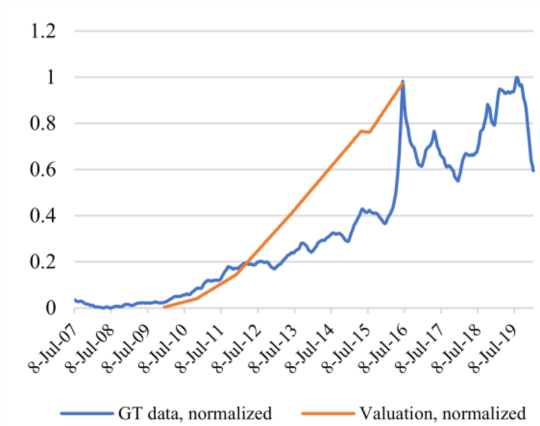
(b)



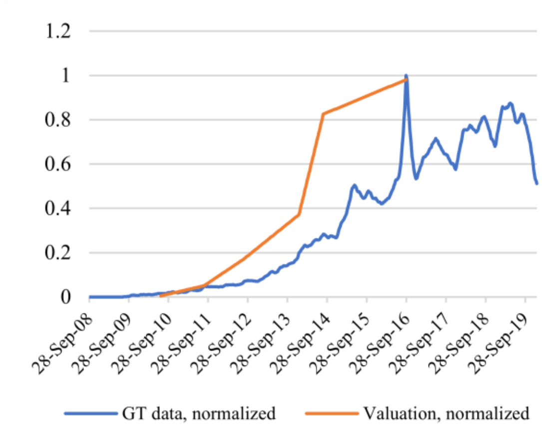
(c)



(d)

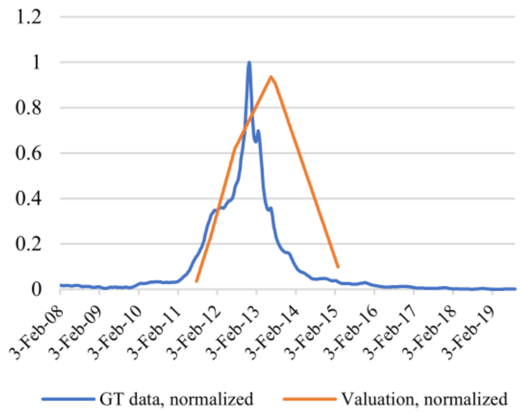


(e)

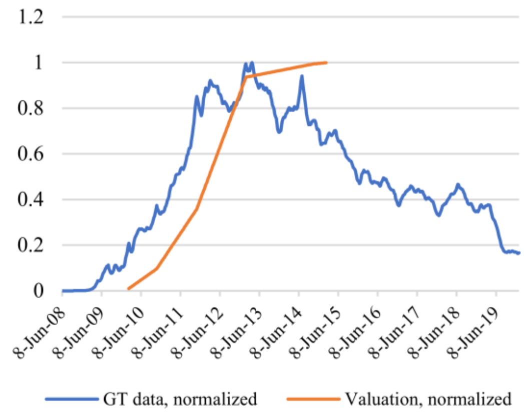


(f)

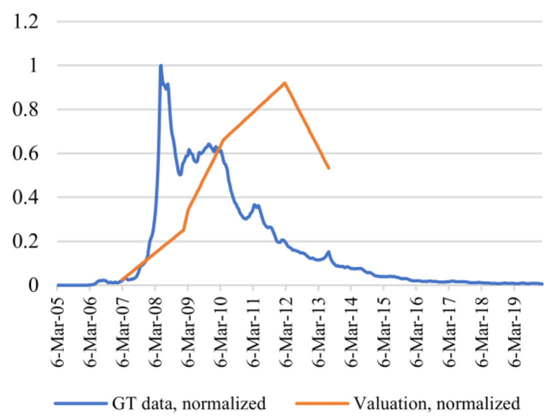
Figure 3.5. GT data and valuation curve of G1 “Strong link, without a shift”: Instacart (a), AppDynamics (b), One Medical (c), GoodData (d), Twilio (e), Nutanix (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.



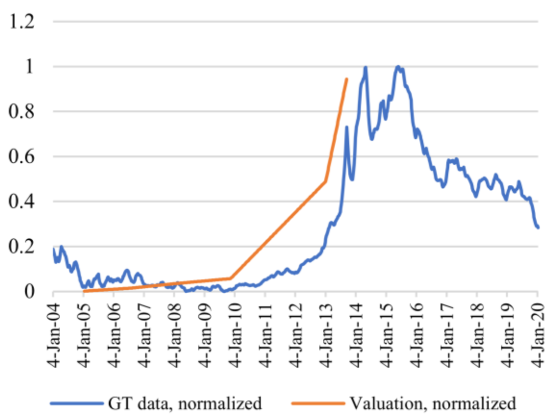
(a)



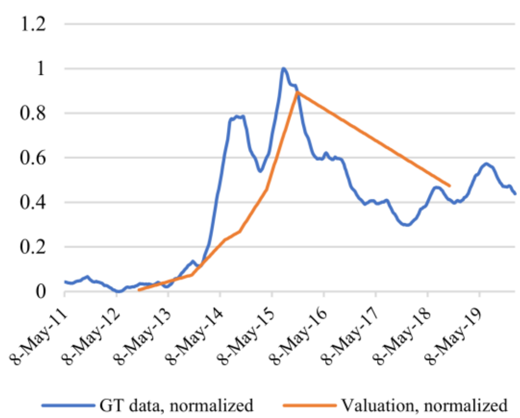
(b)



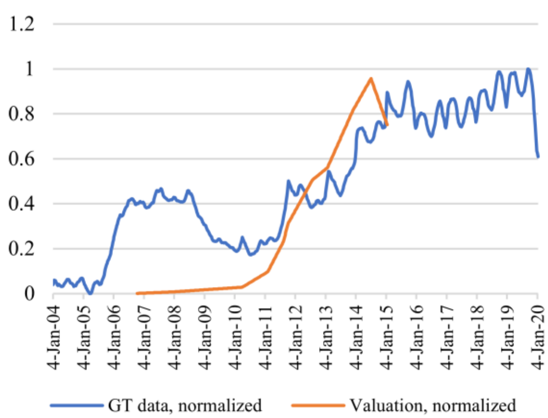
(c)



(d)



(e)



(f)

Figure 3.6. GT data and valuation curve of G2 “Strong link, with a shift”: Fab (a), Urban Airship (b), Xobni (c), FireEye (d), Handy (e), Box (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.

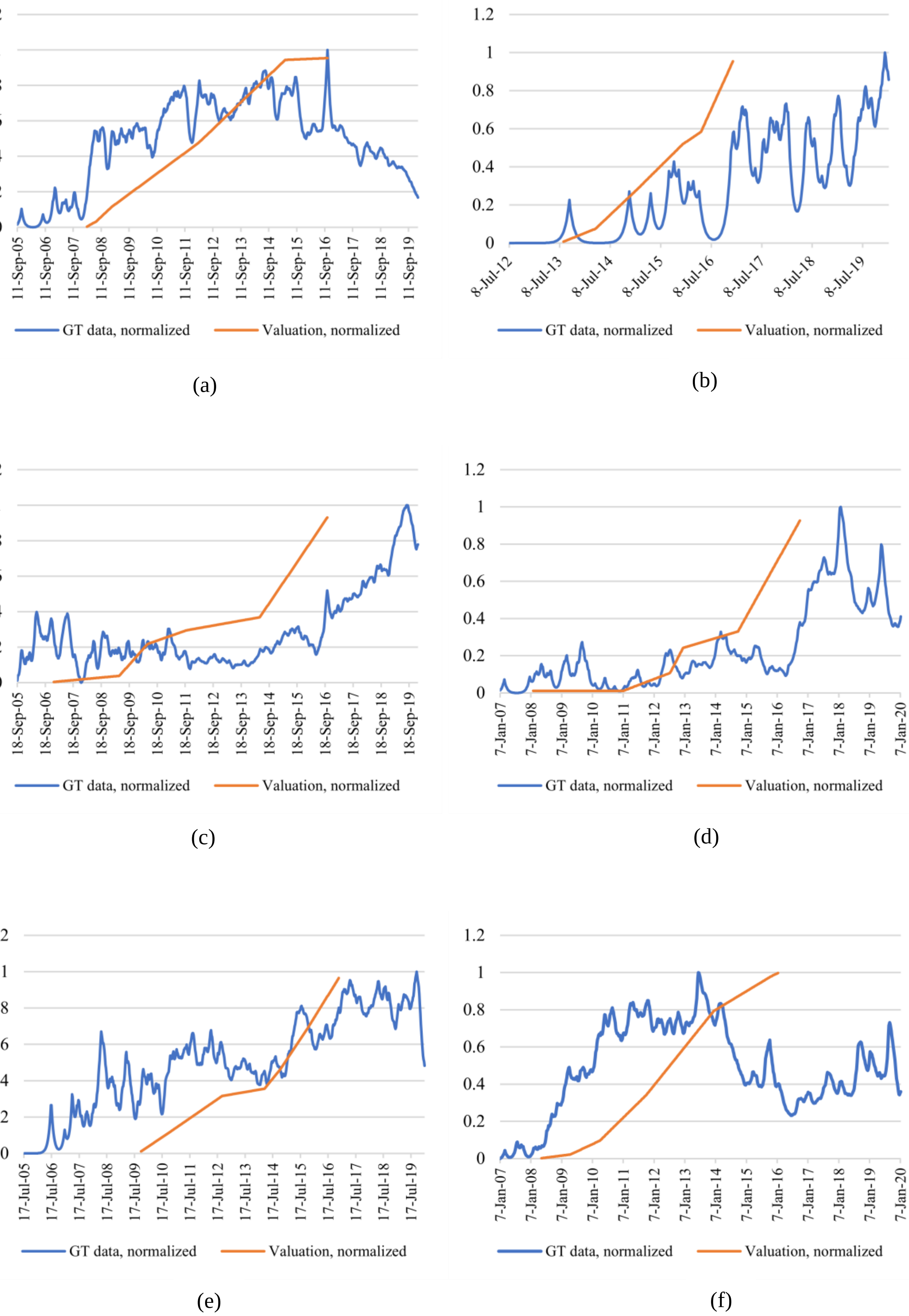


Figure 3.7. GT data and valuation curve of G2 “Strong link, with a shift”: Appirio (a), BlueVine (b), iRhythm Technologies (c), Obalon Therapeutics (d), SnapLogic (e), Knewton (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.

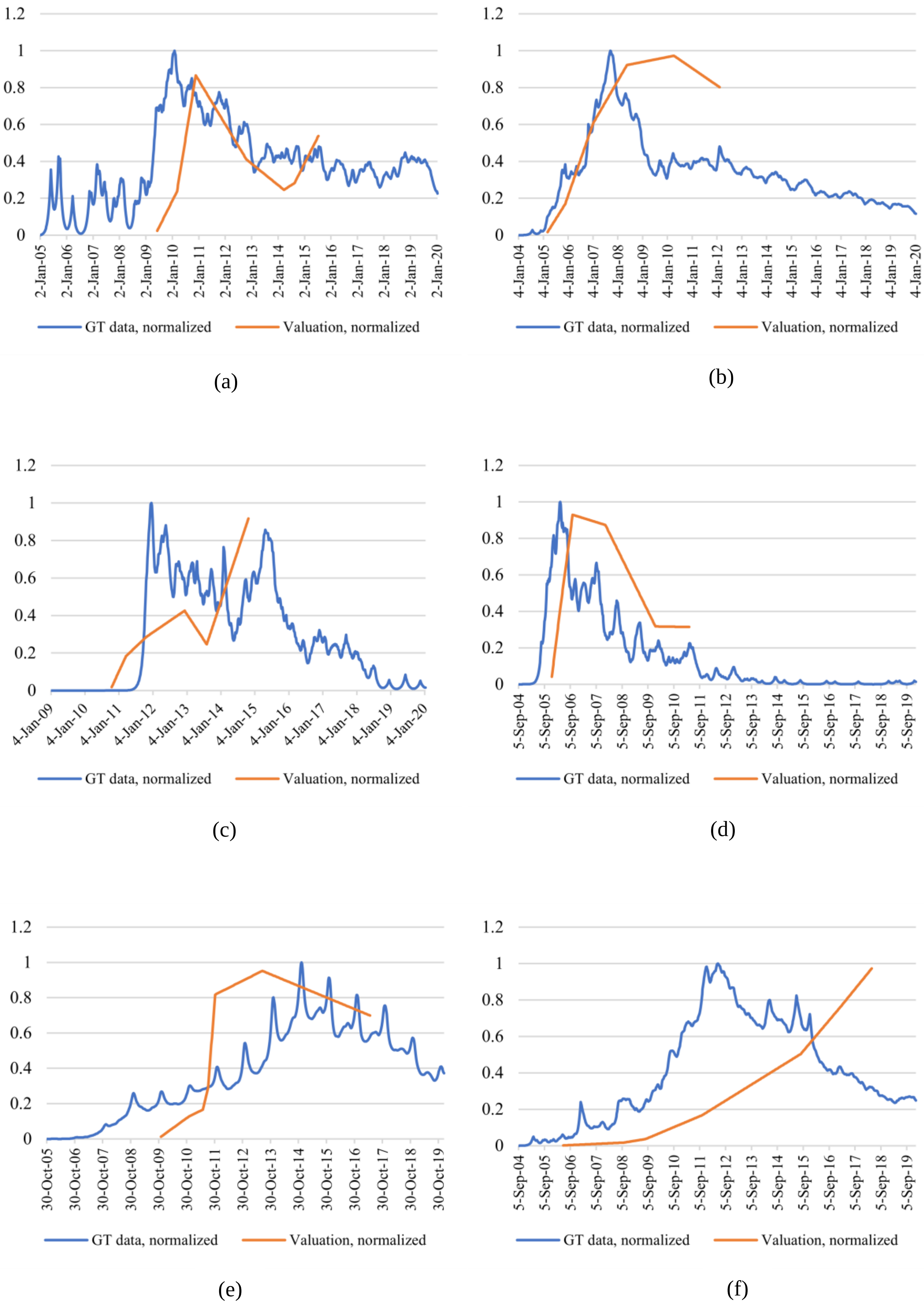
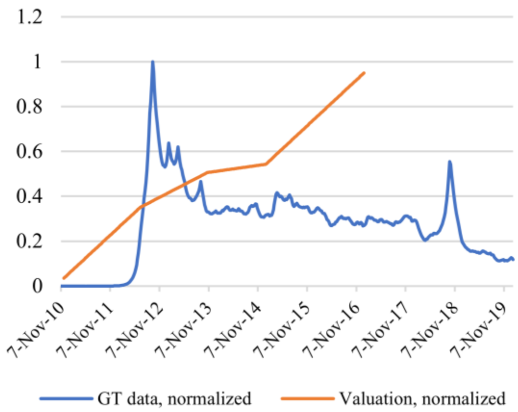
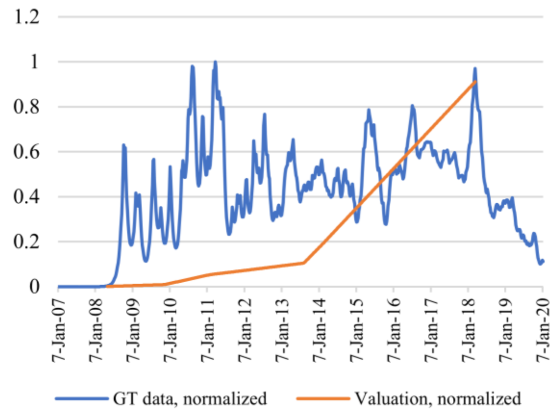


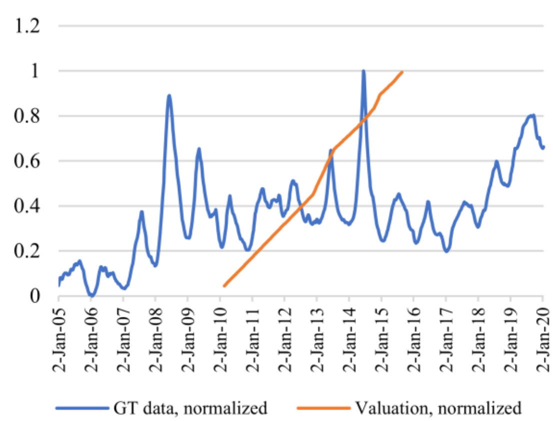
Figure 3.8. GT data and valuation curve of G2 “Strong link, with a shift”: Scale Computing (a), Brightcove (b), Linkable Networks (c), Jingle Networks (d), RetailMeNot (e), SoundHound Inc. (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.



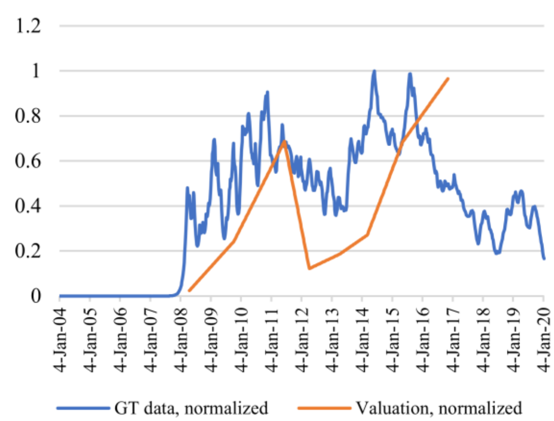
(a)



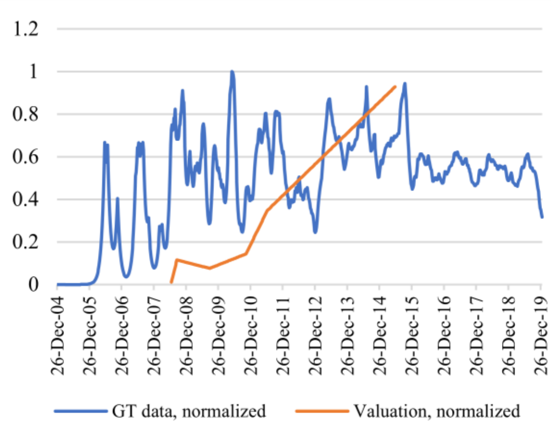
(b)



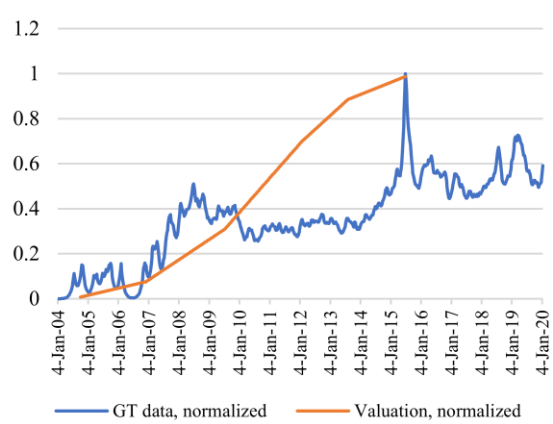
(c)



(d)



(e)



(f)

Figure 3.9. GT data and valuation curve of G3 “Weak link”: Rethink Robotics (a), NxThera (b), Zero Motorcycles Inc. (c), Illumitex (d), Biodesix (e), ConforMIS (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.

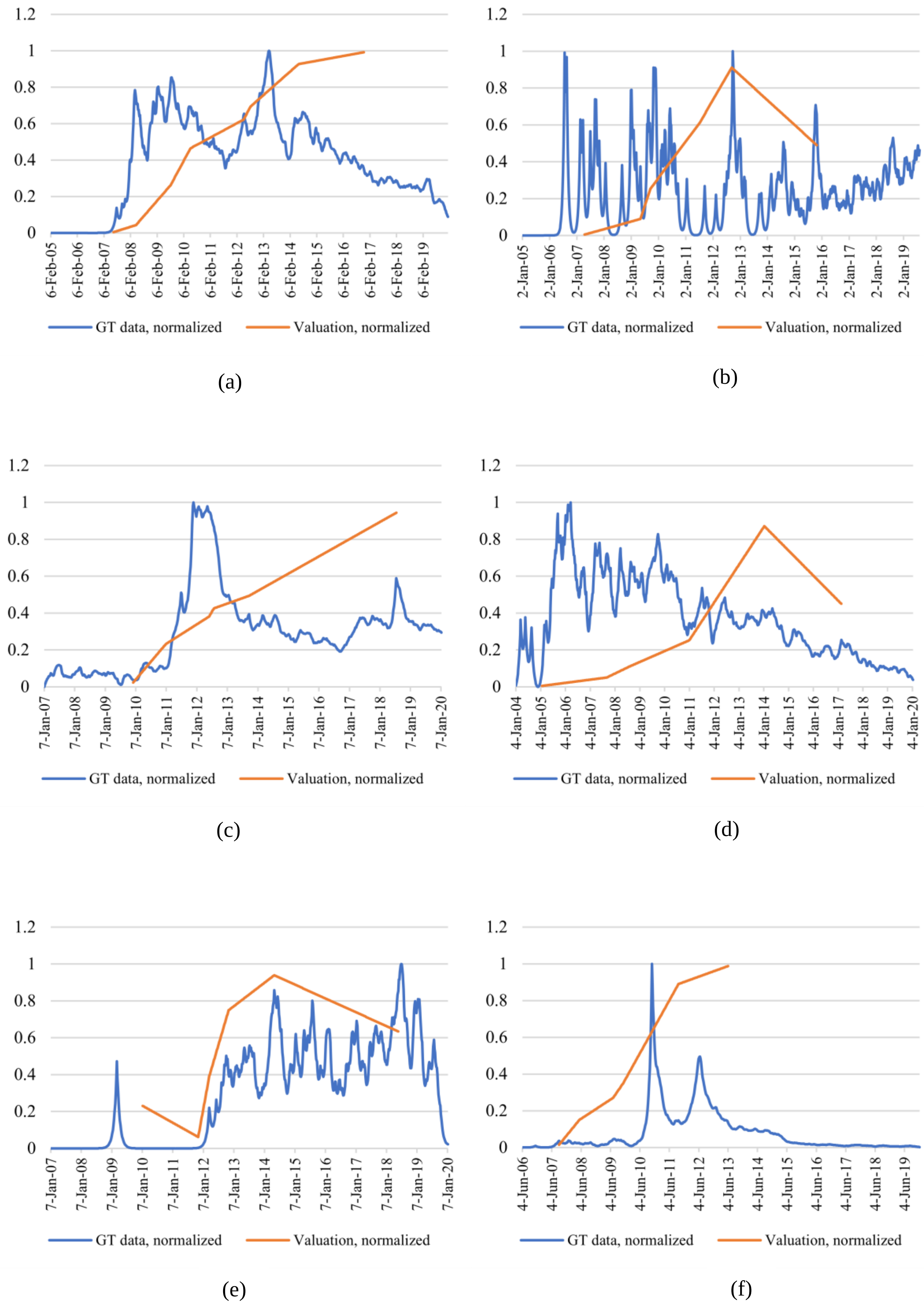
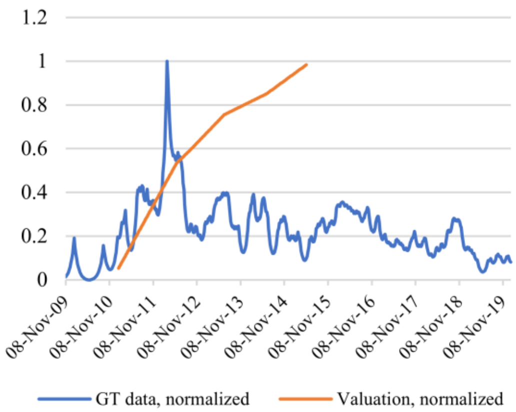
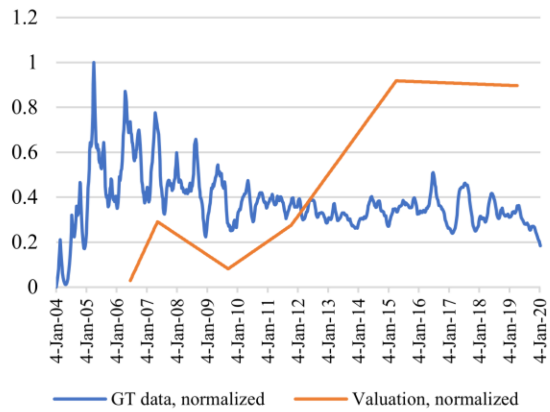


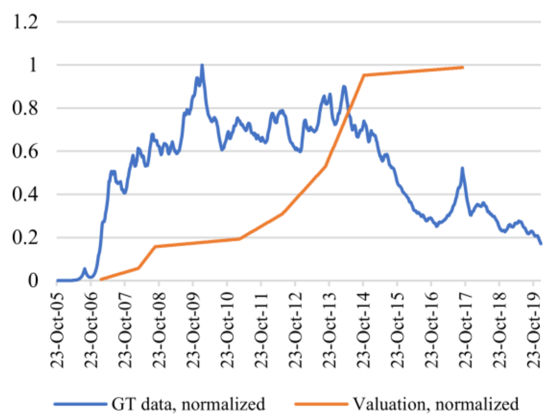
Figure 3.10. GT data and valuation curve of G3 "Weak link": RichRelevance (a), Adesto Technologies (b), LevelUp (c), Turn Inc (d), Counsyl (e), Blekko (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.



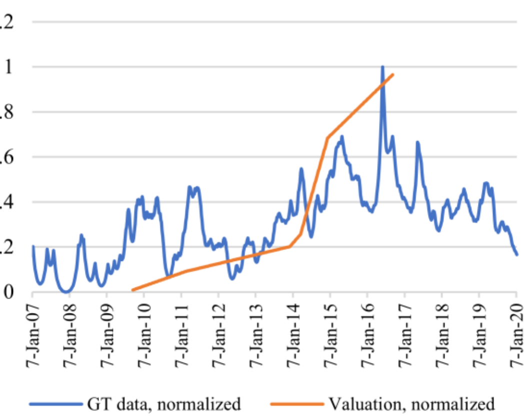
(a)



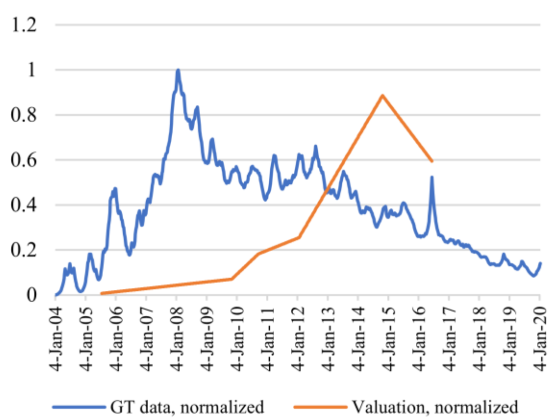
(b)



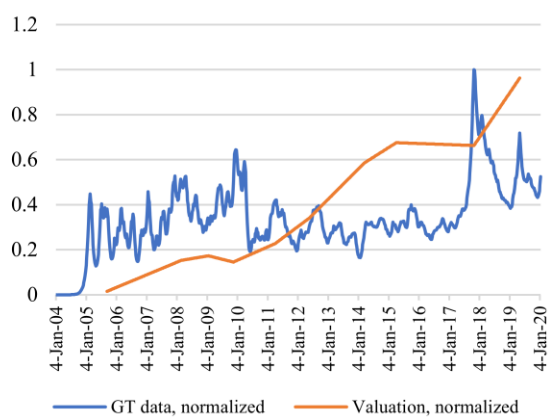
(c)



(d)



(e)



(f)

Figure 3.11. GT data and valuation curve of G3 “Weak link”: Kony (a), RedSeal (b), Gigya (c), Phononic (d), Joyent (e), Aquantia (f). On the X-axis are dates, on the Y-axis is the GT index, normalized.

Descriptive statistics of the obtained results demonstrate the left skew of Kendall's tau values for G1, "Strong link, without a shift," that signals a stronger relationship than the correlation threshold chosen for the current research (Table 3.3). The cases with weak correlation have average Kendall's tau lower than 0.5 but still close to it, pointing at the existence of the particular level of the link between GT and valuation data but below the selected threshold. Among all cases with a time lag, absolute lag values are relatively significant: averages of both positive and negative variants are close to 100, what is approx. two years (Table 3.3, Positive shift, Negative shift). Considering the summarized results across the whole sample, it can be seen that the mean value of Kendall's tau equals 0.66, with 83% of the outcomes laying above the chosen threshold level with all p-values lower than 0.01 (Table 3.3, Total sample). The median value is also slightly higher than the mean that demonstrating the existence of the left skew among the obtained correlations. Percentiles also back this fact: the quarter of correlation coefficients in all cases from G1 lies above 0.88, which is significantly higher than the chosen threshold (Table 3.3, G1). Moreover, even the companies from G3, on average, obtain values quite close to 0.5, what signals about the existence of the particular relationship in weak correlation cases. Altogether, the obtained results present clear evidence that TBNVs' dynamics are positively and significantly correlated with the related to the web search query statistics and, thus, support *H1*.

			Count	Percentage			Kendall's tau			
				Sample	Dimension	Group	Mean	25th pct	Median	75th pct
Success dimension	" Unicorn"	G1	92/106	38%	86%	57%	0.80	0.71	0.82	0.90
		G2	8/106	3%	8%	21%	0.63	0.58	0.59	0.66
		Lag	-	-	-	-	-35	-110	-77	4
		G3	6/106	2%	6%	15%	0.26	0.17	0.27	0.39
	" Non-unicorn"	G1	69/135	29%	51%	43%	0.72	0.64	0.73	0.80
		G2	31/135	13%	23%	79%	0.64	0.56	0.62	0.70
		Lag	-	-	-	-	-73	-121	-88	-46
		G3	35/135	15%	26%	85%	0.25	0.15	0.38	0.45
Customer type dimension	b2c	G1	72/91	30%	79%	45%	0.78	0.68	0.80	0.89
		G2	7/91	3%	8%	18%	0.70	0.61	0.66	0.74
		Lag	-	-	-	-	-37	-99	-54	34
		G3	12/91	5%	13%	29%	0.28	0.25	0.43	0.46
	b2b	G1	89/150	37%	59%	55%	0.75	0.68	0.75	0.85
		G2	32/150	13%	21%	82%	0.62	0.55	0.59	0.70
		Lag	-	-	-	-	-71	-117	-91	-49
		G3	29/150	12%	19%	71%	0.24	0.15	0.36	0.42
Product type dimension	Digital platform	G1	40/53	17%	76%	25%	0.83	0.77	0.87	0.90
		G2	6/53	2%	11%	15%	0.70	0.60	0.70	0.77
		Lag	-	-	-	-	-42	-100	-62	-37
		G3	7/53	3%	13%	17%	0.30	0.21	0.40	0.42
	Traditional product	G1	121/188	50%	64%	75%	0.75	0.67	0.76	0.82
		G2	33/188	14%	18%	85%	0.62	0.56	0.59	0.69
		Lag	-	-	-	-	-70	-130	-90	-42
		G3	34/188	14%	18%	83%	0.24	0.15	0.37	0.45

Table 3.4. Correlation analysis across three dimensions.

max MCAP-GR, $M/year

	Mean	deviation from the sample, %	25th pct	deviation from the sample, %	Median	deviation from the sample, %	75th pct	deviation from the sample, %
Sample	373.60		33.94		92.11		200.37	
G1	512.50	37%	66.86	97%	133.75	45%	296.27	48%
G2	71.59	-81%	18.51	-45%	37.44	-59%	117.70	-41%
G3	115.41	-69%	15.38	-55%	30.55	-67%	62.72	-69%

Table 3.5. Distribution of the *Market Capitalization Growth Rate* measure among the groups.

“Success” dimension

To describe the obtained results from the “success” dimension, I analyze how correlations between TBNVs’ GT data and their valuations are distributed among the chosen groups. It can be observed that 86% of all “unicorns” (92 cases) demonstrate a strong correlation without any time gap, as well as 51% (69 cases) of all “non-unicorns” (Table 3.4, Success dimension, G1, and G2). The fraction of all “unicorns,” which show a high correlation is significantly higher than the fraction of all “non-unicorns.”

At the same time, only 8% of all “unicorns” showed the presence of a time lag (G2), while for “non-unicorns,” percentage is almost three times higher, 23% (Table 3.4, Success dimension, G2). The absolute average value of the time lag for “non-unicorns” is more than two times larger than for “unicorns” in the sample. The difference significantly decreases when median values are compared (87 vs. 77 weeks). A smaller absolute time gap for “unicorns” belonging to G2 suggests better reflection of “unicorns'” growth dynamics in related to them web search queries (Table 3.4, Success dimension, Lag). For G3, the distribution between “unicorns” and “non-unicorns” is more or less the same as for G2: only a few “unicorns” (6%) do not show a strong correlation, while one quarter (26%) of all “non-unicorns” is in this group (Table 3.4, Success dimension, G3).

It can be observed that in G1, the “unicorn” companies obtain on average higher mean and median correlation values than “non-unicorns,” and altogether are closer to one (represented by the 25th and 75th percentiles). At the same time, “unicorns” from G3 obtain a lower median, while “non-unicorns” have more variations with some values falling below zero. This evidence suggests that the growth dynamics of TBNVs with the “unicorn” status are better reflected by their GT data than the growth dynamics of TBNVs with the “non-unicorn” status, proving hypothesis H2.

Next, I analyze companies' *Market Capitalization Growth Rate* (MCAP-GR) – another metric related to a company valuation. I took the maximum valuation of a TBNV, the date of this event, and then calculated how long it took a company to reach it, i.e., the maximum MCAP-GR. Results demonstrate that, on average, companies from Group 1 have a 37% higher MCAP-GR than the sample averages (Table 3.5, G1). In opposite, cases with weak correlations (G3) gain significantly lower average and median values compared to the whole sample: -69% for the average and -67% for the median (Table 3.5, G3). TBNVs from Group 2 obtained significantly lower than average measures of MCAP-GR (Table 3.5, G2) as well. However, it is expected - most companies from G2 have a negative time lag, so their valuation started to grow after some period since their foundation what decreased their MCAP-GR. The overall conclusion is that the faster the company's value is growing, the better its growth dynamics is reflected by GT search query data. It supports hypothesis H2a.

Altogether, considering the "success" dimension, the evidence shows that, as predicted by H2 and H2a, the more successful company is, the better it is reflected by related to it GT data.

Customer type dimension

The distribution of results demonstrates that TBNVs focused on the b2c segment of the market have higher mean and median correlation values than b2b-oriented companies across all three groups. Percentiles (25th and 75th) also suggest that the distribution of correlation values for b2c TBNVs, across all three groups, is more skewed to the left than in the case of b2b companies (Table 3.4, Customer type dimension, G1, G2, and G3).

Further, almost 80% of all b2c companies from the sample demonstrate a high correlation without time lag, compared to only 60% of all b2b companies from the sample. At the same time, less frequently, b2c companies show a weak correlation compared with b2b companies (13% vs. 19%). In the case when high correlations with and without time lag are considered together, again, b2c companies dominate (87% vs. 80%).

When it comes to the time lag, only 8% of b2c companies achieve a high correlation with it, opposed to 21% of b2b companies. The absolute average value of the lag for b2c companies is almost twice smaller than for b2b-oriented companies in the sample. This result suggests the better connection between b2c companies' growth dynamics and related to them web search queries (Table 3.4, Customer type dimension, Lag).

These results present clear evidence that b2c oriented companies have a better correlation between their growth dynamics and related to them web search queries. Hence, *H3* is supported. However, the difference between two "poles" in this dimension (b2c vs. b2b) is smaller than what was observed for the "success" dimension. Thus, it can be concluded that the customer type dimension is less impactful on the link between TBNVs' growth dynamics and related to them web search query data than the "success" dimension.

Product type dimension

Three-quarters of all TBNVs who create digital platforms in the sample (76%) demonstrate a high correlation between their GT and valuation data without any time lag, opposite to 64% of companies, which produce traditional products (Table 3.4, Product type dimension, G1). When high correlations with and without a time lag are considered together, the difference still exists but is significantly smaller (87% platforms vs. 82% traditional products). The higher percent of companies that market traditional products show a weak correlation between their GT and valuation data compared to platform-oriented companies (18% vs. 13%).

The descriptive statistics of G1 and G2 also indicate the higher correlation level in the case of digital platform products. On average, percentiles (25th and 75th), mean and median values of correlation coefficients for digital platforms are 10% higher than for the traditional products (Table 3.4, Product type dimension, G1, and G2). With the exception of the 75th percentile, the situation is the same for G3. That suggests that the distribution of correlation values across all three groups is more skewed to the left in the case of platforms than non-platforms (Table 3.4, Product type dimension, G1, G2, and G3).

A time lag between GT and valuation data leading to high correlation is identified in 11% of platform-oriented companies and 18% of those developing traditional products. That is the smallest difference across three dimensions. The absolute average value of the lag is bigger for non-platforms, but this difference is also smaller than in the other two dimensions.

Summarizing, the results of product type dimension analysis support H4 and demonstrate that growth dynamics of companies that develop their products in the form of a digital platform are better (and stronger) correlated with their web search traffic represented by GT data than companies, which create traditional products.

Industrial designator	Count of high correlation cases	Count of all cases	Ratio
Industrial sectors, the 1st level			
Computer Hardware & Services	8	11	73%
Healthcare	13	24	54%
Internet	120	141	85%
Mobile & Telecommunications	26	33	79%
Industries, the 2nd level			
eCommerce	29	31	94%
Internet Software & Services	91	110	83%
Medical Devices & Equipment	5	9	56%
Mobile Software & Services	20	27	74%
Sub-industries, the 3rd level			
Advertising, Sales & Marketing	12	15	80%
Business Intelligence, Analytics & Performance Mgmt	14	15	93%
Customer Relationship Management	6	8	75%
Marketplace	7	8	88%
Monitoring & Security	6	8	75%
Payments	4	8	50%
Social	5	6	83%

Table 3.6. Results of industrial area dimension

3.2.4.5 Industrial area perspective on the results

The last feature of the sample companies that was employed to examine the obtained results in current research is the company industrial area. Since CB Insights classification of new ventures' industrial sectors has three hierarchical levels (*sector – industry – sub-industry*[18]), the distribution of results was also analyzed separately. As a result, it can be noticed that for the *sectors* level, four areas present significantly higher results (Table 3.6). According to the mentioned previous studies, consumers tend to use the same channel for executing the process of searching for information about the product and subsequent purchasing it (Shim et al., 2001; To et al., 2007). Thus, since the top three sectors in various manners may become mediating channels for a data acquisition process, TBNVs related to them are expected to have a stronger link between web search statistics and sales dynamics that directly influence their valuation.

Considering the second level of classification (*industry*), the top three areas go in line with common sense: people actively use the internet for selecting, searching, and buying goods, and especially mobile or desktop software. The last level – *sub-industry* – may not be very representative due to the many possible options; however, according to the obtained results, areas that focus on the b2b segment of the market overall obtain more high-correlation "hits" (Table 3.6). This outcome is quite notable since, as was mentioned in the previous chapter, being a b2b does not lead to the highest levels of correlation between the company's GT data and its valuation points. It can be assumed that new business products heavily utilize internet technologies, and b2b customers tend to use web search more often to choose the appropriate solution.

3.2.4.6 Configurational analysis

Methodology

To better understand which TBNVs' features (i.e., b2c vs. b2b, "unicorn" vs. "non-unicorn," and digital platform vs. traditional products) lead to achieving high correlation between the TBNV's growth dynamics and related to it web search traffic, fuzzy-set Qualitative Comparative Analysis (fsQCA) is employed on the data sample. This is a set-theoretic, cross-case, and diversity-based research methodology that allows a holistic comparison of individual cases while identifying comprehensive configurations across the sample (Ragin, 2000).

By taking into account multiple conjunctural causations, causal equifinality, and causal asymmetry of different conditions (i.e., TBNVs' features) in relation to the outcome of interest (i.e., high or low positive correlation between the TBNV's growth dynamics and its web search traffic), fsQCA allows us to deal with the extant causal complexity (Fiss, 2007; Misangyi et al., 2016) in this part of the data analysis. It can be assumed that a TBNVs' feature may lead to the outcome of interest only in configuration with other features (i.e., the premise of multiple conjunctural causation), while different configurations of the TBNVs' features may be related to the same outcome (i.e., the premise of causal equifinality). Also, it can be assumed that configurations of TBNVs' features related to the presence of the outcome of interest may not be symmetrical to configurations related to the absence of this outcome (i.e., the premise of causal asymmetry).

Thus, to be able to capture the complexity of relationships between the TBNVs' features and the outcome of interest, I conducted the fsQCA with the support of the algorithm of the R Studio QCA package (Dușa, 2018) for the calibration of measures, analysis of necessity, as well as analysis of sufficiency. As fsQCA may treat varying degrees of case membership in sets, it allows the analysis of both fuzzy sets, from entirely out (value 0) to entirely in (value 1), and crisp sets, as entirely out (value 0) or entirely in (value 1).

Being based on the TBNVs' features, in my analysis, the condition sets have the characteristics of crisp sets (i.e., b2c vs. b2b, "unicorn" vs. "non-unicorn," and digital platform vs. traditional products). Thus, after translating the qualitative data collected for the specific TBNV features into index measures for each of the condition sets, membership in the set is coded in the following way:

1. "1" for b2c-oriented venture and "0" for b2b oriented venture;
2. "1" for digital platform oriented venture and "0" for traditional products oriented venture; and
3. "1" for a venture with the "unicorn" status and "0" for a venture with the "non-unicorn" status.

Having the interest in the TBNVs' features leading to both high and low positive correlation between the TBNV's growth dynamics and related to its web search traffic, in this analysis, two outcome sets are defined, i.e., "high correlation" and "low correlation." As the outcome measure is based on the quantitative scale of correlation scores, the data required calibration for further fsQCA. Calibration is the transformation of raw numerical data into fuzzy-set membership scores that express the degree to which cases belong to a set (Schneider and Wagemann, 2012). Relying on the direct method of calibration (Ragin, 2008), I specified the very high correlation score of 0.9 as the qualitative anchor determining full membership (1) and a very low correlation score of 0.1 as the qualitative anchor determining full non-membership (0) in the "high correlation" set. As the qualitative anchor determining a cross-over point (0.5) for membership in this outcome set, I use the correlation score of 0.499 to avoid the case ambiguity problem that can be caused by the use of the correlation score of 0.5 exactly. The membership in the "low correlation" set is coded as the negation of the correlation scores described above.

Based on the coded and calibrated data, I further conducted analyses of necessity and sufficiency within the fsQCA. To identify the TBNVs' features and/or configurations of these features that are necessary for the outcome of interest, I use the consistency threshold of 0.9 and the relevance threshold of 0.6 in the analysis of necessity (Schneider and Wagemann, 2012). On the other hand, to identify the TBNVs' features and/or configurations of these features that are sufficient for the outcome of interest, I use the consistency threshold of 0.75 and the frequency threshold of 1 in the analysis of sufficiency, supported by the truth table analysis and logical minimization process (Schneider and Wagemann, 2012).

Results of the configurational analysis

The truth table analysis identifies eight different configurations of the three TBNVs' features (equals to 2^3 possible configurations). The minimization process of the fsQCA identified two solutions that are sufficient for the high positive correlation between the TBNV's growth dynamics and related to it web search traffic, covering five different configurations of the TBNVs' features. There is no solution identified to have the sufficiency relation to the low correlation between the TBNV growth dynamics and related to it web search traffic (Table 3.7) – the three remaining configurations are identified to be insufficient for both high and low correlations between a TBNV's growth dynamics and related to it web search traffic. Also, there are no TBNV's features and/or configurations of these features identified to have a relevant necessity relation to neither of the two outcomes of interest.

The first solution (HIGH1) shows that a single TBNV's feature, i.e., the "unicorn" status, is a sufficient condition for a high correlation between its growth dynamics expressed in VC valuations and the related web search traffic. The TBNVs' features of b2c or b2b and digital platform or traditional products appear as conditions of indifference. The solution HIGH1 shows high consistency of 0.82 and substantial coverage of 0.50.

Conversely, the second solution (HIGH2) shows that a combination of b2c and digital platform is sufficient for a high correlation between the TBNV's growth dynamics expressed in VC valuations and related to it web search traffic. In this solution, the TBNV's feature of the "unicorn" or the "non-unicorn" status appears as a condition of indifference. The solution HIGH2 also shows high consistency of 0.81 and significantly lower coverage of 0.20 in comparison to the solution HIGH1.

The overall solution (HIGH1 + HIGH2) shows high consistency of 0.82 and substantial coverage of 0.57.

Solution	**HIGH1**	**HIGH2**	**LOW**
Conditions:			
b2c (●) vs. b2b (○)	-	●	
Digital platform (●) vs. traditional product (○)	-	●	No solution
“Unicorn” (●) vs. “non-unicorn” (○)	●	-	
Solution consistency and coverage:			
Consistency	0.82	0.81	
Raw coverage	0.50	0.20	-
Unique coverage	0.37	0.07	
Overall solution consistency and coverage:			
Overall solution consistency	0.82		
Overall solution coverage	0.57		
No. of cases covered by the overall solution	120		-
No. of cases not covered by the overall solution	121		
Black circles represent core present conditions; dashes indicate conditions of indifference.			

Table 3.7. Results of fsQCA analysis.

3.2.4.7 Outlying cases

In the previous sections, it was examined how various features of new ventures are connected with the correlation between their Google Trends search query data and valuation points. It was demonstrated that the "unicorn" status, focus on the b2c customer segment, and being a digital platform leads to a higher correlation on average that was supported by a configurational analysis. The configurational analysis also demonstrated that to have a strong correlation between a TBNV's dynamics and related to it web search traffic, a company should be a "unicorn" or a b2c-oriented digital platform. Nevertheless, I obtained six "unicorns" (four of which are also b2c-oriented digital platforms) and one "non-unicorn" b2c digital platform that still demonstrates low correlation levels. Bearing in mind that TBNVs, which strongly "unfollow" the taken hypotheses, may provide additional valuable insights, next, an overview of these companies is presented, and the possible reasons for their results are discussed.

In the outliers' list there are seven TBNVs: *Pinterest*[19], max tau = 0.28; *Compass*[20], max tau = 0.14; *Quantenna Communications*[21], max tau = 0.25; *Coinbase*[22], max tau = 0.43; *Coursera*[23], max tau = 0.03; *Marqeta*[24], max tau = 0.42; *TabbedOut*[25], max tau = 0.45. Five of these companies work in the b2c market segment and develop digital platforms (*Pinterest, Compass, Coinbase, Coursera*,

---

[19] Pinterest [WWW Document], n.d. URL https://www.pinterest.com/ (accessed 1.4.20).

[20] Real Estate, Homes for Sale & Apartments for Rent | Compass [WWW Document], n.d. URL https://www.compass.com/ (accessed 1.4.20).

[21] Quantenna – A division of ON Semiconductor Quantenna - A division of ON Semiconductor | [WWW Document], n.d. URL https://www.quantenna.com/ (accessed 1.4.20).

[22] Coinbase - Buy & Sell Bitcoin, Ethereum, and more with trust [WWW Document], n.d. URL https://www.coinbase.com/ (accessed 1.4.20).

[23] Coursera | Build Skills with Online Courses from Top Institutions [WWW Document], n.d. URL https://www.coursera.org/ (accessed 1.4.20).

[24] Payment Processing | Card Issuing | Merchant Services | Marqeta [WWW Document], n.d. URL https://www.marqeta.com/ (accessed 1.4.20).

[25] TabbedOut - Never Wait For Your Check Again [WWW Document], n.d. URL https://www.tabbedout.com/ (accessed 2.13.20).

and *TabbedOut*). I could not derive a straightforward rule why the correlation between GT data and valuation points of companies is weak, but it can be assumed that, for some cases, it may refer to the product and market positioning, while for others, the low correlation may be explained by limitations of the applied methodology.

For instance, the first group includes *Quantenna* Communications developing a discrete technology (Cohen et al., 2000), which is likely to have very low market value if considered separately from the complex product in that it is built-in. Another example is *Coinbase*, which created a digital currency exchange. Due to the novelty of the digital currency and the high volatility of its most famous type (*Bitcoin*[26]), the service's popularity might face a significant influence on the issues related to this topic[27,28]. The company's GT curve obtained a huge peak when digital currency started to bring the attention of a wide audience and dropped almost to the "pre-peak" values when the hype ended.

Coursera and *Pinterest* have similar graphs and, thus, I try to find the same reason for their relatively low results. According to the data from open sources, the former TBNV corrected its business model several times[29] what might influence its perception in the eyes of investors and resulted in a slower speed of growth compared to the dynamics of the GT curve. In its turn, *Pinterest* announced its first revenue-generating instrument *Promoted Pins*[30], late, five years after its launch. In both cases, I can observe evidence of particular problems with value appropriation and, at the same

---

[26] Why Bitcoin Has a Volatile Value [WWW Document], n.d. URL https://www.investopedia.com/articles/investing/052014/why-bitcoins-value-so-volatile.asp (accessed 3.4.20).

[27] Coinbase freezes Ethereum Classic trading following attack | TechCrunch [WWW Document], n.d. URL https://techcrunch.com/2019/01/07/coinbase-ethereum-classic-freeze/ (accessed 3.4.20).

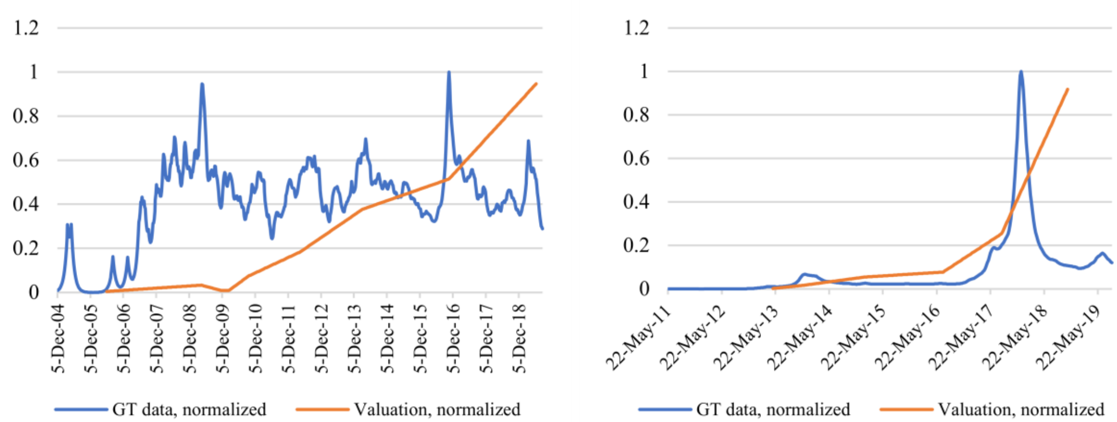
[28] Coinbase may have given away its own Bitcoin Cash surprise | TechCrunch [WWW Document], n.d. URL https://techcrunch.com/2017/12/20/coinbase-bch-bitcoin-cash-api-reddit/ (accessed 3.4.20).

[29] Coursera, Inc. - Wikipedia - Business model [WWW Document], n.d. URL https://en.wikipedia.org/wiki/Coursera#Business_model (accessed 3.4.20).

[30] Pinterest Announces Promoted Pins [WWW Document], n.d. URL https://www.verticalresponse.com/blog/pinterest-announces-promoted-pins/ (accessed 3.4.20).

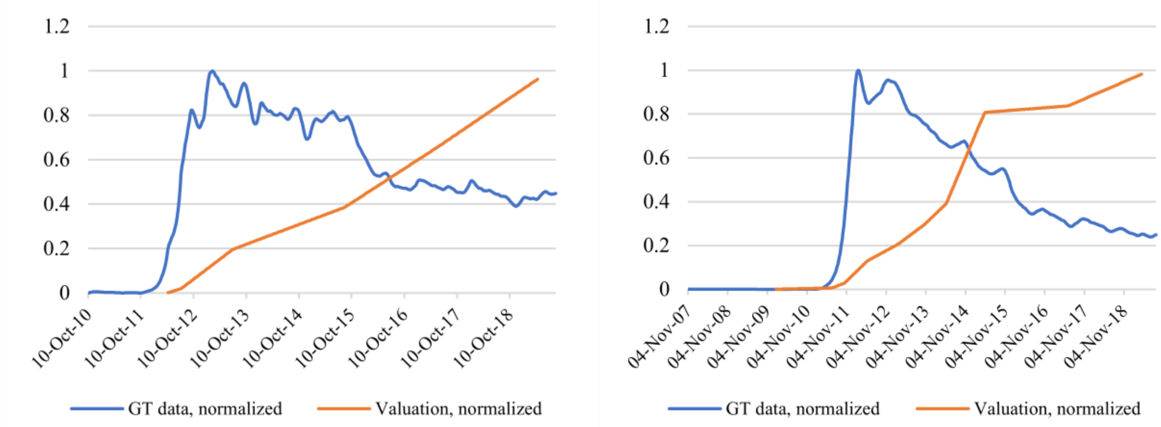
time, successful value creation (Teece, 1986) that resulted in the fast growth of public interest in web search queries with significantly slower growth in valuation.

The second group of cases, whose low correlations may be justified by the methodology limitations, contain the rest of the outlying cases. It can be observed that GT data of *Marqeta* and *TabbedOut* face a significant level of noise, which should have been filtered out by a higher filtering coefficient. However, since one level of filtering was selected for all companies, I cannot manipulate it in specific cases after analysis of the results. In future studies, this issue is planned to be solved by developing a case-dependent filtering algorithm. The last outlier, *Compass*, demonstrates a similar trajectory of GT data and growth in valuation, which is lagged for two hundred weeks. However, in order to exclude a large number of false positives, I limited the maximum shift in the cross-correlation algorithm by the beginning of the valuation curve. Hence, in the case of *Compass*, the lag of two hundred points cannot be reached. This limitation is also planned to be eliminated in future studies. All plots for the outlying cases are presented (Fig. 3.12-3.15).



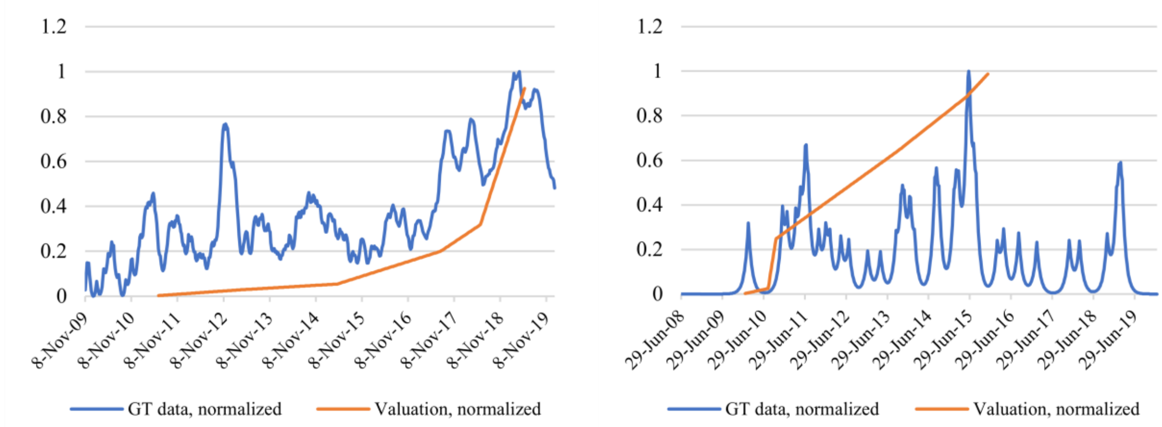
(a) (b)

Figure 3.12. GT data and valuation curve of Quantenna Communications (a) and Coinbase (b). On the X-axis are dates, on the Y-axis is the GT index, normalized.



(a) (b)

Figure 3.13. GT data and valuation curve of Coursera (a) and Pinterest (b). On the X-axis are dates, on the Y-axis is the GT index, normalized.



(a) (b)

Figure 3.14. GT data and valuation curve of Marqeta (a) and TabbedOut (b). On the X-axis are dates, on the Y-axis is the GT index, normalized.

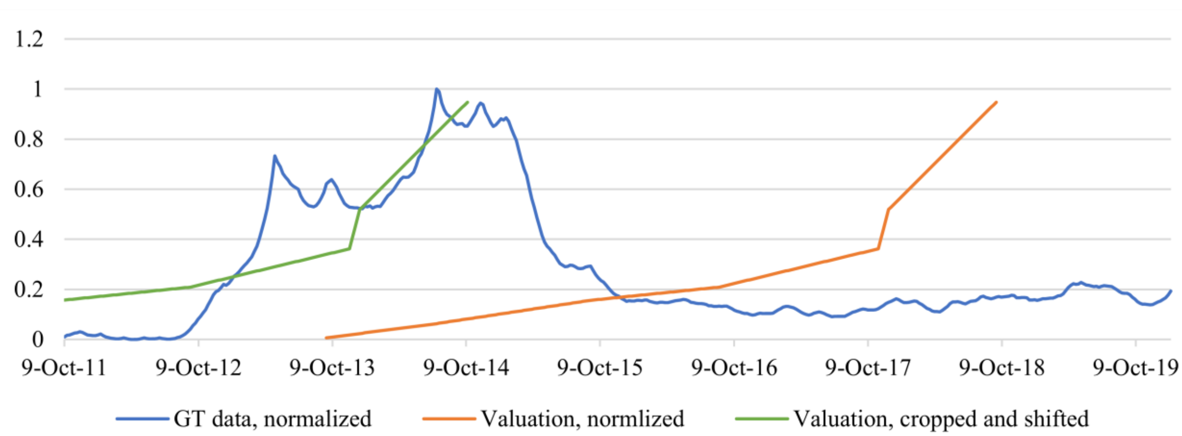


Figure 3.15. GT data and valuation curve of Compass. On the X-axis are dates, on the Y-axis is the GT index, normalized.

3.2.5 Discussion

Across the diverse and large sample of TBNVs that have been analyzed in this paper, I observe that web search traffic generally correlates well with the companies' growth dynamics. Namely, the majority of technology-based new ventures from the sample, 66.8% of them, have a correlation coefficient (Kendall's tau) between Google Trends search queries on their brand name and valuations they achieved through rounds of VC investments, higher than 0.5, without a lag, and statistically significant with all p-levels lower than 0.01. The additional 16.2% of companies from the sample have the same high correlation result when time shift between two sources of data is identified and taken into account through adopted cross-correlation. Altogether, 83% of companies from the sample show a high correlation. Moreover, the average value of the correlation coefficient across the whole sample is 0.66, with a median of 0.70; and the distribution of the correlation levels across the sample is heavily skewed to the left (Fig. 3.16).

To figure out if the obtained correlations do not have spurious character, I implemented the test of the significance level for the obtained correlation coefficients for each case. Due to the fact that in all cases, more than ten points were available for correlational analysis (min = 107 points), I

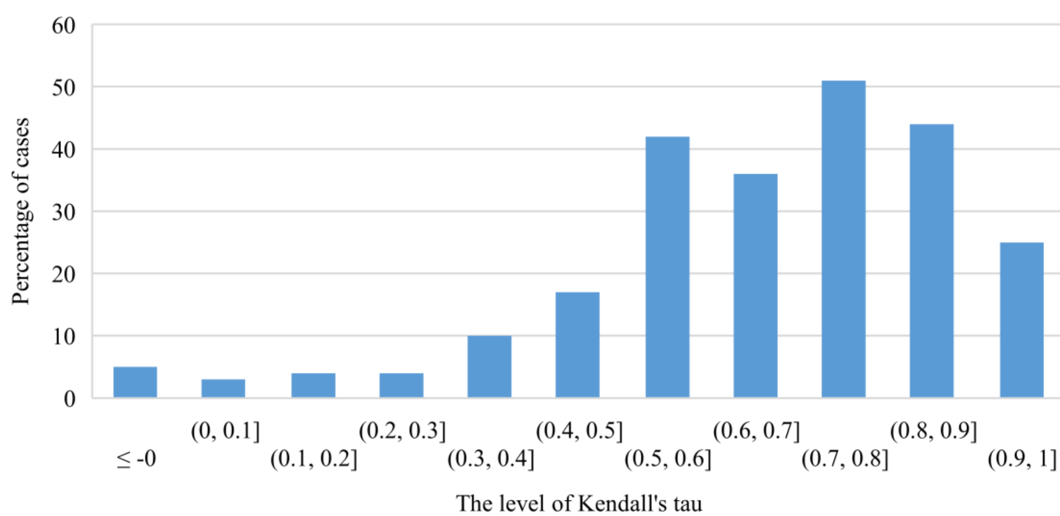


Figure 3.16. Distributions of Kendall's tau coefficient across the sample.

considered the obtained Kendall's tau as the normally distributed random variable (Kobzar, 2006, p. 625). Therefore, to make a conclusion with the chosen level of significance ($\alpha = 0.01$), I compared the obtained correlations with their normally distributed approximations, calculated as a product of the normal distribution α-quantile and the standard deviation of a related normal distribution (Kobzar, 2006). The results demonstrated that in 234 cases (or in 97% of the whole sample), the obtained correlation was not spurious with a 0.99 probability level. Taking to account the high average and median levels of correlation, it can be concluded that the growth dynamics of TBNVs are positively and strongly correlated with the associated web search traffic.

When groups with the high correlation without a time lag (G1) and groups with the low correlation (G3) across the dimensions are analyzed and compared, it was identified that for "unicorns" (success dimension), the difference is highest (86% vs. 6%), then for b2c companies (customer type dimension; 79% vs. 13%) and, then for digital platforms (product type dimension; 76% vs. 13%). In a similar manner, the success dimension exhibits the biggest difference between percentages of companies with the high correlation without a time lag (G1) at its different "poles" ("unicorn" vs. "non-unicorn": 86% vs. 51%), then customer type dimension (b2c vs. b2b: 79% vs. 59%) and, then product type dimension (digital platforms vs. traditional products: 76% vs. 64%). However, when the strength of correlation achieved by the same companies (G1) is compared across the sample, the strongest correlation (highest mean and median values of correlation coefficients) with web search queries were observed for digital platforms (mean = 0.83, median = 0.87), then for "unicorns" (mean = 0.80; median = 0.82), and then for b2c companies (mean = 0.78; median = 0.80).

The results also demonstrate that 39 out of 241 companies in the sample exhibit a strong correlation between their growth dynamics and web search query data once GT data points are shifted in time. A positive time lag signals that a company's value is growing faster than customers' interest in the product they are developing (as approximated by GT data), while the negative signals are the opposite. Experience tells us that negative time lag should be much more frequent in startups' reality. That is reflected in the data – 32 companies display negative time lag, while only seven positive. The positive time lag may happen when VCs are especially enthusiastic about the team, product, or market due to various reasons. On the other hand, the negative lag may be due to the problems with the value appropriation strategy, as it was observed in some outlying cases. In the sample, a time lag is bigger and more frequently identified in "non-unicorns," b2b-oriented companies, and those which market traditional products (Table 3.8). More precisely, in the success dimension, 8% of "unicorns" display a time lag, and 23% of "non-unicorns" (with the average time lag more than two times bigger for "non-unicorns"). In the customer type dimension, it is 8% of b2c companies vs. 21% of b2b companies with a time lag (with the average time lag almost two times bigger for b2b companies), while in the product type dimension, it is 11% of platform companies vs. 18% of non-platform companies (with the average time lag almost two times bigger for b2b companies).

Overall, these results suggest that the correlation between web search traffic and TBNVs growth dynamics is stronger for a) when ventures are more successful in attracting venture capital (H2a), especially for the most successful companies – "unicorns" (H2); b) when companies' customers are individuals (not the other businesses) (H3); and 3) when companies' products are in the form of digital platforms (not traditional products) (H4).

A better understanding of complex causal relationships between the three dimensions (or more precisely poles of these dimensions, i.e., "unicorn" vs. "non-unicorn," b2c vs. b2b, and digital platform vs. traditional products) and high correlation comes after applying fuzzy-set Qualitative

	"Unicorn"		"Non-unicorn"	
	Digital platform	Traditional product	Digital platform	Traditional product
b2c	1/28	0/23	3/14	3/26
b2b	1/6	6/49	1/5	24/90

Table 3.8. The distribution of cases with time lag in relation to all cases.

Comparative Analysis on the dataset. The results of the configurational analysis demonstrate that being a “unicorn” or a b2c-oriented digital platform are sufficient conditions that lead to a high correlation between web search traffic and TBNVs’ growth dynamics, proving H5. Solutions determined by these two conditions cover, with high consistency (0.82) and relatively high coverage (0.57), five out of eight configurations in the analysis (Table 3.9). The three remaining configurations do not lead to low correlations (there are no configurations that lead to low correlation!). They are inconsistently related to both outcome sets, i.e., both high and low correlation between a TBNVs growth dynamics and web search traffic.

The research makes two contributions. First, by demonstrating that changes in web search traffic reflect TBNVs’ growth dynamics well, a new methodology is verified – a tool and data source – for analyzing and researching the growth of recently formed growth-oriented companies. In this way, I contribute to the extant literature on firm growth (Aldrich, 1990; Davila et al., 2003; Greiner, 1972; Kazanjian and Drazin, 1990; Penrose, 1952; Shane and Venkataraman, 2000). Being one of the key topics in the entrepreneurship (and management) literature, growth research has been attracting continuous and significant interest but achieved only limited progress in recent years (Gilbert et al., 2006; McKelvie and Wiklund, 2010; Shepherd and Wiklund, 2009). By focusing on empirical and quantitative analysis (as recommended by Coad, 2007 and Achtenhagen et al., 2010) and verifying new indicators (as recommended by Weinzimmer et al., 1998) derived from open data, I validate new methodology allowing new insights into the “how” aspect of growth. This is a necessary and fundamental question that needs to be better understood to move the field forward (McKelvie and

	“Unicorn”		“Non-unicorn”	
	Digital platform	Traditional product	Digital platform	Traditional product
b2c				
b2b				

Table 3.9. Configurations of features, which lead to the high correlations between TBNVs’ GT data and growth dynamics (*green*), and which are indifferent (*yellow*).

Wiklund, 2010). The results reveal the potential of Google Trends data to be used as a proxy measure of growth instead of non-public and rarely available measures like sales, employee, and market share growth. Overcoming the limitations of the existing approaches, Google Trends data—which are public, free, easy to collect, available from the first day of company existence, and almost for each company — can help in building data-driven trajectories that will more accurately and, even, *in real-time* reflect TBNVs' growth paths. These evolution curves should make it possible to revisit some old answers as well as to ask new questions and to come up with more solid concepts, theories, and predictions. That is especially true in the case of "unicorns" and b2c platform companies, but relevant even in the case of all other high-growth TBNVs, in the case of which search traffic can still be useful for analyzing their growth dynamics, albeit to a more limited degree.

Second, this is a pioneering study to use Google Trends data – big data created from human interaction with the Internet – to analyze startups and high-potential ventures emerging from them. Hence, I contribute to the recent literature using Google Trends data in business research (Chumnumpan and Shi, 2019; France et al., 2021; Jun et al., 2018, 2014b, 2014a). Several previous studies showed that web search traffic information could help in understanding the adoption of technology or the purchase of a product (Chumnumpan and Shi, 2019; Goel et al., 2010; Jun et al., 2014a, 2014b; Jun and Park, 2016). However, all these studies used established (mostly large and well-known) companies and their, more or less, known products. Unlike established firms, which know what they do, for whom, who pays for it and how much, how a solution is delivered, how money is collected, and have an internal organization that serves all these activities, TBNVs during their startup phases are searching to answer all these questions and, thus, behave differently in many aspects (Blank, 2013). Thus, the study extends previous applications of Google Trends data from established companies to technology-based new ventures (including their startup stages) and from technology management to entrepreneurship research.

3.2.6 Study 1 – conclusions, limitations, and implications

Startups and high-potential technology-based new ventures are “black boxes.” They share only a limited amount of data – those they want people to see and have time to make public. This fact makes it hard to study startups. Academic researchers and analysts from venture funds and policy-making bodies use different approaches to connect the pieces of data to explain and predict real-life events. Some of the attempts resulted in successful empirical methods and some in viable theories – but with ample space for improvements. In this study, based on a diverse sample of 241 US-based TBNVs from a variety of industries, it is demonstrated that web-search traffic information, in particular Google Trends data, can serve as a powerful source of high-quality data for analyzing growth trajectories of high potential technology-based new ventures emerged from startups. The results suggest that for the most successful companies (“unicorns”) and consumer-oriented digital platforms (i.e., b2c digital platform companies) proposed approach may become what X-ray chamber is for studying the human body – cheap, easy, and non-invasive way to understand what is going on inside a technology-based new venture.

However, the research is not without limitations. First, the sample is US-centered, and, thus, the results should be carefully generalized in other regions, especially in China. However, this selection was intentional as the US is the world-leading market for successful TBNVs and VCs, and, at the same time, Google is the dominant search engine. Second, “unicorns” are over-represented in the sample (44%). Although this may influence some of the results, such an amount of “unicorns” was obtained “organically” during the companies’ selection process and according to the rules described. Since “unicorns” are more expected to attract several series of funding during their lifecycles, it was more likely to find for them enough valuation data points needed for analysis. Further, the study is based on two valuable sources of data – Google Trends data and companies’ valuation information from the Crunchbase and CB Insights databases, both of which impose some limitations. Google

Trends provides processed rather than raw data points, so the results depend on unknown processing methodology. However, it is the same across the sample, so this issue is seen to be not too influential. The level of noise in some of the GT data was too high, so the level of filtration that was chosen was not enough to avoid noise and detect the main trend. Thus, in future studies, it may be useful to develop a filtering method, which will depend on the parameters of data and vary among cases. Finally, a great majority of companies do not make their valuation data public. Although Crunchbase and CB Insights provide deep insight, it is very hard to verify data objectively. Therefore, I cannot exclude the fact that some valuation points were provided with an error. However, since the source of data is similar for all cases, again, I believe it does not influence the results. I also noticed that the date of the company foundation also varies when taken from different sources. Although the data sources were triangulated, it cannot be undoubtedly claimed that the starting date of analysis should not be earlier.

the research opens a wide range of possibilities for future applications in practice and academia. I see the opportunity for a wide application of Google Trends data as a proxy for analyzing technology-based new ventures' dynamics of development. For instance, TBNVs' GT data may be valuable for a better understanding of marketing strategies, business models, and intellectual property management practices used in technology-based new ventures and their results. That especially may be the case in understanding, in practice, frequently used terms, like product-market fit or business model validation, which still lacks appropriate tools for a fuller explanation. It can be assumed that this can help to predict the future development of the early-stage ventures, what, in turn, will positively influence the development of the entrepreneurship and innovation management areas.

Second, the methodology of using GT data for analyzing the growth dynamics of a particular venture can be slightly modified and applied for growth prediction purposes. Since GT data is a very comprehensive measure (time series can be presented even in the minutes scale) and since I demonstrated its correlation with companies' valuation dynamics, I expect that it can serve as a basis for building company-related mathematical models of evolution and future growth. Whereas mathematical models work in both directions, it can be inferred that next to explaining historical data, they can also be used for predicting companies’ future growth or decline. I aim to tackle this question in future studies.

Third, the methodology developed in this paper can be further improved and studied. Additional mathematical apparatus can be applied to improve achieved results. It can be assumed that more complex statistical analyses may uncover more dependencies or study more dimensions (e.g., appropriability regimes and IP rights), enhancing the generalizability of the results.

Finally, the link found in the current research implies the positive correlation between two sets of data but does not tell anything about causal connection. Does the change in valuation cause stable growth in the public interest? Or, vice versa, the high amount of search queries leads to the rise in valuation? Or, maybe, these two processes reinforce each other? Answering these questions in future studies will lead to a deeper understanding of new ventures' evolution process and the premises of their success or failure.

Considering the implications for practice, the research adds value by verifying the additional source of data that venture capitalists may employ, next to existing sources, in the investment decision-making process. By proposing the objective source of data with a description of use, my study can provide meaningful benefits in identifying potential market leaders and decreasing the information asymmetry and risk. With further improvements and increased ability to make more solid data-driven decisions, the proposed methodology may even make venture capital not so *venture* anymore.

Chapter 4. The character of TBNVs growth trajectory

4.1 Introduction

According to the OLC theory, new ventures (and, especially, technology-based ones) are likely to evolve through several phases, altogether forming an organization lifecycle (Adizes, 1979; Greiner, 1972; Hanks et al., 1994; Kazanjian, 1988; Lester et al., 2008; Miller and Friesen, 1984; Muhos et al., 2010; Penrose, 1952; Rutherford et al., 2003; Scott and Bruce, 1987; Tam and Gray, 2016). The mainstream studies bridge this life cycle to a living organisms' evolution process, according to which an organization experiences birth, several phase transitions directed by the need to solve the management crisis (Greiner, 1972; Kazanjian and Drazin, 1989; Lester et al., 2008), maturity, and sometimes revival or death. Some of these concepts describe the growth part between birth and maturity as having an S-curved shape (Picken, 2017; Söderling, 1998); others insist on the pure exponential growth part (Ismail et al., 2014). The rest – if to go into this question – simply draw curves of various forms divide them into phases without proposing any mathematical reference (Adizes, 1979; Greiner, 1972; Kazanjian, 1988). What is similar in all studies, which question the shape of the new ventures' growth curve, is the fact that none of them provide *any* empirical evidence proving the proposed trajectory. I could not find the discussion of this phenomenon premises and can only speculate that either this question was not possible to research due to the absence of statistically significant data, either it was not seen as important to discuss, either both. At the same time, scholars acknowledge the fact that "there is no basis for conceptual and empirical alignment between stage models" (Garnsey et al., 2006, p. 3), and I believe that proposed in the current study data-driven mathematical model of TBNVs growth can become this basis.

4.2 Theoretical concept

Considering the studies, which discuss exactly the growth curve shape, two approaches can be noted. First is reflected in Soderling's in-depth comparison between the organizational lifecycle and biological systems (Söderling, 1998). Based on the previous OLC models (e.g., Adizes, 1979; Greiner, 1972; Lippitt and Schmidt, 1967; Scott and Bruce, 1987) and on the theory of organic systems (Land, 1973), the author developed the framework, which includes three distinctive phases: Phase 1 - a formative phase, Phase 2 - a normative phase and Phase 3 - an integrative phase (Fig. 4.1).

According to Söderling, in Phase 1, a system emerges when complementary elements are establishing need bonds that happen in order to set up a growth pattern. The key strategy is "see-act" since the system does not have experience on how to survive in the given environment. In Phase 2, a system is searching for security by integrating elements and bonds that are similar to the ones that were found during Phase 1. This period can be divided into two parts due to the fact that the system experiences dependence on its environment during the first part (Fig. 4.1, 2a) and independence during

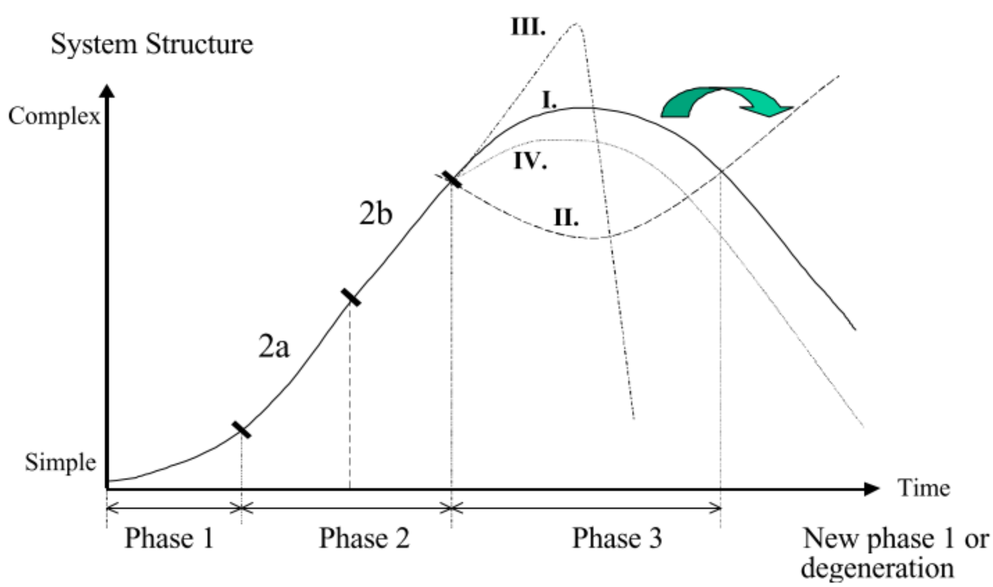


Figure 4.1. Stages of a new venture evolution process by Söderling (Söderling, 1998)

the second (Fig. 4.1, 2b). At the transition to Phase 3, the system has reached a point where it uses more energy to integrate further similarities than its benefit. The right strategy in this phase, according to Söderling, is to maintain an old structure at its present level (Fig. 4.1, curve I) while the excess energy from the previous operations should be used to search for new, dissimilar links. That is needed to break down parts of an old structural pattern and integrate new elements (Fig. 4.1, curve II) (Söderling, 1998). Using an analogy from living systems, Söderling builds the conceptual curve, whose pre-peak part represents an S-curved pattern with three transition points. However, the model is still conceptual, with no empirical evidence provided.

In the second study considering the question of growth trajectory, Picken (Picken, 2017) provides the sigmoid representation of the TBNV's development pattern, which has four phases (Fig. 4.2): Phase 1 – *Startup*, Phase 2 – *Transition*, Phase 3 – *Scaling*, Phase 4 – *Exit*. From his point of view and following the previous concepts (Greiner, 1972; Kazanjian, 1988), these phases are defined depending on the principal challenges faced by the founders. However, the names and the meaning of

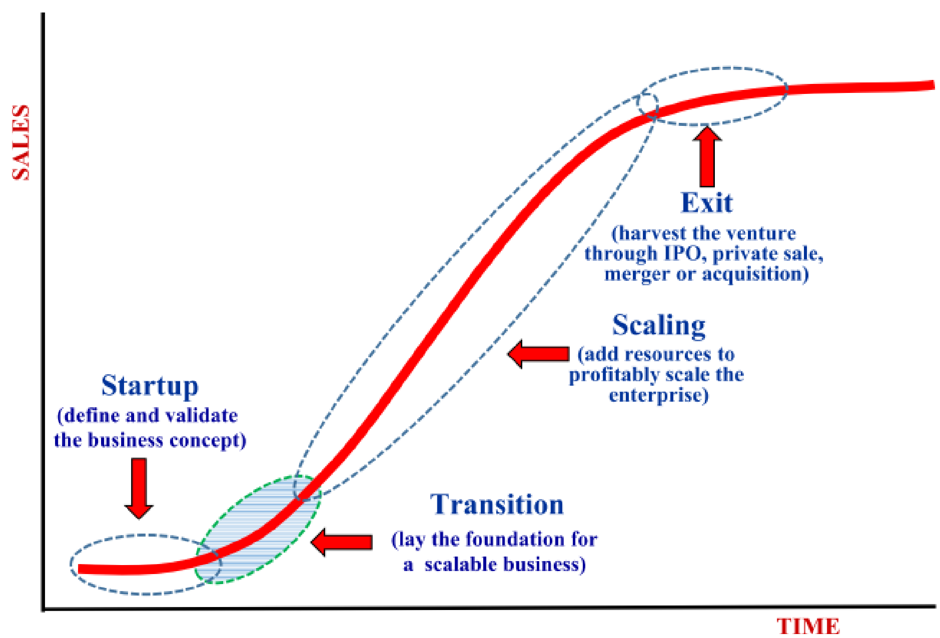


Figure 4.2. Phases of a new venture growth by Picken (Picken, 2017)

the phases and challenges they are driven by are slightly adapted to the actual state of the art. Thus, in the *Startup* phase, this challenge is the identification and validation of the business concept, while in the *Transition* stage, the key challenge is to develop "a bridge" between the informal structure of a startup and a more "disciplined" form required for rapid scaling. Next, according to Picken, in the Scaling phase, founders should allocate significant resources to leverage processes and partnerships in order to grow within the validated during the Startup phase business concept (Picken, 2017). During the last phase – Exit – founders, as well as the investors of a TBNV, harvest the value accumulated during the previous steps and benefit from a venture development. The model of Picken reflects actual core principles of a TBNV's evolution and could be successfully bridged to other existing OLC models. Since it applies the current knowledge of TBNVs evolution phases and utilizes practice-oriented concepts (e.g., business model validation (Blank, 2013)), it can be an extremely useful for nascent startups and TBNVs, which emerge from them.

However, the provided curve, the logic, and the division on phases again lack the empirical evidence and, thus, leave the model with the same drawbacks acknowledged in the previous studies (Lester et al., 2008; Phelps et al., 2007; Rutherford et al., 2003). Adding to that, in the first footnote, the author claims: "the model of organizational development reflects the collaborative contribution of the entrepreneurship faculty at the Jindal School" (Picken, 2017, p. 2) that may not be enough for the rigorous scientific evidence of the model validity.

Agreeing with the biological analogy of growing organisms, in the first part of this thesis chapter, it is questioned whether the shape of TBNVs growth has a sigmoid character (S-curve) with the exponential growth in the beginning, exponential deacceleration, and saturation after some period of company evolution (Cramer, 2002). In the most canonical form, an S-curve can be described by a logistic equation:

$$f(x) = \frac{a}{1+e^{-k(x-x_c)}}, \quad (4.1)$$

where a – the curve's maximum value, i.e., amplitude, k – the logistic growth rate or steepness of the curve, x_c – the x-value of the curve's midpoint, and e – the natural logarithm base.

An S-curve model is often applied by scientists to describe evolution processes of all kinds, including ones related to innovations (Kucharavy and Guio, 2007; Modis, 2007). For instance, in his seminal work, Everett Rogers employs an S-curve model to conceptually explain the process of innovation diffusion (Rogers, 1962). He asserts and further discusses that "(m)ost innovations have an s-shaped rate of adoption" while differing in the slope of "S" (Rogers, 1962, p. 23) and, thus, raise one of the key questions of the innovation diffusion process: why some innovations are adopted in the market faster and have a steeper shape of the adoption S-curve, while other are spreading in a slower pace and, therefore, present the more gradual S-curve. Ideas, models, and concepts proposed by Rogers on the innovation diffusion process gained significant popularity in many innovation areas, resulting in more than 130,000 citations[31]. In the work of another well-known scholar, Clayton Christensen, the concept of innovation value also utilizes a sigmoid model (Christensen, 2011). From his perspective, the first iteration of a new product provides minimum value to a customer, while, during the creation of the base for innovation, this value increases exponentially until reaching the maximum level of improvements, when the value again does not change significantly (Christensen, 2011). This model became fundamental for the concept of "disrupting innovation" and for the whole innovation management science (Shilling, 2013; Tidd et al., 2005). In other examples, logistic curves are heavily employed for explaining and forecasting the evolution of markets (Dyson et al., 1965; Meade, 1985), growth in technology patents (Bengisu and Nekhili, 2006; Gao et al., 2013), and

[31] Rogers: Diffusion of innovations - Google Scholar [WWW Document], n.d. URL https://scholar.google.ru/scholar?cites=7959597870782428962&as_sdt=2005&sciodt=0,5&hl=en (accessed 2.21.22).

diffusion of new products (Bass et al., 1994; Bemmaor, 1992; Rogers et al., 2019). Bearing in mind the innovativeness as the key distinguishable feature of TBNVs and applying the previous discussion, I propose the following hypothesis:

H6: In the majority of cases, TBNVs growth dynamics follow an S-curve pattern.

Next, the previous step analysis results are questioned from the configurational perspective. Regardless of the results of H6 testing, I want to understand whether these outcomes are or are not driven by various configurations of TBNVs features. The majority of taken dimensions, which form these configurations, were applied from the previous study (i.e., “success,” customer type, and product type dimensions), while two are additionally discussed and employed. The first additional dimension considers whether a TBNV develops and sells a stand-alone product or works under the service business model (Hurmelinna-Laukkanen and Ritala, 2010; Seip et al., 2018). Product and service strategies are known to differ by various parameters: in terms of IP management process (Seip et al., 2018), customer engagement (Xue et al., 2005), innovation management (Hurmelinna-Laukkanen and Ritala, 2010), etc. Therefore, I want to check if this variability, while reflected by the product-service dichotomy, may influence the character of the evolution curve. Another dimension added to this part of the research reflects the company exit, in particular, the IPO event (Chang, 2004; Gornall and Strebulaev, 2017). There are two common types of TBNVs’ financial exits, which are often analyzed in the scientific studies (DeTienne, 2010; DeTienne et al., 2015): Merger and Acquisition (M&A) and Initial Public Offering (IPO). The first type is described by selling a venture to a bigger company with the full transfer of founder’s control (Cumming, 2008), but after the first data tryouts, it was found that this type of exit may happen significantly before the venture’s maturation. Since in this study I aim to study the full growth process, i.e., starting with birth and ending with reaching the maximum, I excluded the M&A-experienced TBNVs and focused only on those who had an IPO exit or any other late-stage investment according to the Crunchbase classification[8]. The event of exiting through IPO is quite rare and, hence, characterized by the significant public attention (Chang and Kwon, 2020; Gornall and Strebulaev, 2017) that may influence the level of the public interest and the shape of the GT curve. Summing up, from a configurational perspective, there can be various configurations taken and considered above features, which may lead to the same outcome of an S-curve shape. Thus, I hypothesize:

H7: There is different configurations of TBNVs' features, i.e., the b2c vs. b2b, "unicorn" status vs. "non-unicorn," digital platform vs. traditional product, product vs. service, and IPO vs. non-IPO, that are equifinal in achieving the S-curve growth model.

4.3 Analysis and results

4.3.1 Sample and data collection

To test the proposed hypotheses, I utilize TBNVs Google Trends data understanding it as a valid proxy reflecting the growth dynamics of a TBNV as it was shown in Chapter 3 and discussed in Malyy et al., 2021. I employed the Crunchbase database[8] as a source of companies' data and built the initial sample of US-based TBNVs different from the previous research as I have fewer funding-related limitations. However, TBNVs are still needed to be distinguished from SMEs (both of them are listed in the selected source of data), which can be achieved by filtering only to the companies which have at least *some* funding history. For the purpose of this study, the first applied filter limited the sample to the companies which have had three and more funding events.

Secondly, the fact that I aim to examine whether the TBNVs growth process follows an S-curve trajectory and whether it depends on TBNVs' particular combination of features, I need the sample companies to have a higher probability of presenting the full growth part of their evolution curves. In other words, it is needed to examine TBNVs, which most probably have finished their growth. At the initial stage, I achieve that by including two additional filters: the company should be founded no later than December 31, 2015, and experience the later stage investment evidenced by the Crunchbase tags "Late Stage Venture," "Private Equity," and "IPO." Of course, there can be a situation when a company is still on the growth period of its lifecycle but experienced any of these events, but I aim to overcome this issue in the last phase of the sampling process.

The third filter limited the sample only to the TBNVs, which were founded after January 1, 2009. Since in the previous study, it was demonstrated that company foundation date may not be precise and bearing in mind that Google Trends provides search query statistics only from January 1, 2004, a five-year pre-foundation window was taken to overcome this issue. All the applied Crunchbase filters led us to the initial sample of 1527 TBNVs (Fig. 4.3).

During the next step, taking to account that a smaller sub-sample can significantly increase the speed of data analysis while keeping the core effects[32], I randomly selected the 500 TBNVs from the initial sample and controlled the distribution of the funding events number by the Pearson's chi-square test in order to eliminate possible sampling bias (Moore et al., 2009). After that, following the procedure described in Appendix C[33], I assessed the TBNVs GT data, excluded those cases, which were scored as "bad," and obtained a sample of 343 new ventures. As for the final step, I collected GT data of these companies in a weekly dimension and tested if there were enough data points in the pre-foundation and in the post-maximum windows with the thresholds of 5 and 0.5 years respectively. The former check is needed to overcome the risk of inaccurate foundation date identification, while

[32] What Is the Purpose of Sampling in Research? | CloudResearch [WWW Document], n.d. URL https://www.cloudresearch.com/resources/guides/sampling/what-is-the-purpose-of-sampling-in-research/ (accessed 9.27.21).

[33] The criterion "Amount of the related search queries" was slightly adopted for this study. Despite having the three levels as in the original assessment criteria described in the Appendix C, I added the additional fourth level. Thus, the grading for the criterion "Amount of the related search queries" for the second study looks as follows: Good case, 1 point - ≥ 10 related search queries; Fine case, 0.7 points – from 5 to 10; Suspicious case, 0.3 points – from 0 to 5; Bad case, 0 points – 0 related search queries.

Also, the criterion "Systematic noise in the time series" was also calculated slightly differently. Instead of the 1st year GT data mean, I took the mean of the pre-foundation period, i.e., approx. 5 years in the majority of cases.

the latter confirms that the company has reached the maximum by starting to demonstrate the declining trend and, thus, has a complete growth period. The last selection step provided us with the 246 TBNVs (Appendix B), which were further used for analysis. In each step of selecting, I controlled the distribution of a number of funding rounds across the sample with the *Pearson's chi-square test*, for instance, the probability that the randomly selected at step 2 sample of 500 TBNVs and the final sample of 246 cases have the same distribution of funding rounds amount equals 1.0 with equal medians of 5.

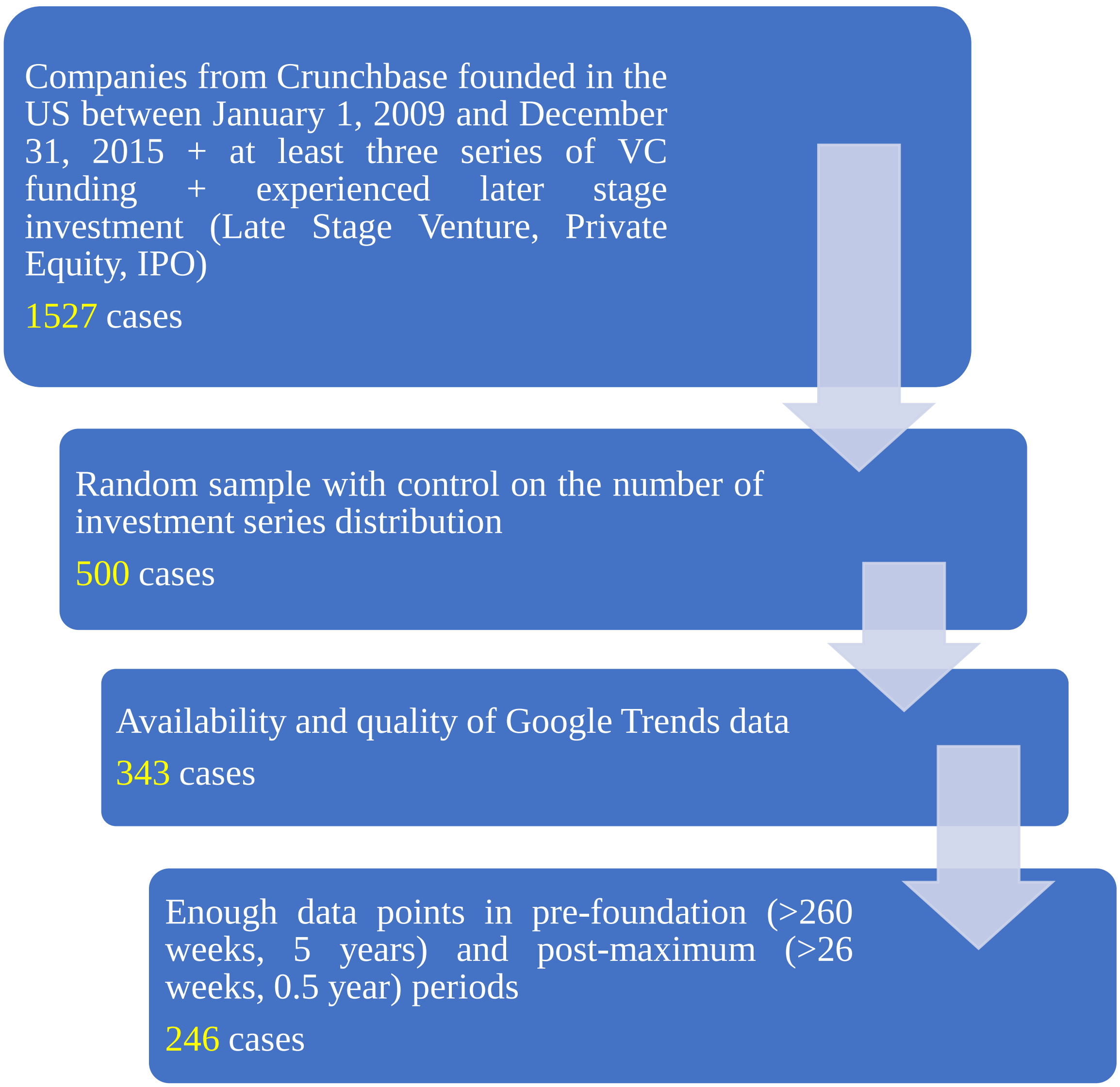


Figure 4.3. The sample collection process

4.3.2 Methodology

With a goal to test the proposed hypotheses, I first prepared the data to be analyzed. I employed the double exponential smoothing by the Holt-Winters method, which is known to demonstrate the rigid results for the time-series data (Armstrong, 2002; Chatfield, 1975). For the purpose of this study, I focused on the fast noise filtering by the simple exponential smoothing (i.e., *smoothing_level = 0.03*) and the smoothing of the trend component while accepting that the trend does not depend on timing and, thus, has an additive rather than multiplicative character (i.e., *smoothing_trend = 0.01, trend = 'add'*). The smoothing was implemented by the *statsmodels.tsa.holtwinters.ExponentialSmoothing.fit* function of the *statsmodels* Python library[34]. Next, the GT data were normalized between 0 and 1.

In the data preparation final step, I cropped the sample TBNVs GT data, which lay outside the selected boundaries, i.e., in more than five years before foundation and in 0.5 years after reaching the maximum (Fig. 4.4). Since the S-curve and some of the other models selected for comparison cannot describe the after-maximum decline partly due to their nature, the long decline tail after

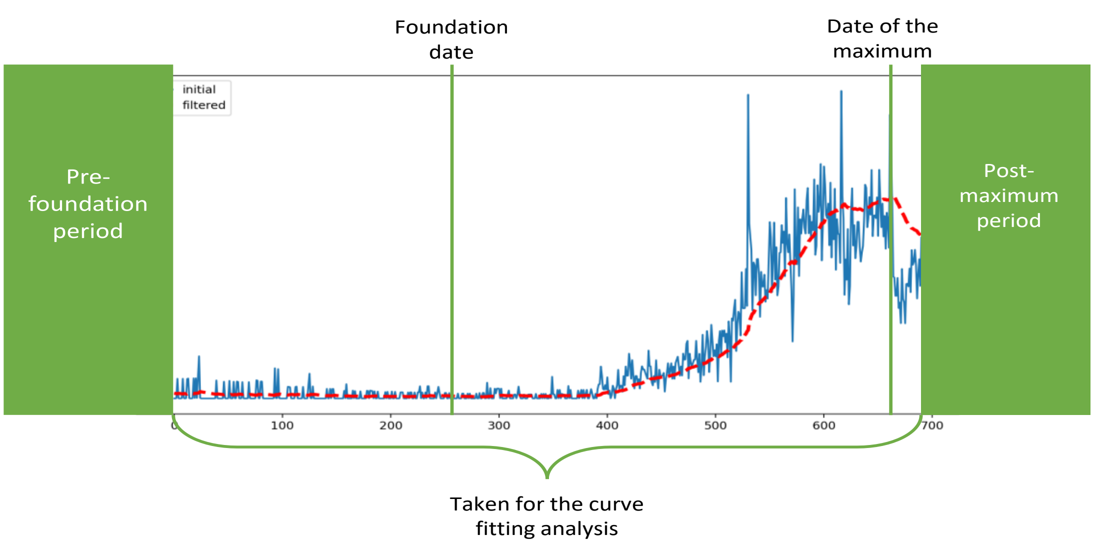


Figure 4.4. The data-cropping process and identification of the final "piece" taken for analysis

[34] statsmodels.tsa.holtwinters.ExponentialSmoothing.fit — statsmodels [WWW Document], n.d. URL https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.fit.html#statsmodels.tsa.holtwinters.ExponentialSmoothing.fit (accessed 9.1.21).

reaching the maximum may demonstrate strong false tendencies. The long pre-foundation data, in its turn, may be distorted by some related search term, which was captured by Google before the company of interest entered the market and started to gain significant attention from the public audience to be accounted by the GT algorithms and separated from the other terms. During the algorithms' tests, I have also evidenced the cases with some other-term noise in the pre-foundation period, which could have heavily influenced the further curve-fitting process. To overcome this issue, I applied the additional filter, which weighted the pre-foundation time-series points by the value exponentially depending on the distance between the point and the foundation date: the further the point from the foundation date, the lower its weight will be. By this method, I kept accounting for the possible error in the foundation date by including the pre-foundation five-year period to the model and, at the same time, eliminated the possible noise in the pre-foundation GT data.

The main analysis is based on the curve fitting process and measuring the quality of the fit by various measures. For the curve fitting, six simple models were selected. Four of them are often employed for describing the various growth processes (exponential, 2nd order polynomial regression,

Name of the model	**Equation**	
Polynomial regression, 1st order	$\widehat{GT}(t) = intercept + b_1 t$	(4.1)
Polynomial regression, 2nd order	$\widehat{GT}(t) = intercept + b_1 t + b_2 t^2$	(4.2)
Polynomial regression, 3rd order	$\widehat{GT}(t) = intercept + b_1 t + b_2 t^2 + b_3 t^3$	(4.3)
Exponential curve	$\widehat{GT}(t) = y_0 + Ae^{R_0 t}$	(4.4)
Logistic model	$\widehat{GT}(t) = \frac{a}{1 + e^{-k(t-t_c)}}$	(4.5)
Gompertz model	$\widehat{GT}(t) = ae^{-e^{-k(t-t_c)}}$	(4.6)

Table 4.1. The models taken for the curve-fitting procedure

Gompertz, and logistic (Kaufmann, 1981)), and the additional two models (1st and 3rd order polynomial regressions) are taken as the reference and exploratory models respectively (Table 4.1). The curve fitting process was implemented in Python by the *curve_fit* function of the *scipy* library[35] and by the *polyfit* function of the *NumPy* library[36].

After the selected models were fit to each sample TBNV GT data, the scores reflecting the quality of the fit were calculated. These qualities included six measures: root mean squared error (RMSE), adjusted R^2 (adj. R^2), Akaike Information Criterion (AIC), adopted Hellinger Distance (aHD), dynamic time warping (DTW), and cross-validation adjusted R^2 (adj. R^2 cv). If the first two are quite common measures in the regression analysis (Armstrong, 2002; Chandler et al., 2011; Kuhn and Johnson, 2013), the rest may raise some questions. The Akaike Information Criterion is based on the information theory (Cover and Thomas, 2006) and assesses the amount of information each compared model loses during the fitting process (Akaike, 1973); thus, the smaller values of AIC reflect the more accurate models. Additionally, since AIC accounts for the number of models' estimated parameters, it deals with the risk of overfitting as well. The Hellinger distance, in its turn, is applied to assess the similarity between two probability distributions, expressed in terms of the Hellinger integral[37] (Wetherill and Weiss, 1962), and was evidenced as the reliable metric for comparing the quality of probability distributions fitted to the Google Trends data (Bauckhage and Kersting, 2016). For the purpose of the current study, I applied the slightly adapted version of this test, bearing in mind that the measured models are not statistical distributions and, thus, the test will not provide absolute distances located between 0 and 1 but only relative values (Wetherill and Weiss, 1962). The next

---

[35] SciPy.org — SciPy.org [WWW Document], n.d. URL https://www.scipy.org/ (accessed 9.27.21).

[36] numpy.polyfit — NumPy v1.21 Manual [WWW Document], n.d. URL https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html (accessed 9.27.21).

[37] Hellinger distance - Encyclopedia of Mathematics [WWW Document], n.d. URL https://encyclopediaofmath.org/index.php?title=Hellinger_distance (accessed 9.2.21).

measure, *dynamic time warping (DTW)*, is often applied in the time-series analysis for measuring similarities between two processes, which may present the same dynamical pattern but may vary in speed (Müller, 2007). The core distinguishing feature of this algorithm is its ability to compare the time-distorted patterns that are particularly useful for the purposes area of voice recognition (Efrat et al., 2007; Salvador and Chan, 2007). In the current study, DTW is intended to measure the similarity between the GT data and the fitting curves while adjusting the possible lags and other time-dependent distortions, which may happen between them (Last, 2016; Papavlasopoulos, 2019; Patel et al., 2018).

The last measure that was employed quantifies the predictive power of the fitted model: cross-validation adj. R^2. It was obtained during the cross-validation procedure implemented with the help of *sklearn.model_selection* function of *scikit-learn* python library[38]. Since the TBNVs GT data has a form of time-series, during the first step of the cross-validation procedure, I implemented a *time series split* to subsamples being employed for further calculations. A time-series split is a type of *k-fold* cross-validation with the difference that subsamples are formed not from randomly selected data points but from sequential folds of train and test data so that each next split of data contains the previous one[39] (Chen et al., 2019; Verly Lopes et al., 2021). The trainsets are used to calculate the model parameters (i.e., "to train") and test sets employed for the cross-validation procedure measured by an adj. R^2. I selected twenty splits for the current analysis as enough; more splits would result in a smaller number of test points, while fewer splits could not be enough to derive rigid averaged results. Next, for each model, I calculated the average adj. R^2 and identified the maximum one. That is to note, the adj. R^2 measures obtained during the cross-validation procedure can be lower than minus one but not higher than one. In addition, coming from the understanding that one of the key features of a time

---

[38] scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation [WWW Document], n.d. URL https://scikit-learn.org/stable/ (accessed 9.2.21).

[39] 3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.24.2 documentation [WWW Document], n.d. URL https://scikit-learn.org/stable/modules/cross_validation.html#time-series-split (accessed 9.2.21).

series fitting model is its forecasting power (Box et al., 2008; Brockwell and Davis, 2016; Shumway and Stoffer, 2017), I have selected *adj.* R^2 *cv* as a benchmark measure of the S-curve tendency in each TBNV and employed in the further configurational analysis. In other words, if GT data of a particular TBNV demonstrate the S-curve tendency by the cross-validation test, this case was counted as having the S-curved growth trajectory. Next, following the methodology described in Malyy et al., 2021, I implemented the configurational analysis taking the selected TBNVs features as an input and related to its S-curve benchmark as an output.

4.3.3 Results

In order to compare the taken growth models by the selected measures, the median values for each model and proposed measure were calculated. The results demonstrate that by five out of six measures, the logistic model outperforms the other growth models. Only the Hellinger distance measure demonstrated that median values have the higher tendency towards the 3rd order polynomial regression, however, differing from the logistic model by less than 1% (Fig. 4.5). According to the AIC metric representing the assessment from the information theory point of view, the logistic fit loses less information from the initial data than the other taken models. DTW coefficient, in its turn, shows that the sigmoid models – both logistic and Gompertz – are significantly better than the other (the median values, respectively, by 17% and 13% are lower than their closest runner-up) explain TBNVs growth data with possible time-dependent distortions like lags and squeezes. In other words, even if a TBNV's growth does not follow an ideal mathematical pattern (like it happens naturally) and sometimes outruns or lags behind the predicted model, it still more tends to the sigmoid shape.

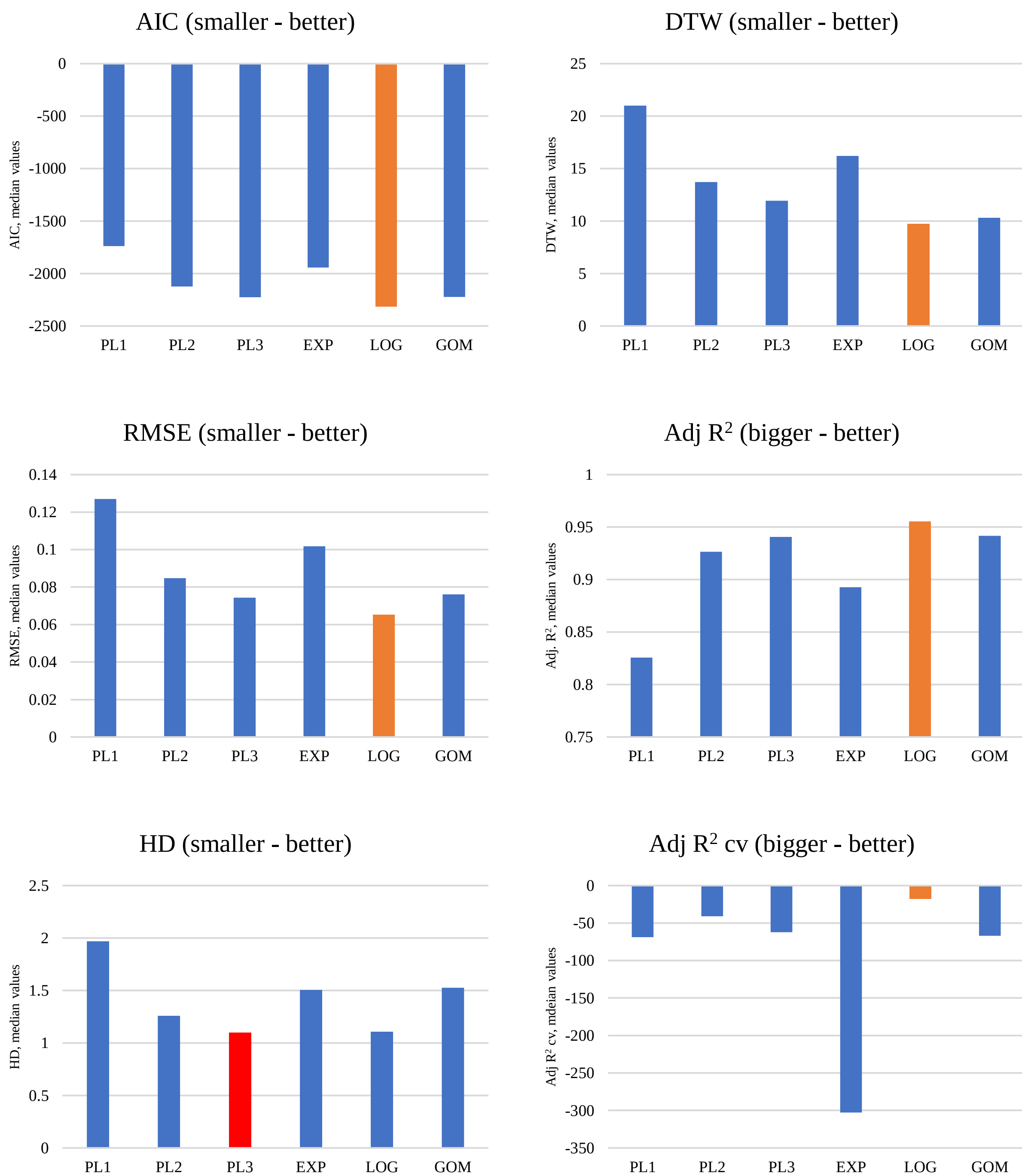


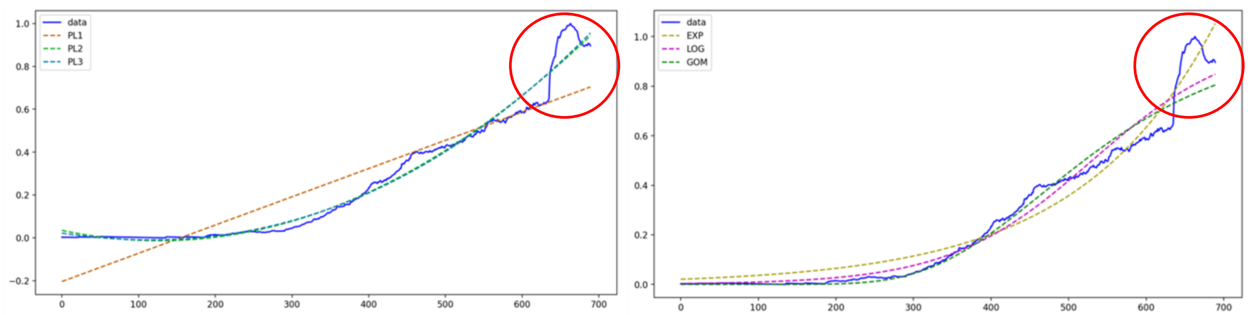
Figure 4.5. The sample median goodness-of-fit measures for the compared models. The orange color denotes the cases where logistic model outperforms the other models, while the red one shows another "winning" model.

According to the results of the cross-validation procedure, it can be observed that the logistic model has the highest forecasting power across the selected growth models with more than twice the difference compared to the runner-up 2nd order polynomial regression model. The S-curve models (i.e., logistic and Gompertz) resulted with the highest adj. R^2 cv values in 78% (192) cases and, thus, it can be inferred that in the majority of sample cases, the growth is better forecasted by the sigmoid curves. In addition, the worst forecasting result demonstrated an exponential model with more than 17 times smaller median adj. R^2 cv than for the logistic one. That makes it possible to conclude that TBNVs growth can hardly be forecasted by the exponential curve and, thus, most likely do not follow the pure exponential trajectory as it was proposed by Ismail (Ismail et al., 2014). The taken more common goodness-of-fit measures (i.e., RMSE and adj. R^2) of the logistic model also outperform the other models' results and have the acceptable absolute values: RMSE is 0.065 (6.5%) and adj. R^2 is 0.96 (Fig. 4.5, RMSE and Adj R^2). In particular, this level of RMSE shows that the logistic model is rather good at predicting the observed TBNVs growth data and explains 96% of its variance according to adj. R^2.

The obtained curve-fitting results were also tested for the statistical significance for each taken model and measure proposed. While accounting for the possible non-normality in the distribution of the results, the Wilcoxon signed-rank test for matched pairs was employed (Moore et al., 2009). During the testing procedure, I took each measure's results obtained for the logistic model as the first sample and the results of other taken growth models for the second. The null hypothesis for this test claims that there is no statistically significant difference between medians of the two taken samples. The significance test results demonstrate that the null hypothesis can be rejected with the 0.05 level of confidence for the majority of the curve-fitting measures and taken growth curves. Only in two cases – Adj R^2 while comparing LOG model to PL3 and HD comparing LOG model to PL2 – the confidence level is higher: 0.08 for the former case and 0.13 for the latter. Nevertheless, these confidence levels are still acceptable (Hair et al., 2010) that, taking to account the significance test results for other growth models and measures, makes it possible to conclude on the high statistical significance of the outcomes.

Summing up, it can be concluded that the logistic model explains, predicts, and forecasts TBNVs growth more accurately than other commonly used growth models while losing less information and accounting for the possible data distortions. Therefore, bearing in mind that 78% of cases demonstrated a strong tendency to S-curve models by the Adj R^2 cv measure, it can be inferred that H6 is supported, and most TBNVs follow a sigmoid trajectory while growing.

In order to test whether the S-curve character of the growth curve is linked to a particular TBNV's feature or combination of features, I implement the configurational analysis while taking the results of the cross-validation procedure as an outcome. According to the truth table (Table 4.2), there are $2^5 = 32$ configurations that can be identified from the five considered TBNVs features, and twenty-one of them are presented by the sample TBNVs. The results demonstrate that seventeen configurations lead to an S-curve (Consistency and Proportional Reduction in Inconsistency (PRI) $\geq$ 0.75), three can either lead or not ($0.5 \leq$ Consistency and PRI < 0.75), and only one does not lead to the sigmoid shape (Consistency and PRI < 0.5): *PagerDuty*, which, according to the selected features, is a b2b unicorn developing a service, monetizing through traditional product scheme, and exited through an IPO. Of course, since the case-based method was applied in this part of the research, I cannot derive any statistically significant conclusions and generalizations (Rumble and Mangematin, 2015; Yin, 2009) and, thus, it cannot be said that companies of the similar configuration to *PagerDuty* will lead to the same result of a "non-S-curviness." Moreover, going deeper into the case of *PagerDuty*, it can be seen that the absence of tendency to an S-curve can be explained by its sudden "fast" stop of growth, which is better approximated by the non-S-shaped models (Fig. 4.6). While the S-curve models have a decreased growth speed closer to the maximum, the *PagerDuty* GT data show the “fast” growth cessation without the deceleration period that is better forecasted by a parabola during the cross-validation procedure. Finding the general rule for the cases similar to *PagerDuty* (not only in relation to its features configuration) requires additional research, which is planned for further studies. Nevertheless, since the majority of configurations are equifinal in reaching the taken outcome, it can be concluded that the S-shaped growth trajectory is not driven by some particular feature or any combination of them, and H7 is supported.



(a) (b)

Figure 4.6. The results of the curve-fitting procedure for *PagerDuty* by the selected equations: (a) polynomial of the 1st, 2nd, and 3rd orders; (b) exponential, logistic, and Gompertz

	Dimensions					Outcome	Results		
Tag	“Unicorn” (1) vs. “non-unicorn” (0)	b2c (1) vs. b2b (0)	Digital platform (1) vs. traditional product (0)	Service (1) vs. product (0)	Exit through IPO (1) vs. not-exited through IPO (0)	S-curve (1) vs. non-S-curve (0)	Consistency	Proportional Reduction in Inconsistency (PRI)	No of cases
Leading to an S-curve	0	1	0	0	1	1	1	1	1
	1	0	0	0	1	1	1	1	1
	1	1	0	0	1	1	1	1	1
	1	1	0	1	1	1	1	1	1
	0	1	0	1	1	1	1	1	2
	0	1	1	1	1	1	1	1	2
	1	0	0	0	0	1	1	1	2
	1	1	1	1	0	1	1	1	2
	1	1	1	1	1	1	1	1	2
	0	1	0	0	0	1	0.889	0.889	9
	0	0	0	0	1	1	0.871	0.871	31
	1	0	0	1	0	1	0.818	0.818	11
	0	1	0	1	0	1	0.818	0.818	33
	1	1	0	1	0	1	0.8	0.8	5
	0	0	0	1	0	1	0.771	0.771	105
	0	0	0	1	1	1	0.75	0.75	4
	0	0	1	1	0	1	0.75	0.75	4
Ambiguous	0	1	1	1	0	0	0.571	0.571	7
	0	0	0	0	0	0	0.571	0.571	21
	1	0	1	1	0	0	0.5	0.5	2
Not leading to an S-curve	1	0	0	1	1	0	0	1	1

Table 4.2. The configurational analysis results.

4.4 Theoretical propositions

In the previous chapter, it was demonstrated that the majority of TBNVs follow the sigmoid pattern while reaching their lifecycle maximum. However, the sigmoid models commonly applied to describe this type of growth (e.g., logistic and Gompertz) have significant limitations, which crucially decrease their utility.

In particular, as it was first mentioned by Theodore Modis in his review of Ray Kurzweil's book *The Singularity is Near* relatively the exponential model representing the beginning of the S-curve pattern, "one-parameter mathematical function" cannot be used "to single out a particular region" on it (Modis, 2006, p. 107). Further, Fred Phillips extends Modis' logic to the logistic model while explaining that this rule is applicable in a similar vein to the "two-parameter curves in which one parameter serves only to locate the asymptote" (Phillips, 2007, p. 717). From Philips' perspective, this is caused by the fact that the logistic model is intrinsically autocatalytic, meaning that the change in the growth speed is driven by the model's parameters and does not depend on any external factors (Phillips, 2007, p. 720). That leads to a conclusion that if applying the logistic model (and the Gompertz one, since it also does not account for the external factors, which can influence the growth speed), it is not possible to identify the point on the curve where the exponential growth began, i.e., the so-called *tipping point* (Gladwell, 2000; Phelps et al., 2007; Phillips, 2007).

At the same time, the tipping point concept is known to play a crucial role in describing the evolution of various "social epidemics" phenomena and, specifically, for a rise of a particular new product. For instance, Malcolm Gladwell, in his book "The Tipping Point: How Little Things Can Make a Big Difference," discusses the dynamics of different social and business processes, like crime level in New York, growth of suicide rates in Micronesia, or the unexpectable rise of the *Hush Puppies* shoes, considering the event of a sudden "tip" as a critical moment "when everything can change at once" (Gladwell, 2000). According to Gladwell, the tipping point is the event in some socio-related process where happens the complete and self-driven shift of paradigms from the constant or linear dynamics to one with the progressive viral growth. The author understands this event as highly effective and influential since the great change is driven by a little factor and, thus, concludes that the tipping point is “a reaffirmation of the potential for change and the power of intelligent action” (Gladwell, 2000).

Despite the detailed and multi-dimensional discussion of the tipping point by Gladwell, he misses one crucial aspect, which significantly lowers the practical applicability of the concept: its mathematical representation. From Gladwell’s book, it is hardly possible to build a precise and trustworthy mathematical model, which would explain the premises of the tipping point appearance and propose the mechanism to identify its location on the curve. The directions to solve this issue were offered by Phillips (Phillips, 2007), who proposed to utilize the Bass model instead of the logistic one and, hence, to overcome the issue of the intrinsic autocatalyticity of the latter. The author demonstrated that, under the *innovation/imitation* paradigm (Bass, 1969), the model’s parameters p and q have the power to explain the tipping point occurrence and, thus, help to identify its location at the S-shaped time series curve. The Bass model and the related paradigm were developed by Frank Bass in 1969 (Bass, 1969), and since those times got huge popularity among innovation forecasting researchers with almost 10,000 citations[40]. The core assumption of this model is that the probability of new product purchases at any time linearly depends on the number of previous buyers. In this concept, all buyers are separated into two groups: (1) *innovators* who decide to adopt a new product without the influence of other individuals but only when subjected to external factors (e.g., advertising), and (2) *imitators* who are adopting a new product due to the pressure of *innovators* and,

[40] A new product growth model for consumer durables - Google Scholar [WWW Document], n.d. URL https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=A+new+product+growth+model+for+consumer+durables+Bass&btnG= (accessed 9.26.20).

therefore, subjected to the internal influence proportional to the *innovators'* number. For the late adopters, this pressure is the highest due to the great number of previous buyers (Bass, 1969). Thus, to explain and predict the likelihood of purchasing a new product, the Bass model incorporates the *innovation/imitation* paradigm and applies two parameters: *p* (coefficient of *innovation* or *external* influence) and *q* (coefficient of *imitation* or *internal* influence) (Eq. 4.7) Under this logic, according to Phillips, the Bass model leads to the meaningful tipping point, happening at the first time period where the total number of adopters become greater than the *p/q* ratio, i.e., *F(t-1)* > *p/q* (Phillips, 2007).

$$f(t) = m\frac{[(p+q)^2/p]e^{-(p+q)t}}{(1+\left(\frac{q}{p}\right)e^{-(p+q)t})^2}, \qquad (4.7)$$

Where *m* – the total number of potential new product adopters, *p* – coefficient of *innovation* or *external* influence, *q* – coefficient of *imitation* or *internal* influence, *t* – time of innovation adoption, *e* - the natural logarithm base.

I assume and aim to demonstrate in this part of the study that the new product diffusion model (i.e., Bass distribution) can be beneficially applied to explain the TBNVs growth dynamics and able to provide a trustworthy mechanism of identifying the tipping point that will make it more beneficial for describing TBNVs growth than the commonly used S-shaped models (i.e., logistic and Gompertz).

To reach this goal, I utilized the same sample of US-based companies, which was obtained at the final step of the sample collection process for the previous part of the empirical research, i.e., containing 246 TBNVs (Appendix B). The GT data related to these companies and taken for further analysis were also the same. Next, again, similarly to the previous part, I applied the curve-fitting procedure with the help of the *curve_fit* function of the Python *scipy* library ("SciPy.org — SciPy.org," n.d.) and fitted the Bass model to TBNVs data (Fig. 4.7).

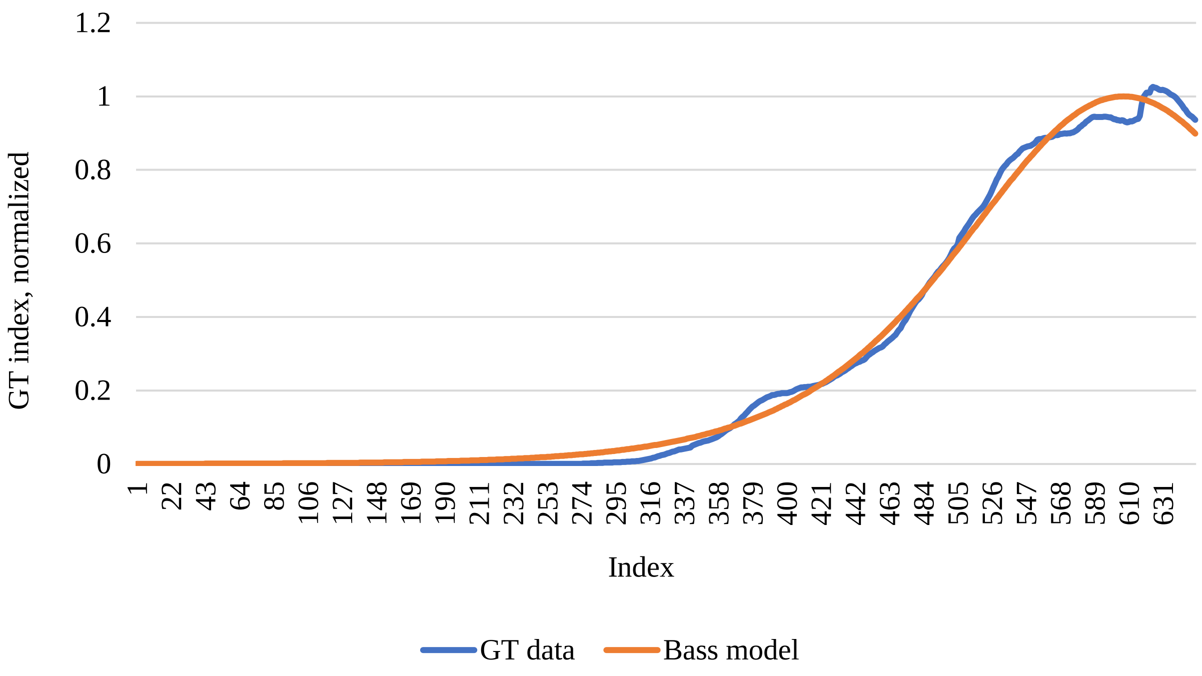


Figure 4.7. Approximation of the *Lyft* GT data by the Bass model

The results demonstrated that the Bass distribution in the form of Eq. 4.7 can be beneficially utilized to explain companies' growth. The adj. R^2 and RMSE measures of the Bass model provide rather similar values if to compare with the logistic and Gompertz curves (Fig. 4.8): the adj. R^2 differences are less than 1% for the Bass/logistic ratio and less than 2% for the Bass/Gompertz, while the RMSE differences are approx. 3% for Bass/logistic and 11.5% for Bass/Gompertz. Therefore, it can be concluded that the Bass model has the same (and, by some results, even higher) explanatory power than the logistic and Gompertz curves; that is, however, so far does not mean that it is more advantageous to apply it for studying TBNVs.

The more important result is that the Bass model can be beneficially applied to identify the growth-preceding tipping point, as it was proposed in the previous part of this chapter. First, I employed the logic proposed by Phillips (Phillips, 2007), saying that the tipping point happens when the cumulative number of adopters exceeds the *p/q* ratio, i.e., $F(t\text{-}1) > p/q$, but the curve-fitting procedure demonstrated that the empirical cases do not follow it: the x-coordinate of the *p/q* ratio

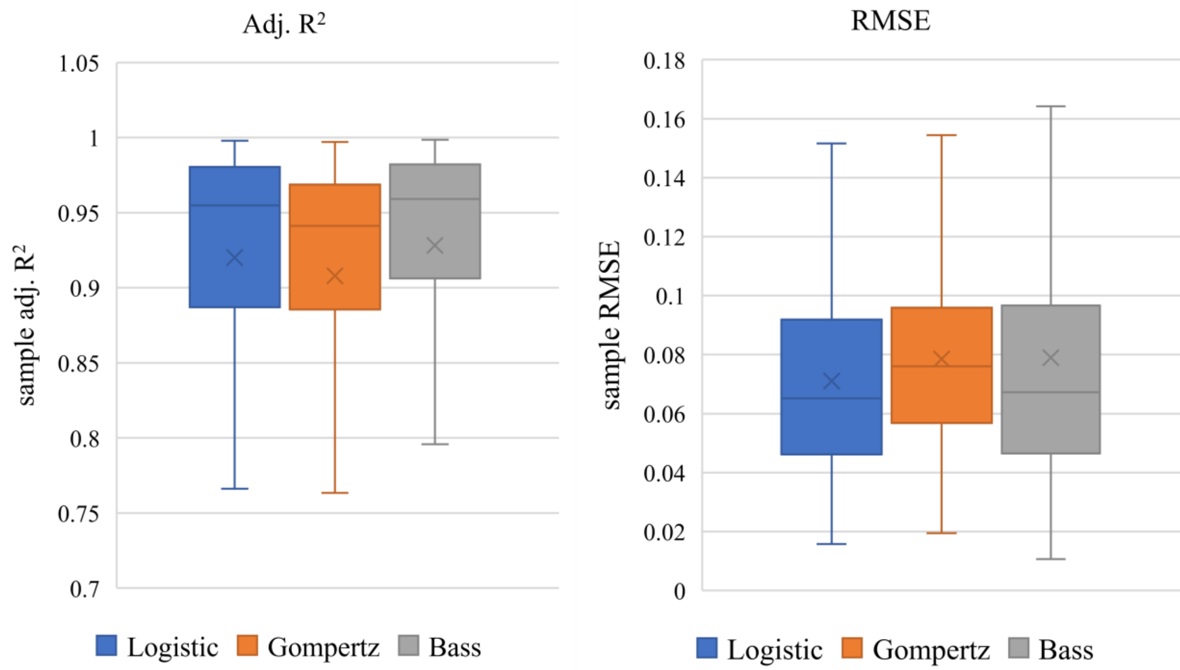


Figure 4.8. Distribution of Adj. R^2 and RMSE measures across the sample for three S-curve models, logistic, Gompertz, and Bass

point of the cumulative curve tends to be close to zero in the majority of cases that do not provide any valuable insights since the beginning of the curve cannot divide it into the pre- and post-beginning parts. However, I tested whether the event of reaching the level of *q* can precede the start of the accelerated growth and, after visual observation, confirmed it. All cases demonstrated that when the Bass model, fitted to the particular company's GT data, reaches the level of *q*, the curve starts to grow exponentially that is making the *q-point* a meaningful TBNV's tipping point. I also applied the formulas proposed by Orbach (Orbach, 2016) to identify the peak and inflection points (Phillips et al., 2016) of the TBNVs growth curves. After all, for each sample TBNV, I resulted with its Bass curve, containing *p*, *q*, and *m* parameters, the related tipping *q-point*, inflection point, and peak (Fig. 4.9-4.20).

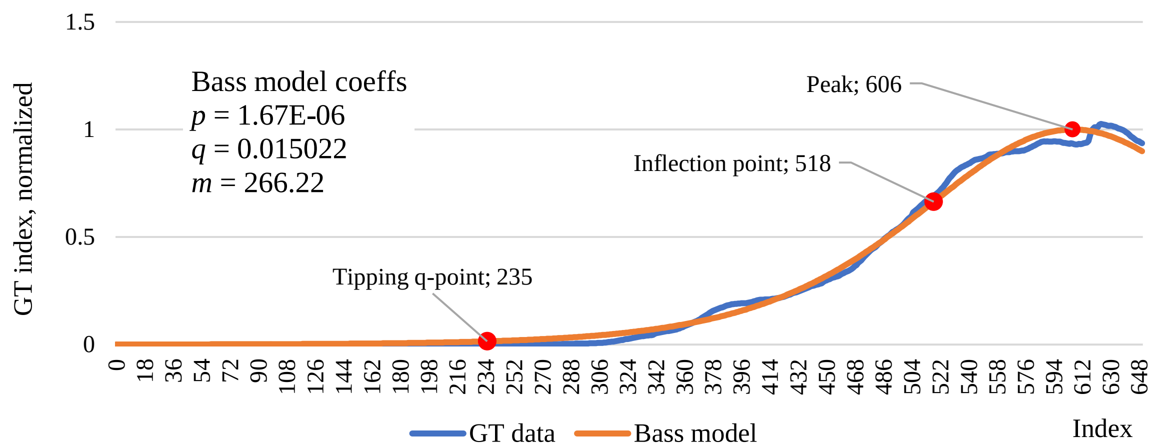


Figure 4.9. *Lyft* GT data, the related Bass model, and the critical points

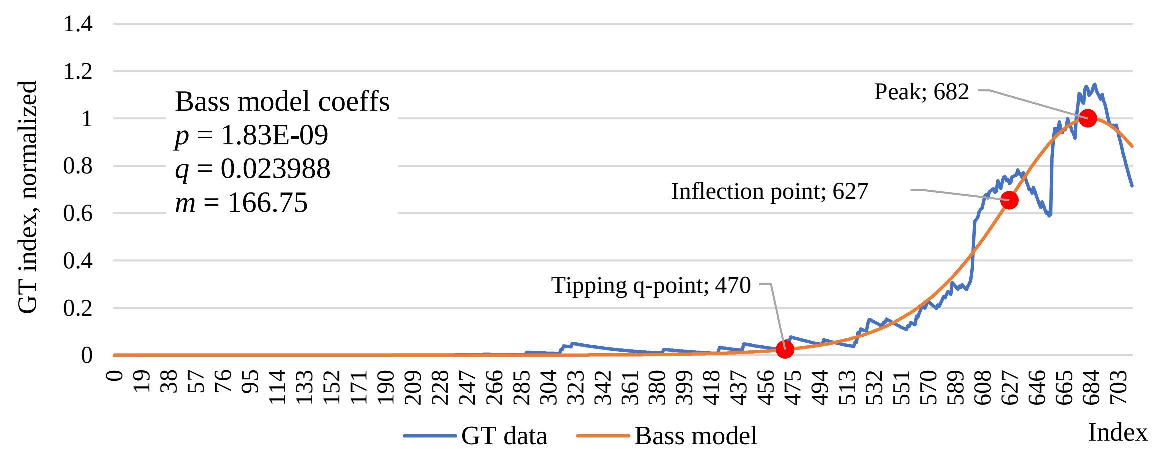


Figure 4.10. *Allakos* GT data, the related Bass model, and the critical points

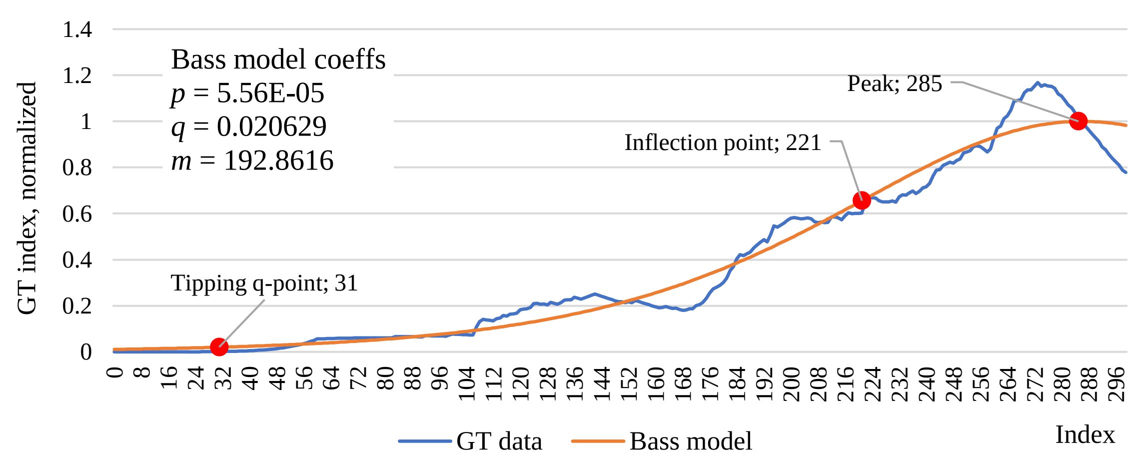


Figure 4.11. *Yerdle* GT data, the related Bass model, and the critical points

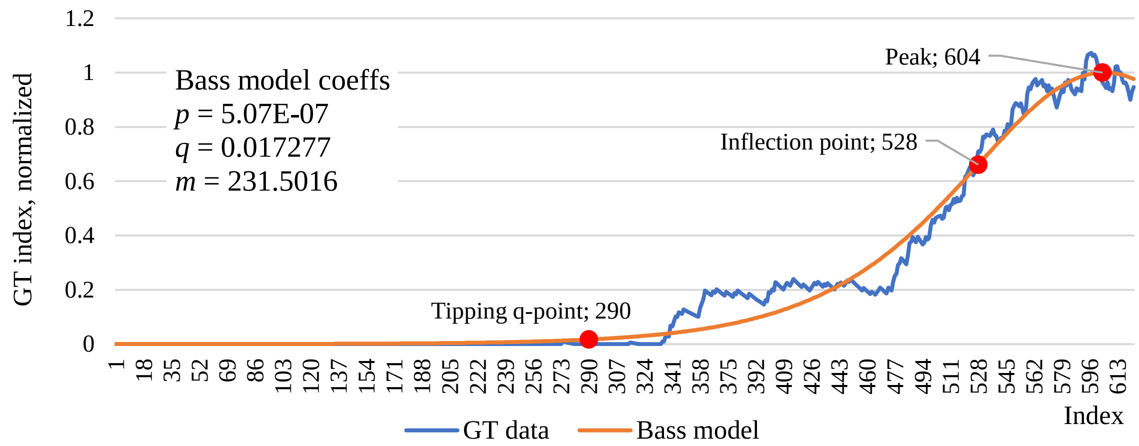


Figure 4.12. *FloQast* GT data, the related Bass model, and the critical points

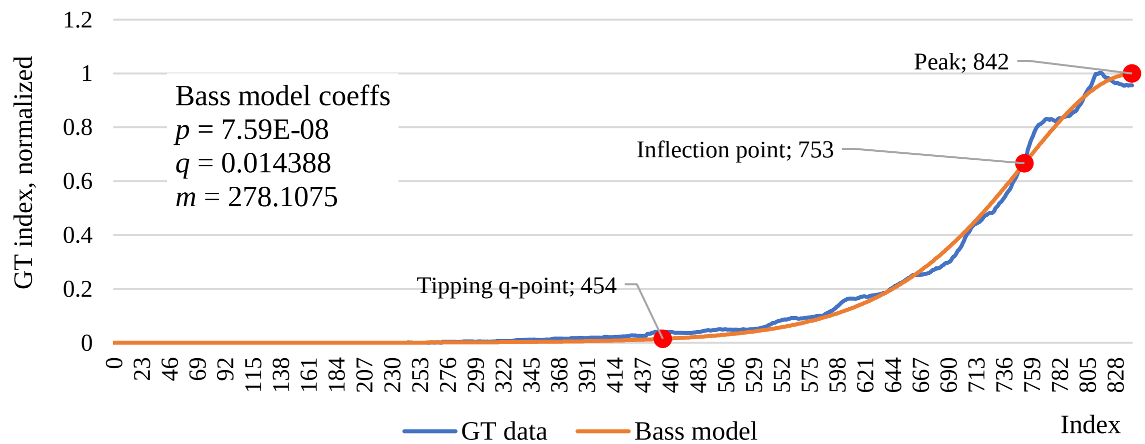


Figure 4.13. *Turo* GT data, the related Bass model, and the critical points

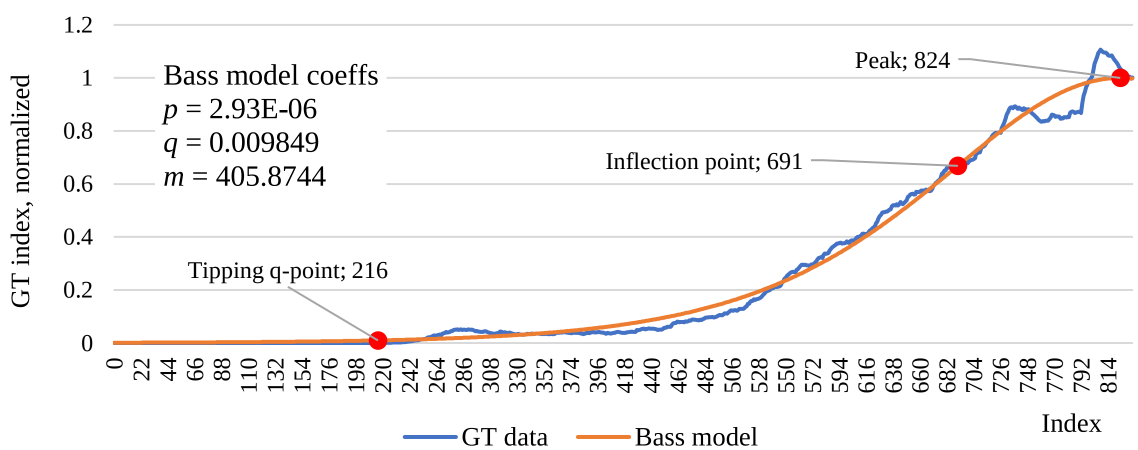


Figure 4.14. *Headspace* GT data, the related Bass model, and the critical points

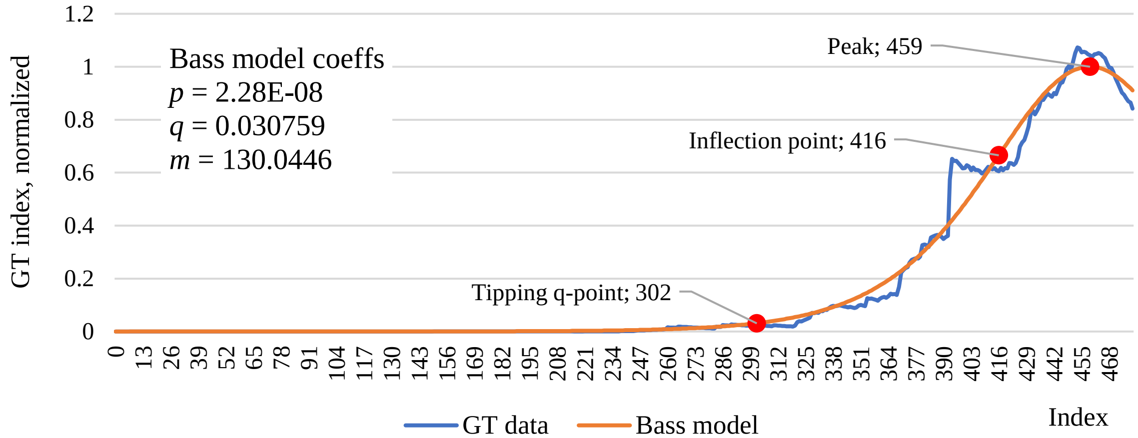


Figure 4.15. *Happify* GT data, the related Bass model, and the critical points

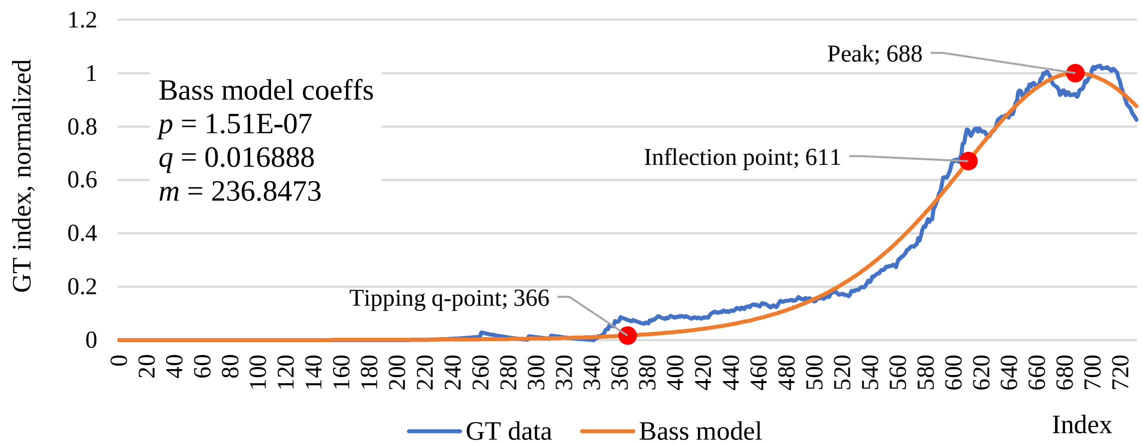


Figure 4.16. *Ygrene Energy Fund* GT data, the related Bass model, and the critical points

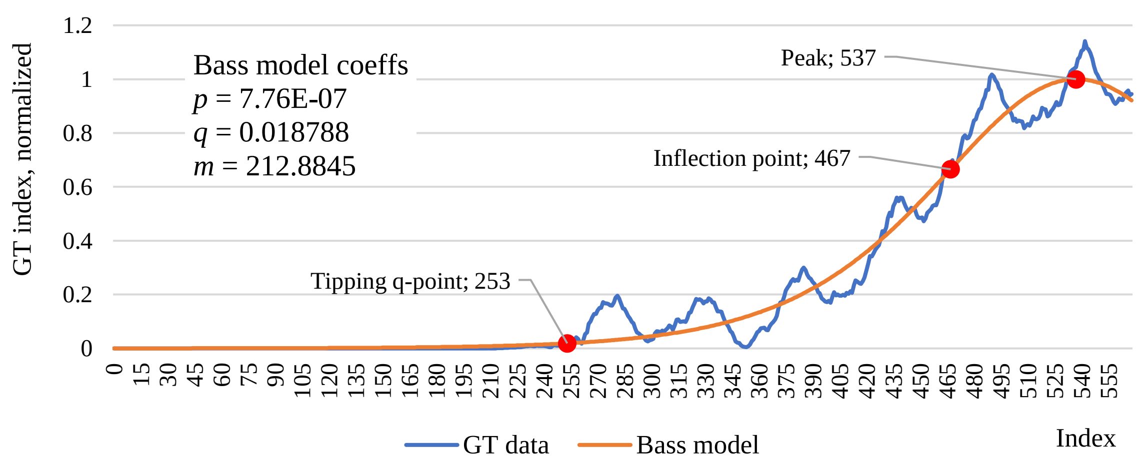


Figure 4.17. *Lever* GT data, the related Bass model, and the critical points

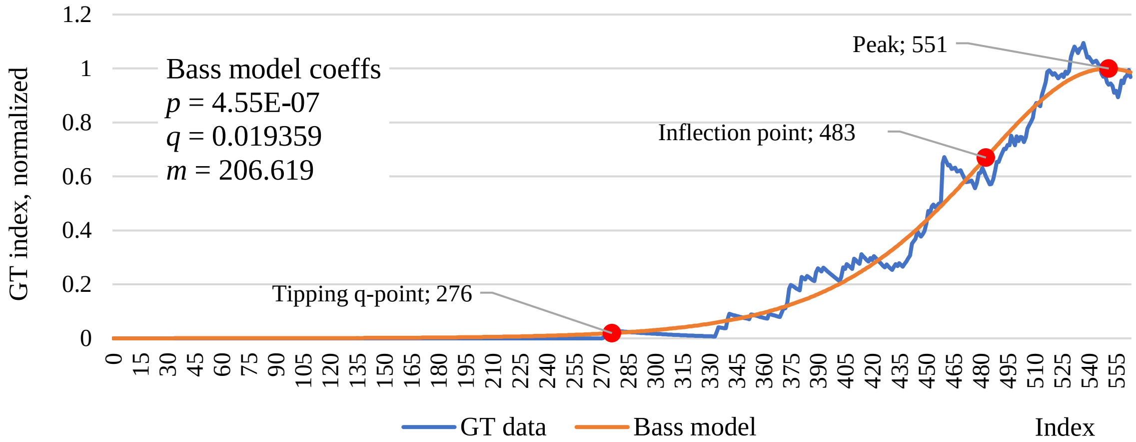


Figure 4.18. *Quanergy Systems* GT data, the related Bass model, and the critical points

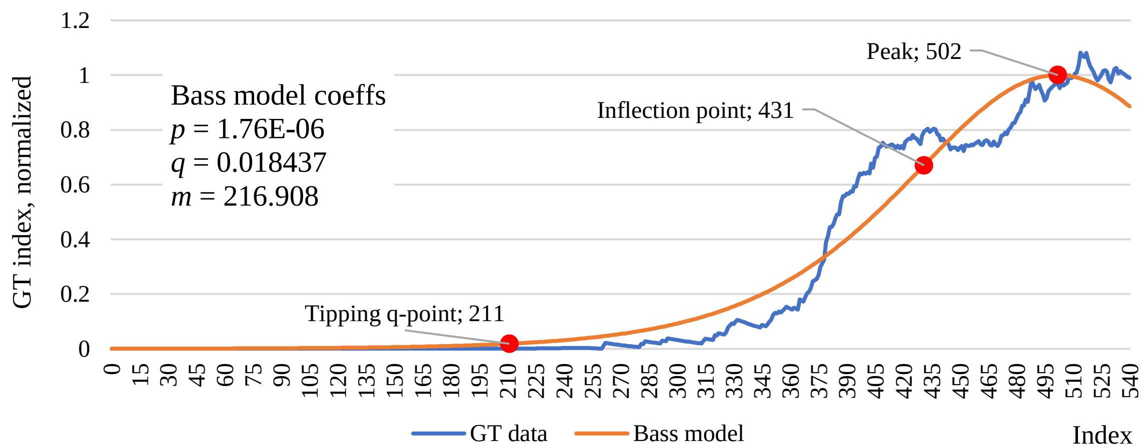


Figure 4.19. *LendingHome* GT data, the related Bass model, and the critical points

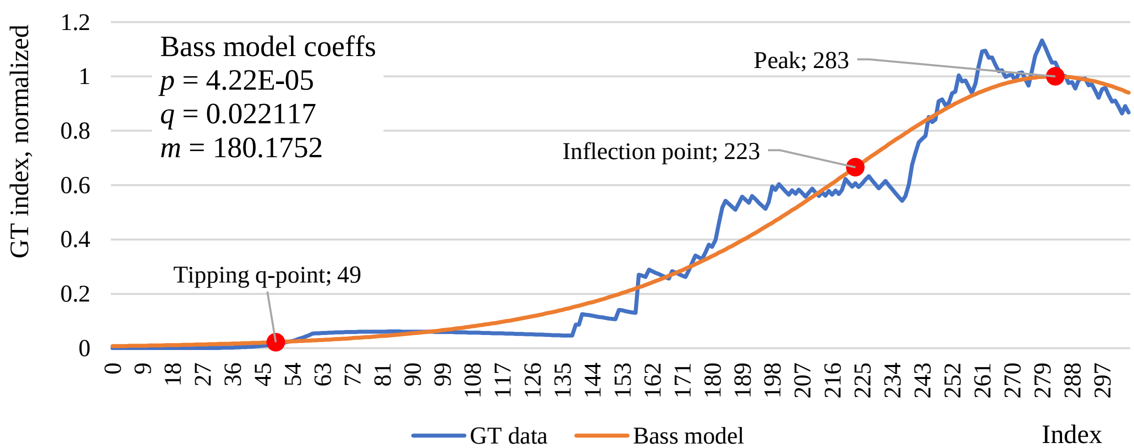


Figure 4.20. *Magenta Therapeutics* GT data, the related Bass model, and the critical points

In some cases (i.e., 87 or 35%), I evidenced q-points with x-coordinate equal to zero, similarly to the x-coordinates of the *p/q* points. However, in contrast to the *p/q* logic, these cases presented correct coordinates of the tipping points: visually, the related to them curves started to grow exponentially faster than they should according to TBNVs evolution principles, i.e., years before companies' foundation dates. That is to note, the zero-equal x-coordinates of these points are not necessarily the *true* coordinates of the q-points but the closest to them and available on the curve. In other words, the *true* q-points in these cases happened earlier than the taken beginnings of the related GT time series (Fig. 4.21). Despite the fact that visually these q-points are located quite correctly just at the beginning of the accelerated growth, this phenomenon raises some questions and leads to possible, mostly speculative, explanations. For instance, does it mean that some TBNVs have begun to "experience" an accelerated growth far before their foundation dates? If yes, which theoretical or practical concept or particular TBNVs features can explain it? Maybe some areas of technology or business types are so "hype" that just starting there brings immediate and fast growth? Or, maybe, the more "painful" the solved by a TBNV problem for the market is, the faster it will start to grow exponentially? I aim to examine these questions in future studies.

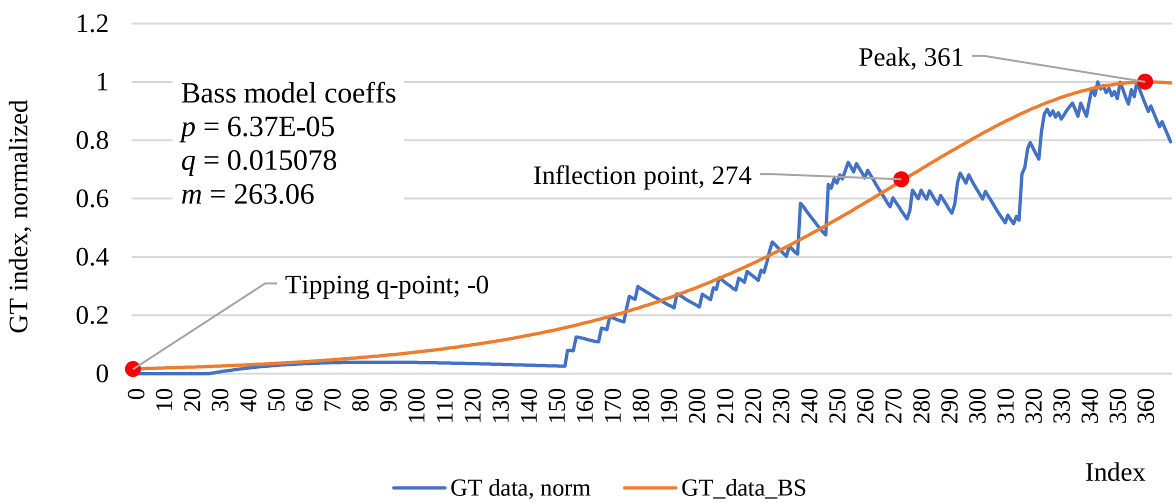


Figure 4.21. *Chorus.ai* GT data, the related Bass model, and critical points. Tipping q-point y-value is smaller than in the point (0, 0.01675318); thus, it is located in 0 point.

Nevertheless, from the innovation/imitation paradigm perspective, the fact that the accelerated growth starts just after the Bass curve reaches the level of *q*, which itself depicts the influence of imitators on the new product diffusion process (Bass, 1969; Phillips, 2007), probably leads to a conclusion that imitators play the key role in the launching the fast growth of a TBNV. I can explain it from the notion of "viral marketing" (Leskovec et al., 2007): according to the results, a company starts to experience accelerated growth only when it reaches a particular level of sales "virality" reflected by the imitators coefficient. Otherwise, according to the results, the growth will be moderate and close to linear rather than to the exponential form. Since, by their definition, TBNVs are entities, which are able to demonstrate high growth potential, and since VCs are exclusively focused on this growth (Zider, 1998), it can be concluded that a company, which is not able to launch the accelerated growth, may be treated as not perspective for VC investment and "unsuccessful" in the long term. Therefore, I can infer that one of the main tasks of a TBNV in the startup stage is to reach the level of *q* in their projected adoption curve that seems to be similar to what is called in practice *product/market fit* (Andreesen, 2007; Göthensten and Hellström, 2017). To describe this event, practitioners often apply to the particular subjective states of a product marketing process, describing it as when "the customer pulls the product out of their hands" (Rachleff, 2019), your company "in a good market with a product that can satisfy that market" (Andreesen, 2007), or "you have found a group of customers and a market that reacts positively to your product: you solve a problem, and you get paid for it" (Cooper and Vlaskovits, 2010). From the methodological point of view (and especially when you are an external observer), these states are quite hard, if possible, to formalize and, thus, to analyze. In other words, it is not possible to make any firm conclusions on whether a company reached the product/market fit and, if yes when it did it. Under this perspective, the q-point proposed in the current research has a great potential to provide a methodological and theoretical backing for the product/market fit. Just similarly to how a product/market fit understood, the q-point defines the beginning of the accelerated growth but, in contrast to it, can be precisely identified and located in the TBNV's growth curve. Therefore, it can be used in practical applications and, hence, solve the existing ambiguity of terms while providing a hands-on instrument with potential to control the growth speed.

To summarize, in Chapter 4, I have demonstrated that the Bass diffusion model provides the comparable (and by some measures – better) explanatory power for TBNVs growth data taken from Google Trends than the simple sigmoid curves. The more important is that, with the q-point, it also has the power to identify the predecessor of the accelerated growth, i.e., the tipping point, which is known to be the "holy grail" of new products diffusion models and, with a different name of product/market fit, playing the crucial role in the practical startup-management frameworks. The Bass model can be built for almost any TBNV that has high-quality GT data and, with the help of other critical points (i.e., inflection point and peak), can be precisely divided into three parts: no/moderate growth, accelerated growth, and deaccelerated growth. Altogether, these qualities lead us to propose the new product diffusion Bass model as the model describing TBNVs growth trajectory and the q-point as the tipping point or product/market fit, reaching of which makes it possible to launch an exponentially accelerated growth. Thus, I develop two propositions, which aim to test further:

<u>Proposition 1</u>

The q-point, identified by the Bass model calculated for a particular TBNV and considered under the innovation/imitation paradigm, is the most valid (so far) solution for defining the exponentially accelerated growth starting point on this TBNV's growth trajectory.

<u>Proposition 2</u>

The Bass model is the most beneficial (so far) analytical representation of a particular TBNV growth trajectory due to its high fitting quality and ability to provide the precise phase division of this TBNV's growth trajectory.

Chapter 5. Discussion

In general, the results of the current research demonstrate that (1) Google Trends is a valid source of information able to provide growth dynamics data for the majority of TBNVs (H1 is supported); (2) in the majority of cases, TBNVs growth dynamics follow an S-curve pattern that is not driven by the particular TBNVs' features or combination of them (H6 and H7 are supported). In the response to the question on *which S-curve model(s) is able to provide the most fruitful description of the TBNVs growth trajectory*, I propose the Bass new product diffusion model as the most beneficial so far. It has a comparable to the simple S-curve models quality of fit (and by some measures - higher) and has a built-in mechanism of identifying the tipping point, i.e., the q-point, which is known to be crucial for a new venture evolution (Gladwell, 2000; Phelps et al., 2007; Phillips, 2007).

In Chapter 3, it was demonstrated that Google Trends data have a strong connection to TBNVs valuations, which, in their turn, are known to be related to the companies' sales (Gompers et al., 2016; Köhn, 2018) or sales-related measures, e.g., daily/weekly/monthly active users, the number of downloads or subscriptions (Morwitz et al., 2007; Oberholzer-Gee and Strumpf, 2016). Compared to the previous empirical OLC studies, which investigated the relationship between companies' stages and various abstract state indicators (Hanks et al., 1994; Kazanjian and Drazin, 1990; Lester et al., 2003; Quinn and Cameron, 1983), GT data seems to be more advantageous. First, it is more transparent than the combination of features taken as particular states: while states may highly vary across the sample companies due to the unseen external or internal influence, Google Trends data have the same meaning for all TBNVs, i.e., its popularity in the public domain (Jun et al., 2018). And since it is connected to the public interest level, which may be more important for some TBNVs (e.g., social networks), it is even more beneficial than taken alone sales dynamics that are often analyzed as the growth measure of mature companies (Köhn, 2018; Miloud et al., 2012). Secondly, the higher transparency and simpler underlying meaning will lead to greater trustworthiness from the practitioners' side. I can hardly believe that a new venture founder will ignore the recommendation of strengthening the internal company's structure (for example, the valid recommendations need an additional study) after approaching the accelerated growth of public popularity. Therefore, the overall practical utility of the OLC theory will increase. Thirdly, a very few of the observed OLC empirical studies cover the early stages of a TBNV evolution (i.e., up to 5 years of operations), which are known to be the most critical in a new ventures' lifecycle since 86% of them fail exactly at this stage (Cantamessa et al., 2018). At the same time, related to a company, Google Trends data can be observable from its official foundation date and even earlier. So far, only by GT it is possible to retrospectively examine new venture's dynamics while it was mostly hidden for the market analysists, VCs, and competitors and derive specific factors, which influenced its further evolution. Finally, Google Trends data combine the best qualities of both empirical approaches applied in OLC research, longitudinal and cross-section, and overcome their frequently mentioned limitations. Indeed, each TBNV founded after January 1, 2004, which is identified by Google Trends algorithms and, thus, having a high-quality GT data, can be tracked from the very beginning to the date of its maximum growth point in the minute scale. For the oldest ones this timeframe can cover at least 18 years that can hardly be achievable with the common longitudinal approach (Abetti, 2001; Kazanjian and Drazin, 1989). In the similar vein, with the help of GT, it is possible to build various and wide cross-sectional samples of TBNVs, which will demonstrate the presence or absence of a studied effect(s) across different industries/locations and/or at specifically selected point(s) in time. Overall, Google Trends is a source of data, which can be seen as "an ideal setting" for empirical OLC research, as "in an ideal setting, one would study the OLC with a time-series sample, as a temporal component is inherent in the phenomenon" (Rutherford et al., 2003, p. 332).

The results of Chapter 4 go in line and, hence, empirically support the previous conceptual proposals of the TBNVs growth trajectories, e.g., (Picken, 2017; Söderling, 1998), while enhancing them in terms of the strong data-driven backing and straight-forward mathematical representation. In other words, I obtain reliable evidence that most TBNVs grow by the S-curves and propose a clear mathematical model to describe and apply this growth pattern. Of course, it cannot be claimed that the Bass model is the *only* model which has the power to describe companies' evolution trajectories. On the contrary, I believe that other models with comparable accuracy and benefits can be found or created further. For instance, the current literature suggests that the *Shifted Gompertz model* taken under some specific conditions and in particular contexts may be more beneficial than the Bass one (Bauckhage and Kersting, 2016; Bemmaor and Zheng, 2018). I aim to examine this question in future studies. Next, the Bass model provides a clear analytical algorithm for identifying the curve's critical points, i.e., tipping q-point, inflection, and peak points, which break the curve into three parts. If to bridge these parts to the existing OLC theory, the first phase (before the q-point) may relate to the "startup" or "inception," the second (between the q-point and inflection) to the "scaleup" or "growth," and the third (between the inflection and peak) to the "exit" phase or "maturity" (Miller and Friesen, 1984; Picken, 2017; Scott and Bruce, 1987). Hence, by utilizing the proposed mathematical model, it is possible to precisely identify when one phase ends and another begins and unambiguously relate the proposed OLC theory-specific states of a TBNV to each period of its analytical growth curve (Fig. 5.1). Some studies (e.g., Adizes, 1979; Greiner, 1972; Hanks et al., 1994; Miller and Friesen, 1984) claim that TBNVs have more growth stages, but this position is not supported by the proposed mathematical model and data and, therefore, so far cannot be reliably identified. Additionally, it cannot be ignored that the Bass distribution has the "bell" shape, i.e., a declining part after reaching the maximum growth level (Bass, 1969). Despite the fact that the after-maximum period lies aside the current study, I noticed the same behavior of some TBNVs' uncropped GT data that also corresponds to the majority of the previous OLC concepts (e.g., Kazanjian and Drazin, 1990; Lippitt and Schmidt, 1967; Quinn and Cameron, 1983).

Conception and development + Commercialization	Growth	Stability	*Kazanjian, 1988*
Birth	Growth	Maturity	*Miller & Friesen, 1984*
Inception + Survival	Growth + Expansion	Maturity	*Scott & Bruce, 1987*
Existence	Survival	Success	*Lester & Parnell, 2008*
Startup + Transition	Scaling	Exit	*Picken, 2017*
Formative	Normative	Integrative	*Söderling, 1998*
Seed stage + Early stage	Expansion (Series A, B)	Later stage (Series C, D, ...)	*Gompers, 1995*

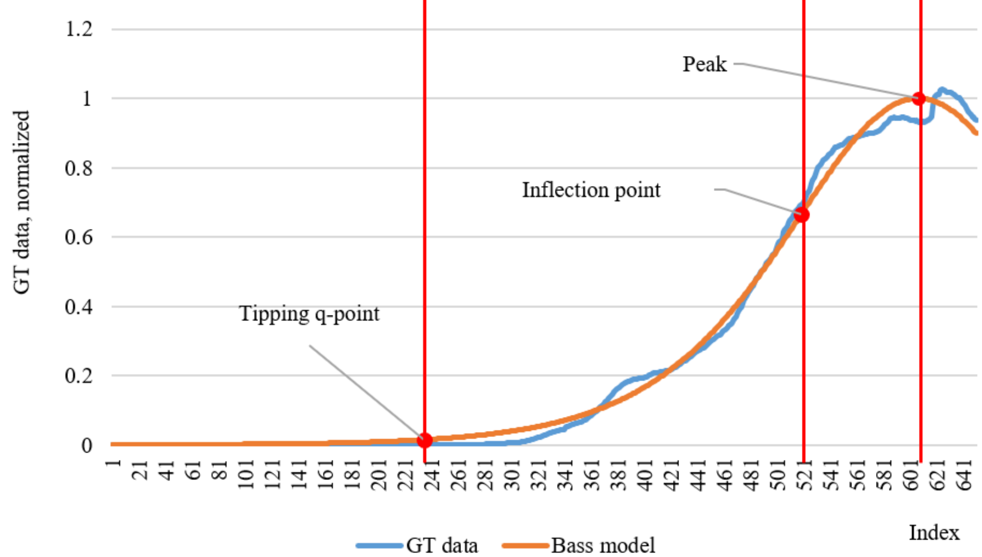


Figure 5.1. An example of bridging the empirical growth stages to some conceptual OLC models

However, is the OLC theory the most suitable organizational development theory for backing the proposed TBNVs' growth model? I assume it is. This conceptual basis was selected due to the core similarities between OLC and the analyzed process of TBNVs growth. Since, as it was described earlier in the thesis, venture capital seeks for TBNVs with high growth potential, it ultimately focuses on the growth, while investing in serial principle, when each series reflects a particular state of a company's evolution. Although the OLC theory has particular limitations, which are argued to be critical for the actual development of TBNVs, the proposed data-driven growth model may be considered from the position that accounts for these limitations. For instance, one of the key hypotheses of the organizational lifecycle concepts is the predetermined non-chaotic character of company development (Mosca et al., 2021), which is to some extent supported by the taken logic of the VC investment and the obtained model. Indeed, under the TBNVs' growth-driven evolution perspective, the process of movement from zero to the maximum possible company valuation looks predetermined. Also, the serial character of investment implies the similarity of conditions and states needed to be reached by a TBNV to move to the next investment round. However, at the same time, each company may need to take specific actions related to its business model and series of funding and demonstrate the case-dependent results to overcome the barriers for further investment rounds. Another core limitation of the OLC theory is related to the fact that the boundaries of an organization are understood to be clear and given (Van De Ven and Poole, 1995), while the actual studies argue that the boundaries are blurred due to the influence of ecosystems and a more open global market (He et al., 2020; Santos and Eisenhardt, 2005). Applying again to the adopted logic, it can be inferred that the proposed growth model is not influenced by one or another position in this question. Since VC valuations (mirrored by TBNVs' search query data) reflect a cumulative assessment of companies' actual status, including sales, team, partnerships, network, and others (Davila et al., 2003; Gompers et al., 2016; Hoenig and Henkel, 2015), it can be concluded that any – given or blurred – boundaries may be present in a particular TBNV and related to its specific stage of evolution. While the impact of one or another TBNV's feature on its valuation is managed, these features may take various values, sometimes form unique configurations, and serve for the predetermined move to the next investment round. In other words, from the macro perspective, the evolution of a TBNV under the context of VC investment is not chaotic and has particular predefined stages as proposed by OLC, while on the micro level, it may follow its own to some extent chaotic and reactive path taking various and most suitable configurations of features (Mosca et al., 2021). Therefore, it can be concluded that the OLC theory has high explanatory power for the actual research, at least at the macro level of the company growth process.

Chapter 6. Contributions to theory

From my perspective, the achieved results have five major contributions. First, they bring value for retrospective (i.e., *post-hoc*) analysis of TBNVs. Since the dates of the transition points (x-coordinates) can be precisely identified and located on the evolution curve, they can be directly related to the particular preceding events in companies' histories. This analysis can help to understand better the premises of specific decisions, relate them to the growth dynamics, and assess their effects by measuring the change in the rate of growth. The particular TBNV's states (or *gestalts*), which combine the conditions of various features and are often considered as the proxies for identifying the lifecycle stages in the empirical OLC studies (Al-Taie and Cater-Steel, 2020; Gupta and Chin, 1993; Hanks et al., 1994; Kazanjian and Drazin, 1990, 1989; Lester et al., 2008; Miller and Friesen, 1984; Quinn and Cameron, 1983; Rutherford et al., 2003), are now can be accurately related to the phases of a company's growth curve. Since the proposed source of data (i.e., Google Trends) can be applied to almost any developed venture, it is possible to derive statistically significant conclusions on these states and their variability across industries, geographies, and market segments. I am sure that these new opportunities are able to crucially advance the existing OLC concepts.

Second, each TBNV of interest will have its own distinct model's coefficients (i.e., *p*, *q*, and *m*), which themselves can be of particular interest for researchers. For instance, coefficients' variation can be used to carry out between companies' cluster comparisons in order to identify potentially more successful practices, features, or combinations of features under diverse contexts. The rigid mathematical model provides the valuable ability to directly compare growth dynamics of completely different TBNVs by moving their lifecycles under one transparent denominator.

Third, as it was mentioned earlier, the proposed tipping q-point may provide a solid theoretical backing for an ambiguous practical term of the *product/market fit* (Andreesen, 2007; Göthensten and Hellström, 2017). Since its invention in the mid-2000s by a number of US VCs, no study or book have offered a straight-forward and methodologically valid way of its identification; only subjective and blurred descriptions (Andreesen, 2007; Cooper and Vlaskovits, 2010; Rachleff, 2019), which are hard to formalize and utilize for scientific research or a transparent framework. Saying differently, outside observers (and, probably, inside as well) cannot be completely sure whether the company of interest has reached the product/market fit or not. Taking to account the high importance of this event for the VC practitioners[41], the task to propose the objective mathematical backing to it becomes even more extensive. There are some anecdotal cases where new ventures' founders were able to precisely point this event during the evolution of their companies[42], and their definitions look quite similar to what is tipping q-point defines: the point in time, which precedes the beginning of the accelerated growth. Therefore, it can be inferred that the proposed model bridges the two, academic and practitioners, worlds by providing the precise theoretically-backed instrument for the identification of the product/market fit by the q-point.

Fourth, the proposed source of data (i.e., Google Trends) eliminates the limitations, specific for the questionnaire-based (in other words, almost for all) OLC empirical studies focused on TBNVs. Since the related to TBNVs GT data exist independently from the willingness of the studied subjects and cannot be artificially manipulated, it avoids self-reporting, one-person, low response rate, and other biases, which are typically present in the majority of empirical studies, either longitudinal or cross-section (Garnsey et al., 2006; Levie and Lichtenstein, 2010). Due to this quality, GT data can be called a source of *objective* TBNVs information and, therefore, the valid scientific instrument for the organizational lifecycle field of research.

---

[41] What is Product/Market Fit and why is it so important? [WWW Document], n.d. URL https://bcombinator.com/what-is-product-market-fit-and-why-is-it-so-important (accessed 9.24.21).

[42] David Rusenko - How To Find Product Market Fit [WWW Document], n.d. URL https://www.youtube.com/watch?v=0LNQxT9LvM0 (accessed 12.31.18).

Finally, by proposing and describing the model of TBNVs growth, I solve an existing ambiguity, which occurred in the field of the organizational lifecycle theory. Since my model is quite straightforward in its description and rather simple in utilization, it can become a mathematical “basis” able to be applied to all existing OLC concepts and, thus, aligning stages of growth proposed by both theoretical and empirical works. From my perspective, this contribution will help scientists to validate some of the existing OLC concepts and, maybe, create new ones but, this time, with a direct link to data, thus, extending the applicability of the OLC theory that is seen as the key drawback of this area of knowledge (Levie and Lichtenstein, 2010). Moreover, the OLC theory may not be the only one to benefit from the achieved results. I assume that the alternative theories, providing frameworks for explaining organizational development, may also be applicable. For instance, if to consider well-known concepts proposed by Van De Ven and Poole (1995), *teleological* understanding can also benefit from the results obtained in the current study. According to this theory, organizational development is driven by the actions taken by management to reach the goal or a particular state (Hayes, 2014), which, adopting to TBNVs, may mirror the aim of VCs to reach a maximum valuation of a portfolio company during each series of funding including the exit. To underline, I can assume that existing OLC concepts, as well as the other organizational development theories, both discussed here and not, can be reviewed from the proposed growth trajectory perspective and accordingly advanced.

Chapter 7. Implications for practice

The results of this study bring significant value to managerial practice as well. First, since the key OLC theory issue, for which the solution was proposed in the current research, is its evidenced inconsistency leading to the limited practical applicability (Garnsey et al., 2006; Levie and Lichtenstein, 2010), the main practical implication is the increase of the theory utility for the real-life applications. With the proposed mathematical model, the existing theoretical OLC concepts can be reviewed, bridged to the practical cases, and built into the existing or new management frameworks. The plethora of conceptual models created during the 70 years of the OLC theory evolution (Levie and Lichtenstein, 2010) can now bring direct value to practice, i.e., to new venture founders, venture capitalists, investment analysts, and more.

Second, TBNVs related GT data itself may bring a significant value to the deep analysis of the particular companies' growth dynamics. Since it is able to provide time-series information with various resolutions – from minute to months[43] – practitioners can utilize it for at least three scale-dependent tasks. Low-resolution historical data may be utilized for studying the general evolution trends and deriving broad conclusions, more recent high-resolution statistics can be applied for short-term predictions, so-called "predicting the present" (Choi and Varian, 2012), and, when combined with the proposed mathematical apparatus, for the long-term forecasting. Currently, all these tasks are hardly (if possible) solvable with the existing sources of TBNVs information but of great value for practitioners who are highly interested in developing the current assessments of the companies' future dynamics (Brealey et al., 2012; Miloud et al., 2012).

[43] Google Trends: Understanding the data [WWW Document], n.d. Google News Initiat. URL https://newsinitiative.withgoogle.com/training/lesson/4876819719258112?image=trends&tool=Google%20Trends (accessed 7.8.20).

Finally, the proposed source of TBNVs data and the describing it mathematical methodology can be employed for building the direct valuation assessments of the companies of interest during the investment decision-making (Appendix D). Since it was demonstrated in the first study that GT data has a strong connection with TBNVs' valuations, the mathematical model built on it basically reflects the valuation dynamics as well. Since in this study I was focused on the growth period only, the post-maximum period of the company's evolution still requires additional research, but it can be assumed that the *valuation growth* model can be reliably rebuilt from the TBNV's GT data. Moreover, bearing in mind that Google Trends search query statistics is possible to obtain for the majority of the active ventures, I suggest it can be used in building the valuation models for the complete market segments, which further can be utilized to assess the valuations of the no-valuation-data companies.

Chapter 8. Limitations

As with every study, there are some limitations. Up to now, the proposed data source and utilized model are applicable only to *high-potential TBNVs* located in the *United States*, which build and sell their products *applying* new technologies (and *do not sell* the new technologies themselves through the licensing or another intellectual-property-based mechanism), and were *backed by VCs* for at least *three* times. This limitation happened naturally due to the research focus and methodological constraints set to study data. One may be interested in testing the results for a wider sample in the future. I can assume that the research can be repeated for TBNVs not backed by VCs while taking a different valuation-dependent proxy for analysis. However, for non-TBNV companies (e.g., SMEs), the results should be treated very consciously since their growth pathway may significantly differ due to intrinsic qualities. The analysis can also be repeated for TBNVs from other geographies. Since Google is dominating search engine with 92.01% of the world's search engine market[44], I assume that the found link and identified model will work for other jurisdictions in a similar vein. This assumption can be tested in future studies if reliable sources of VC valuations data are found for these specific countries or regions.

Second, to reach a considerable level of external validity, it was decided not to apply additional levels of clustering related, for example, to the variation in employed technologies and technologies' development lifecycles (e.g., Carden et al., 2010; Hoenen et al., 2014), levels of partnerships and affiliation to various technological and innovative ecosystems (e.g., Cavallo et al., 2019; Mason and Brown, 2014; Witt, 2004), geographies of target markets and market penetration goals (e.g., McDougall, 2003; McDougall et al., 1994; Spence et al., 2011), etc. This study presents the first step in the direction of employing TBNVs' related Google Trends data for analyzing their growth

---

[44] Search Engine Market Share Worldwide | Statcounter Global Stats [WWW Document], n.d. URL https://gs.statcounter.com/search-engine-market-share (accessed 3.1.22).

dynamics and, thus, targets to set the basis and provide instruments for further research. Therefore, the obtained results may not be applicable (or applicable in a lower degree) to TBNVs clustered by their particular features. In-depth applicability analysis of the proposed source of data and described methodology for such clusters may be implemented in future studies.

Third, TBNVs' related GT data and the proposed mathematical model so far can be applied mainly for implementing the retrospective analysis as its forecasting power was not examined in the current research. It was mentioned in plenty of the previous studies that Google Trends can be beneficially employed for forecasting and "nowcasting" of various economic (Niesert et al., 2020; Preis et al., 2013; Vosen and Schmidt, 2011), entrepreneurial (Carrière-Swallow and Labbé, 2013; Duwe et al., 2018), and social (Bangwayo-Skeete and Skeete, 2015; Bing et al., 2012; Polgreen et al., 2008) phenomena; however, in the current research, this question was not considered due to the stated goals. Nevertheless, based on the previous knowledge on building forecasts with S-curves (Bengisu and Nekhili, 2006; Easingwood et al., 1981; Kucharavy and De Guio, 2011; Meade, 1985; Modis, 2007), I can assume that the proposed model has a particular forecasting power. Supporting this assumption, fast tests of forecasting power resulted in a positive outcome: from a 1-year horizon and for several TBNVs from the sample on which tests were implemented, it was possible to forecast their maximum valuations with an error varying from 0.8% to 30% of the real maximum valuation. Making a forecasting horizon bigger resulted in a significant error increase. Therefore, additional research is required to assess the model's forecasting capability and describe the methodology for implementing such kind of analysis, maybe with the help of a machine and deep learning techniques. I plan to implement it in the near future.

Next, Google Trends provides the pre-processed rather than raw data that makes it impossible to control the built-in data manipulation algorithms and filters. Although I do not think that the access to the raw data would significantly influence the obtained results since I was interested in the main trends and applied noise-filtering procedures, the fully transparent data processing procedures may increase the accuracy of the outcomes or hint more research questions. Currently, I do not see how this limitation can be overcome without the assistance of *Google* but assume that there may be no need in doing that at all: according to the *Scopus* database statistics[7], more than 1,800 scientific studies successfully utilized GT as a research instrument in its actual form that, from my perspective, implies its high validity.

In addition, Google Trends may have limited power for some countries, which have their own dominating web search engines[45], i.e., China and North Korea with Baidu[46], Russia with Yandex[47], and South Korea with Naver[48]. Even though Google may still catch some proportions of search queries in locations without the full ban (i.e., Russia and South Korea), one cannot be completely sure that it correctly reflects main trends making the hypothetical research results reliable. To some extent, this limitation may be solved by adopting search query data from these competing sources (e.g., in the ongoing study, it was already demonstrated that search query data from Yandex Wordstat[49] can provide high-quality insights comparable to Google Trends and sometimes even better (Parfenov et al., 2022)); however, it will depend on the availability and quality of the similar to Google Trends tools and the related API.

Another limitation is connected to the fact that the maximums observed on the TBNVs' GT curves in Study 2 may be the local ones. To identify the companies, which have reached the maximum in their GT data and, thus, have finished their growth, I first applied the stage-related selection of the

---

[45] 4 Countries Where Google Doesn't Dominate [WWW Document], n.d. URL https://www.makeuseof.com/countries-google-doesnt-dominate/ (accessed 9.26.21).

[46] Click on Baidu and you will know [WWW Document], n.d. URL https://www.baidu.com/ (accessed 9.26.21).

[47] Yandex [WWW Document], n.d. URL https://yandex.com/ (accessed 9.26.21).

[48] NAVER [WWW Document], n.d. URL https://www.naver.com/ (accessed 9.26.21).

[49] Yandex Wordstat [WWW Document], n.d. URL https://wordstat.yandex.com/ (accessed 3.1.22).

Crunchbase database and then checked if the maximum followed by a steadily decline happened by 26 weeks (i.e., half a year) before the last available GT point. However, we cannot completely avoid the probability that the observed maximum is not local, and a TBNV of interest will not continue to grow further. Currently, two ways of overcoming this issue are seen. First, one may take the longer period after the maximum decline that, however, may significantly decrease the sample size since, for some TBNVs, this information simply will not be available due to their youth. Second, one can develop a procedure for identifying the global growth threshold connected to the overall market potential, the level of competition, or another related metric. Although this procedure may require additional research with its own limitations, it may significantly increase the reliability of the growth analysis results.

Finally, the proposed mechanism of the tipping q-point identification was validated only by visual observation since, to the best of my knowledge, there are no methods that would allow us to accurately identify the point preceding the accelerated growth on the S-curve. The only solution I can propose here so far is to compare the accuracy of the q-point identification with the tipping points in the well-known cases discussed in the book of Gladwell (Gladwell, 2000). Of course, this method will still be subjected to particular biases, but in the case of having no better alternatives, it can provide *at least some* results. I aim to work on this question in future research.

Chapter 9. Future research

Aside from the studies, which may evolve from the need to solve the mentioned limitations, I see at least nine topics that look promising for future research:

1. Analysis of the link between TBNVs' GT and valuation data after reaching the maximum valuation: in the current research, I was focused on the *growth* period of companies' evolution, and, thus, the question "what happens with the GT-valuation data link after the company reaches its maximum?" remains open. However, answering this question may shed light on the full TBNVs lifecycle, containing both growth and decline (or revival?) periods.

2. Examination of the causal relationship between TBNVs' GT and valuation data: the first study of this research was focused on *evidencing the link* between companies' GT and valuation data while ignoring the nature of this link. Does a valuation cause the change in GT data? Or do GT data drive the TBNVs' growth in valuation? Or, maybe, these two measures depend on another, so far unobserved, confounding variable? From my perspective, answers to these questions will advance our understanding of the found in the first study phenomenon.

3. Estimation of TBNVs' valuations from the segment-specific GT data and a few known valuation points: this topic is intended to increase the practical applicability of the proposed source of data and methods of its analysis. Coming from the facts that (1) TBNVs-related GT and valuation data have a strong positive correlation, and (2) valuations of various TBNVs can be directly compared, I aim to examine whether the GT data of the competitive TBNVs can be *directly compared* and applied for rebuilding the unknown companies' valuations (the first tryout in this direction is presented in Appendix D).

4. Test of the other new product diffusion models and their comparison to the Bass model: as it was previously noticed, I admit that the TBNVs' GT data can be described by *more than one*

new product diffusion model under the innovation/imitation paradigm (e.g., the Shifted Gompertz model). Therefore, I aim to aggregate, fit, and compare the results of the other possible models that will probably increase the proposed mathematical model's trustworthiness.

5. Review of the existing OLC conceptual frameworks through the prism of the proposed TBNVs growth model: going deeper into the task of aligning the existing empirical and conceptual OLC frameworks (Levie and Lichtenstein, 2010), I intend to aggregate the popular OLC growth models and consider them from the *proposed phase-division* perspective. The results of this study will be able to significantly increase the practical applicability of the existing OLC frameworks through their tight connection to the actual empirical data.

6. Comparison of the models’ shape-defining parameters p and q across the sample of TBNVs: according to the Bass model formula and the results achieved, each TBNV’s growth trajectory can be described by the interplay between the coefficient of innovation p and the coefficient of imitation q. I assume that these coefficients can be *directly compared* between various groups of TBNVs and, thus, provide the currently hidden group-dependent insights.

7. Analysis of the model’s forecasting power: in this proposed future study, the focus is on the *forecasting quality* of the Bass (or another new product diffusion) model. How accurate and long can the forecasting be built from the TBNVs’ GT data? Is it possible to reliably assess the future TBNVs’ valuation dynamics having only its web search query data and a few reference valuation points? Finding answers to these questions, from my point of view, will help practitioners to make more informed (and, thus, less risky) investment decisions.

8. Proposal and validation of the applied methodology for TBNVs growth analysis, description, and forecasting: so far, the described TBNVs growth research methodology utilizes advanced

data science methods, which may be *of hard use* for practical applications. Under this study, I aim to develop, validate, and offer the end-user-friendly framework or API for employing it in practical analytical scenarios. The proposal of such an instrument will heavily increase the practical utility of the proposed approach and enhance the quality of managerial decisions.

9. Examination of the startup accelerators' influence on the companies' growth dynamics: in this possible future research, I aim to examine whether the startup accelerators are *as good as they are thought to be*. Startup-accelerators gained significant attention from the venture and startup markets, and their most known representatives frequently report on a rise of "other several millions of dollars" by the company, which was accelerated during participation in their programs[50, 51, 52]. However, employing the TBNVs GT data and the proposed mathematical model, I would like to test to which extent the alumni's success is driven by their built-in qualities and by the help of an accelerator.

10. Studying the influence of TBNVs' features on their growth: this research targets to employ various companies' features (e.g., level of success, type of product, industry, customer segment, exit strategy, etc.) as independent variables for predicting the parameters of the related to them Bass growth curves (i.e., p, q, m, and location of the critical points) by machine learning techniques. As an outcome, it will be possible to identify how TBNVs' features influence their growth character.

[50] 2ndKitchen Completes $4.35M Seed Round so Places Like Bars Can Serve Food from Nearby Restaurants [WWW Document], n.d. URL https://thespoon.tech/2ndkitchen-completes-4-35m-seed-round-so-places-like-bars-can-serve-food-from-nearby-restaurants/ (accessed 9.26.21).

[51] Y Combinator-backed Karbon Card raises $12 million in latest funding round - The Economic Times [WWW Document], n.d. URL https://economictimes.indiatimes.com/tech/funding/y-combinator-backed-karbon-card-raises-12-million-in-latest-funding-round/articleshow/86453937.cms (accessed 9.26.21).

[52] Y Combinator-Backed Spenmo Closes US$34 Million Series A Fundraise | Fintech Singapore [WWW Document], n.d. URL https://fintechnews.sg/55624/payments/y-combinator-backed-spenmo-closes-us34-million-series-a-fundraise/ (accessed 9.26.21).

11. Examining the short-term forecasting power of TBNVs' related GT data: in this study, I plan to employ machine and deep learning techniques to assess the power of TBNVs' related GT data for building their short-term valuation projections. In fact, search query statistics are available in various scales (can be taken even every minute) and for the majority of TBNVs. Thus, I assume that these time series may be employed as a source of the training dataset for advanced data science techniques (like logistic regression, random forests, artificial neural networks, and others) with a goal to predict fast change in TBNVs' valuations.

Chapter 10. Conclusions

The organizational lifecycle theory has been developing for the last 70 years and, under some perspective, has moved from the “primitive” stage to “validated” (Tam and Gray, 2016). Dozens of OLC concepts have been offered; some gained more attention due to their “elegance”[53] (Adizes, 1979), some due to their intuitive simplicity[54] (Greiner, 1972), some because of the (thought-to-be) strong empirical evidence (Levie and Lichtenstein, 2010; Scott, 1973), and other due to the (probably) revolutionary setting (Ismail et al., 2014). However, according to the opinion of some researchers, “stages of growth modeling has hit a dead end” since “the management literature showed no consensus on basic constructs of the approach, and no empirical confirmation of stages theory” and resulting in a fact that “stages theory is not appropriate for understanding business growth” (Levie and Lichtenstein, 2010, pp. 318, 317, 329). Other researchers propose the less critical point of view, saying that “there is no basis for conceptual and empirical alignment between stage models” (Garnsey et al., 2006, p. 3). Anyhow, what is definitely clear is that the problem with the OLC theory exists; it is recognized, but so far is not solved.

In the current doctoral thesis, I went deeper into the understanding of this issue and tried to identify and solve the underlying questions, which led to the OLC theory stagnation. In particular, for the first step, I determined that without a detailed, comprehensive, and objective source of data, any empirical test will contain pitfalls similar to the previous works. Therefore, I proposed Google Trends – the big data instrument, which provided search query statistics on the particular term and was beneficially applied as a scientific instrument in more than 1,800 studies since its public launch in

---

[53] Adizes Ten Stages - Corporate Life Cycle Model [WWW Document], n.d. URL https://www.businessballs.com/organisational-culture/adizes-ten-stages-corporate-life-cycle-model/ (accessed 9.24.21).

[54] Guide To Greiner’s Growth Model | Lucidity [WWW Document], n.d. URL https://getlucidity.com/strategy-resources/guide-to-greiners-growth-model/ (accessed 9.24.21).

2006[7]. After the deep examination presented in Study 1, I concluded that TBNVs' related GT data have a high correlation with their valuation points and, thus, can be utilized as a reliable proxy for examining companies' growth dynamics.

For the second step, I questioned the mathematical representation of TBNVs' growth trajectory and came up with a conclusion that the S-curve model reflects the companies' evolution pattern in the best possible way compared to the other typically used growth curves. Bearing in mind the inherent limitations of the commonly used S-curve models (i.e., logistic and Gompertz models) driven by their intrinsic autocatalyticity, I employed the Bass new product diffusion model (Bass, 1969) and demonstrated its benefits for the task of TBNVs growth description. Specifically, while keeping the quality of fit on the high level, this model provides a precise analytical tool for three critical points identification, which together divide the growth curve into three parts: no/moderate growth, accelerated growth, and deaccelerated growth. Since this division is driven by a straightforward mathematical principle, which also has a clear theoretical backing, I assume that it can solve the existing ambiguity in the OLC theory and provide the basis for alignment of various empirical and theoretical concepts and, thus, increase their practical utility.

To underline, in this study, I identified two secondary research questions together, leading to the main one: *How do technology-based new ventures grow?* And the answer to this question is: *TBNVs grow by an <u>S-curve trajectory</u> with <u>one tipping q-point</u>, <u>one inflection</u>, and <u>one peak point</u>, which together divides the growth curve into <u>three general stages</u>.* I believe that this result will take the OLC theory from stagnation, breathe new life into it, and open a new, *data-driven* period of theory evolution.

Bibliography

Abetti, P.A., 2001. Accelerated growth: helping companies get and stay on the fast track. Int. J. Manuf. Technol. Manag. 3, 15. https://doi.org/10.1504/IJMTM.2001.001393

Achtenhagen, L., Naldi, L., Melin, L., 2010. “Business growth” - Do practitioners and scholars really talk about the same thing? Entrep. Theory Pract. https://doi.org/10.1111/j.1540-6520.2010.00376.x

Acs, Z.J., Armington, C., 2006. Entrepreneurship, Geography, and American Economic Growth, Entrepreneurship, Geography, and American Economic Growth. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511510816

Addison, E.L., Beazley, D.A., 2007. Life Cycle Challenges of Online Learning Support Systems, in: SAM International Business Conference. Las Vegas, Nevada, p. 14.

Adizes, I., 1979. Organizational passages-Diagnosing and treating lifecycle problems of organizations. Organ. Dyn. 8, 3–25. https://doi.org/10.1016/0090-2616(79)90001-9

Aidin, S., 2015. New Venture Creation: Controversial Perspectives and Theories. Econ. Anal. 48, 101–109.

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle, in: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Akoglu, H., 2018. User’s guide to correlation coefficients. Turkish J. Emerg. Med. 18, 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Akter, S., Wamba, S.F., Gunasekaran, A., Dubey, R., Childe, S.J., 2016. How to improve firm performance using big data analytics capability and business strategy alignment? Int. J. Prod.

Econ. 182, 113–131. https://doi.org/10.1016/j.ijpe.2016.08.018

Al-Taie, M., Cater-Steel, A., 2020. The Organisational Life Cycle Scale: An Empirical Validation. J. Entrep. 29, 293–325. https://doi.org/10.1177/0971355720930572

Aldrich, H.E., 1990. Using an ecological perspective to study organizational founding rates. Entrep. Theory Pract. 14, 7–24. https://doi.org/10.1177/104225879001400303

Allen, D.M., 1974. The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics 16, 125–127. https://doi.org/10.1080/00401706.1974.10489157

Allen, M., 2017. The SAGE encyclopedia of communication research methods. SAGE Publications, Inc., Thousand Oaks, California. https://doi.org/10.4135/9781483381411 NV - 4

Amir, E., Lev, B., 1996. Value-relevance of nonfinancial information: The wireless communications industry. J. Account. Econ. https://doi.org/10.1016/S0165-4101(96)00430-2

Andreesen, M., 2007. Product/Market Fit [WWW Document]. Stanford Univ. URL https://web.stanford.edu/class/ee204/ProductMarketFit.html

Antretter, T., Blohm, I., Grichnik, D., Wincent, J., 2019. Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy. J. Bus. Ventur. Insights 11, e00109. https://doi.org/10.1016/j.jbvi.2018.e00109

Armstrong, J.S., 2002. Principles of Forecasting. Kluwer Academic Publishers, Dordrecht, Netherlands.

Asadullah, A., Faik, I., Kankanhalli, A., 2018. Digital platforms: a review and future directions, in: PACIS 2018 Proceedings. Yokohama, Japan, pp. 1–14.

Assael, H., 1992. Consumer behavior and marketing action, 4th ed. PWS-KENT Pub.

Audretsch, D.B., Bönte, W., Mahagaonkar, P., 2012. Financial signaling by innovative nascent ventures: The relevance of patents and prototypes. Res. Policy 41, 1407–1421. https://doi.org/10.1016/j.respol.2012.02.003

Bangwayo-Skeete, P.F., Skeete, R.W., 2015. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tour. Manag. 46, 454–464. https://doi.org/10.1016/j.tourman.2014.07.014

Bass, F.M., 1969. A New Product Growth for Model Consumer Durables. Manage. Sci.

Bass, F.M., Krishnan, T. V., Jain, D.C., 1994. Why the Bass Model Fits without Decision Variables. Mark. Sci. https://doi.org/10.1287/mksc.13.3.203

Batistič, S., der Laken, P., 2019. History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations. Br. J. Manag. 30, 229–251. https://doi.org/10.1111/1467-8551.12340

Bauckhage, C., Kersting, K., 2016. Collective Attention on the Web. Found. Trends® Web Sci. 5, 1–136. https://doi.org/10.1561/1800000024

Baum, J.A.C., Silverman, B.S., 2004. Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. J. Bus. Ventur. 19, 411–436. https://doi.org/10.1016/S0883-9026(03)00038-7

Bemmaor, A.C., 1992. Modeling the Diffusion of New Durable Goods: Word-of-Mouth Effect Versus Consumer Heterogeneity, in: Research Traditions in Marketing. pp. 201–229. https://doi.org/10.1007/978-94-011-1402-8_6

Bemmaor, A.C., Zheng, L., 2018. The diffusion of mobile social networking: Further study. Int. J. Forecast. 34, 612–621. https://doi.org/10.1016/j.ijforecast.2018.04.006

Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technol. Forecast. Soc. Change 73, 835–844. https://doi.org/10.1016/j.techfore.2005.09.001

Bing, P., Doris Chenguang, W., Haiyan, S., 2012. Forecasting Hotel Room Demand Using Search Engine Data. J. Hosp. Tour. Technol. 3, 3.

Blair, E., Zinkhan, G.M., 2006. Nonresponse and generalizability in academic research. J. Acad. Mark. Sci. 34, 4–7. https://doi.org/10.1177/0092070305283778

Blank, S., 2013. Why the lean start-up changes everything. Harv. Bus. Rev.

Bocken, N.M.P., 2015. Sustainable venture capital - catalyst for sustainable start-up success? J. Clean. Prod. 108, 647–658. https://doi.org/10.1016/j.jclepro.2015.05.079

Bowman, E.H., Singh, H., Thomas, H., 2012. The Domain of Strategic Management: History and Evolution, in: Handbook of Strategy and Management. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, pp. 31–52. https://doi.org/10.4135/9781848608313.n2

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2008. Time Series Analysis, 4th ed, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey. https://doi.org/10.1002/9781118619193

Bradbury, D., 2011. Data mining with LinkedIn. Comput. Fraud Secur. https://doi.org/10.1016/S1361-3723(11)70101-4

Brdulak, A., Chaberek, G., Jagodziński, J., 2021. BASS Model Analysis in "Crossing the Chasm" in E-Cars Innovation Diffusion Scenarios. Energies 14, 3216. https://doi.org/10.3390/en14113216

Brealey, R.A., Myers, S.C., Allen, F., Mohanty, P., 2012. Principle of Corporate Finance, 10th ed.

Mcgraw-Hill Education.

Brockwell, P.J., Davis, R.A., 2016. Introduction to Time Series and Forecasting, Applied Physics Letters, Springer Texts in Statistics. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-29854-2

Campos, E.B., Somoza, M.L., Salmador, M.P., 2011. CoPs & organizational identity: five case studies of NTBFs, in: Handbook of Research on Communities of Practice for Organizational Management and Networking: Methodologies for Competitive Advantage. IGI Global, pp. 308–336. https://doi.org/10.4018/978-1-60566-802-4.ch018

Cantamessa, M., Gatteschi, V., Perboli, G., Rosano, M., 2018. Startups' roads to failure. Sustain. 10, 1–19. https://doi.org/10.3390/su10072346

Carden, C.W., Chamberlain, T., Hill, J.W., 2010. The brave new world of valuing life sciences and healthcare enterprises. Bus. Horiz. 53, 183–197. https://doi.org/10.1016/j.bushor.2009.11.004

Carrière-Swallow, Y., Labbé, F., 2013. Nowcasting with Google Trends in an Emerging Market. J. Forecast. 32, 289–298. https://doi.org/10.1002/for.1252

Carter, N.M., Gartner, W.B., Reynolds, P.D., 1996. Exploring start-up event sequences. J. Bus. Ventur. 11, 151–166. https://doi.org/10.1016/0883-9026(95)00129-8

Cavallo, A., Ghezzi, A., Dell'Era, C., Pellizzoni, E., 2019. Fostering digital entrepreneurship from startup to scaleup: the role of venture capital funds and angel groups. Technol. Forecast. Soc. Change 145, 24–35. https://doi.org/10.1016/j.techfore.2019.04.022

Chandler, G.N., DeTienne, D.R., McKelvie, A., Mumford, T. V., 2011. Causation and effectuation processes: A validation study. J. Bus. Ventur. 26, 375–390. https://doi.org/10.1016/j.jbusvent.2009.10.006

Chang, S.J., 2004. Venture capital financing, strategic alliances, and the initial public offerings of internet startups. J. Bus. Ventur. 19, 721–741. https://doi.org/10.1016/j.jbusvent.2003.03.002

Chang, Y.B., Kwon, Y.O., 2020. Attention-grabbing IPOs in early stages for IT firms: An empirical analysis of post-IPO performance. J. Bus. Res. 109, 111–119. https://doi.org/10.1016/j.jbusres.2019.10.061

Chatfield, C., 1975. The Analysis of Time Series: Theory and Practice, Climate Change 2013 - The Physical Science Basis. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-2925-9

Chen, J., Dunn, A., Hood, K., Driessen, A., Batch, A., 2019. Off to the Races : A Comparison of Machine Learning and Alternative Data for Nowcasting of Economic Indicators. Big Data 21st Century Econ. Stat.

Chen, Y., Xie, J., 2005. Third-party product review and firm marketing strategy. Mark. Sci. https://doi.org/10.1287/mksc.1040.0089

Choi, H., Varian, H., 2012. Predicting the present with Google Trends. Econ. Rec. 88, 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Christensen, C.M., 2011. The Innovator's Dilemma, The Innovator's Dilemma. https://doi.org/10.15358/9783800642816

Chumnumpan, P., Shi, X., 2019. Understanding new products' market performance using Google Trends. Australas. Mark. J. 27, 91–103. https://doi.org/10.1016/j.ausmj.2019.01.001

Churchill, N.C., Lewis, V.L., 1983. The Five Stages of Small Business Growth. Harv. Bus. Rev. 61, 30–50.

Coad, A., 2007. Firm Growth: A Survey. Pap. Econ. Evol.

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Routledge, New

York. https://doi.org/10.4324/9780203771587

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2000. Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not) (No. 7552), NBER Working Paper Series, Working Paper Series. New York. https://doi.org/10.3386/w7552

Colombo, M.G., Grilli, L., 2010. On growth drivers of high-tech start-ups: exploring the role of founders' human capital and venture capital. J. Bus. Ventur. 25, 610–626. https://doi.org/10.1016/j.jbusvent.2009.01.005

Connelly, B.L., Certo, S.T., Ireland, R.D., Reutzel, C.R., 2011. Signaling Theory: A Review and Assessment. J. Manage. 37, 39–67. https://doi.org/10.1177/0149206310388419

Cooper, B., Vlaskovits, P., 2010. The entrepreneur's guide to customer development : a "cheat sheet" to The four steps to the epiphany. https://doi.org/10.1017/CBO9781107415324.004

Courtney, C., Dutta, S., Li, Y., 2017. Resolving Information Asymmetry: Signaling, Endorsement, and Crowdfunding Success. Entrep. Theory Pract. https://doi.org/10.1111/etap.12267

Cover, T.M., Thomas, J.A., 2006. Elements of information theory, 2nd ed. Wiley-Interscience.

Cramer, J.S., 2002. The origins of logistic regression. Tinbergen Inst. Work. Pap. No. 2002-119/4. https://doi.org/http://dx.doi.org/10.2139/ssrn.360300

Cumming, D., 2008. Contracts and Exits in Venture Capital Finance. Rev. Financ. Stud. 21, 1947–1982. https://doi.org/10.1093/rfs/hhn072

Dahiya, S., Ray, K., 2012. Staged investments in entrepreneurial financing. J. Corp. Financ. 18, 1193–1216. https://doi.org/10.1016/j.jcorpfin.2012.07.002

Davila, A., Foster, G., Gupta, M., 2003. Venture capital financing and the growth of startup firms. J. Bus. Ventur. 18, 689–708. https://doi.org/10.1016/S0883-9026(02)00127-1

Delmar, F., Davidsson, P., 2000. Where do they come from? Prevalence and characteristics of nascent entrepreneurs. Entrep. Reg. Dev. 12, 1–23. https://doi.org/http://dx.doi.org/10.1080/089856200283063

Delmar, F., Davidsson, P., Gartner, W.B., 2003. Arriving at the high-growth firm. J. Bus. Ventur. https://doi.org/10.1016/S0883-9026(02)00080-0

DeTienne, D.R., 2010. Entrepreneurial exit as a critical component of the entrepreneurial process: Theoretical development. J. Bus. Ventur. 25, 203–215. https://doi.org/10.1016/j.jbusvent.2008.05.004

DeTienne, D.R., McKelvie, A., Chandler, G.N., 2015. Making sense of entrepreneurial exit strategies: A typology and test. J. Bus. Ventur. 30, 255–272. https://doi.org/10.1016/j.jbusvent.2014.07.007

Dowling, G.R., 1986. Perceived risk: the concept and its measurement. Psychol. Mark. 3, 193–210. https://doi.org/https://doi.org/10.1002/mar.4220030307

Dubey, R., Gunasekaran, A., Childe, S.J., Blome, C., Papadopoulos, T., 2019. Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. Br. J. Manag. 30, 341–361. https://doi.org/10.1111/1467-8551.12355

Dufour, Y., Steane, P., Corriveau, A.M., 2018. From the organizational life-cycle to “ecocycle”: a configurational approach to strategic thinking. Asia-Pacific J. Bus. Adm. 10, 171–183. https://doi.org/10.1108/APJBA-05-2018-0095

Duşa, A., 2018. QCA with R: a comprehensive resource. Springer International Publishing AG, Cham, Switzerland. https://doi.org/10.1007/9783319756684

Duwe, D., Herrmann, F., Spath, D., 2018. Forecasting the diffusion of product and technology innovations: Using google trends as an example. PICMET 2018 - Portl. Int. Conf. Manag. Eng.

Technol. Manag. Technol. Entrep. Engine Econ. Growth, Proc. 1–7. https://doi.org/10.23919/PICMET.2018.8481971

Dyson, R.G., Gregg, J. V., Hossell, C.H., Richardson, J.T., 1965. Mathematical Trend Curves-An Aid to Forecasting. Stat. https://doi.org/10.2307/2987396

Easingwood, C., Mahajan, V., Muller, E., 1981. A nonsymmetric responding logistic model for forecasting technological substitution. Technol. Forecast. Soc. Change 20, 199–213. https://doi.org/10.1016/0040-1625(81)90021-4

Eckenrode, J., Friedman, S., 2017. Fintech by the numbers. Deloitte Serv. LP 24.

Efrat, A., Fan, Q., Venkatasubramanian, S., 2007. Curve matching, time warping, and light fields: new algorithms for computing similarity between curves. J. Math. Imaging Vis. 27, 203–216. https://doi.org/10.1007/s10851-006-0647-0

Eisenhauer, J.G., 2009. Explanatory Power and Statistical Significance. Teach. Stat. 31, 42–46. https://doi.org/10.1111/j.1467-9639.2009.00364.x

Ellis, N., 2010. Business to business marketing: relationships, networks and strategies, 1st ed. Oxford University Press, Oxford.

Fan, J.S., 2016. Regulating unicorns: disclosure and the new private economy., Boston College Law Review. https://doi.org/10.2139/ssrn.2763650

Field, A., 2013. Discovering Statistics Using IBM SPSS Statistics, 4th ed. SAGE Publications, Inc.

Fieller, E.C., Hartley, H.O., Pearson, E.S., 1957. Tests for Rank Correlation Coefficients. I. Biometrika 44, 470. https://doi.org/10.2307/2332878

Fiss, P.C., 2007. A set-theoretic approach to organizational configurations. Acad. Manag. Rev. 32, 1180–1198. https://doi.org/10.2307/20159362

France, S.L., Shi, Y., Kazandjian, B., 2021. Web Trends: A valuable tool for business research. J. Bus. Res. 132, 666–679. https://doi.org/10.1016/j.jbusres.2020.10.019

Frei, P., Leleux, B., 2004. Valuation--what you need to know. Nat. Biotechnol. https://doi.org/10.1038/bioent814

Gaibraith, J., 1982. The Stages of Growth. J. Bus. Strategy 3, 70–79. https://doi.org/10.1108/eb038958

Gao, L., Porter, A.L., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W., Huang, L., 2013. Technology life cycle analysis method based on patent documents. Technol. Forecast. Soc. Change 80, 398–407. https://doi.org/10.1016/j.techfore.2012.10.003

Garnsey, E., Stam, E., Heffernan, P., 2006. New Firm Growth: Exploring Processes and Paths. Ind. Innov. 13, 1–20. https://doi.org/10.1080/13662710500513367

Gawer, A., 2014. Bridging differing perspectives on technological platforms: toward an integrative framework. Res. Policy 43, 1239–1249. https://doi.org/10.1016/j.respol.2014.03.006

Gilbert, B.A., McDougall, P.P., Audretsch, D.B., 2006. New venture growth: a review and extension. J. Manage. 32, 926–950. https://doi.org/10.1177/0149206306293860

Gilpin, A.R., 1993. Table for conversion of kendall's tau to spearman's rho within the context of measures of magnitude of effect for meta-analysis. Educ. Psychol. Meas. 53, 87–92. https://doi.org/10.1177/0013164493053001007

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014. https://doi.org/10.1038/nature07634

Gladwell, M., 2000. The Tipping Point: How Little Things Can Make a Big Difference, 1st ed.

LITTLE, BROWN AND COMPANY, New York.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., Watts, D.J., 2010. Predicting consumer behavior with web search. Proc. Natl. Acad. Sci. U. S. A. 107, 17486–90. https://doi.org/10.1073/pnas.1005962107

Gompers, P., 1995. Optimal investment, monitoring, and the staging of venture capital. J. Finance 50, 1461–1489. https://doi.org/10.1111/j.1540-6261.1995.tb05185.x

Gompers, P., Gornall, W., Kaplan, S.N., Strebulaev, I.A., 2016. How Do Venture Capitalists Make Decisions? Stanford Univ. Grad. Sch. Bus. Res. Pap. No. 16-33; Eur. Corp. Gov. Inst. - Financ. Work. Pap. No. 477/2016 95. https://doi.org/10.3386/w22587

Gompers, P., Lerner, J., 2001. The venture capital revolution. J. Econ. Perspect. 15, 145–168. https://doi.org/10.1257/jep.15.2.145

Gornall, W., Strebulaev, I., 2017. Squaring Venture Capital Valuations with Reality. Cambridge, MA. https://doi.org/10.3386/w23895

Göthensten, V., Hellström, A., 2017. Finding product-market fit. University of Gothenburg, School of Business, Economics, And Law.

Graybill, F.A., Iyer, H.K., 1994. Regression Analysis: Concepts and Applications, 1st ed. Duxbury Pr.

Greiner, L.E., 1997. Evolution and revolution as organizations grow: A company’s past has clues for management that are critical to future success. Fam. Bus. Rev. 10, 397–409. https://doi.org/10.1111/j.1741-6248.1997.00397.x

Greiner, L.E., 1972. Evolution and revolution as organizations grow. Harv. Bus. Rev. 50.

Gulati, R., Higgins, M.C., 2003. Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. Strateg. Manag. J. https://doi.org/10.1002/smj.287

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B., Akter, S., 2017. Big data and predictive analytics for supply chain and organizational performance. J. Bus. Res. 70, 308–317. https://doi.org/10.1016/j.jbusres.2016.08.004

Gupta, P., Guha, S., Krishnaswami, S., 2013. Firm growth and its determinants. J. Innov. Entrep. 2, 15. https://doi.org/10.1186/2192-5372-2-15

Gupta, Y.P., Chin, D.C.W., 1993. Strategy making and environment: an organizational life cycle perspective. Technovation 13, 27–44. https://doi.org/10.1016/0166-4972(93)90012-K

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R., 2010. Multivariate Data Analysis, 7th ed. Pearson Education, Upper Saddle River, New Jersey.

Hanks, S.H., 1990. An Empirical Examination of the Organization Life Cycle in High Technology Organizations. Auburn University.

Hanks, S.H., Watson, C.J., Jansen, E., Chandler, G.N., 1994. Tightening the Life-Cycle Construct: A Taxonomic Study of Growth Stage Configurations in High-Technology Organizations. Entrep. Theory Pract. 18, 5–29. https://doi.org/10.1177/104225879401800201

Hayes, J., 2014. The Theory and Practice of Change Management. Macmillan Education UK, London. https://doi.org/10.1007/978-1-137-28902-5

He, L., Cheng, Y., Su, X., 2020. Research on the sustainability of the enterprise business ecosystem from the perspective of boundary: The China case. Sustain. 12. https://doi.org/10.3390/SU12166435

Hein, A., Schreieck, M., Riasanow, T., Setzke, D.S., Wiesche, M., Böhm, M., Krcmar, H., 2019.

Digital platform ecosystems. Electron. Mark. https://doi.org/10.1007/s12525-019-00377-4

Henrekson, M., Johansson, D., 2010. Gazelles as job creators: a survey and interpretation of the evidence. Small Bus. Econ. 35, 227–244. https://doi.org/10.1007/s11187-009-9172-z

Hoenen, S., Kolympiris, C., Schoenmakers, W., Kalaitzandonakes, N., 2014. The diminishing signaling value of patents between early rounds of venture capital financing. Res. Policy 43, 956–989. https://doi.org/10.1016/j.respol.2014.01.006

Hoenig, D., Henkel, J., 2015. Quality signals? The role of patents, alliances, and team experience in venture capital financing. Res. Policy 44, 1049–1064. https://doi.org/10.1016/j.respol.2014.11.011

Hsu, D.H., 2007. Experienced entrepreneurial founders, organizational capital, and venture capital funding. Res. Policy. https://doi.org/10.1016/j.respol.2007.02.022

Hsu, D.H., Ziedonis, R.H., 2013. Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. Strateg. Manag. J. 34, 761–781. https://doi.org/10.1002/smj.2037

Hsu, D.H., Ziedonis, R.H., 2008. Patents as quality signals for entrepreneurial ventures. Acad. Manag. Proc. 2008, 1–6. https://doi.org/10.5465/ambpp.2008.33653924

Huang, J., Li, C., Yu, J., 2012. Resource prediction based on double exponential smoothing in cloud computing, in: 2012 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet 2012 - Proceedings. pp. 2056–2060. https://doi.org/10.1109/CECNet.2012.6201461

Hurmelinna-Laukkanen, P., Ritala, P., 2010. Protection for profiting from collaborative service innovation. J. Serv. Manag. 21, 6–24. https://doi.org/10.1108/09564231011025092

Ismail, S., Malone, M.S., Geest, Y. van, 2014. Exponential Organisations: why new organisations are ten times better, faster, and cheaper than yours (and what to do about it). Divers. Publ. Corp 1–24.

Jarratt, D.G., 1996. A comparison of two alternative interviewing techniques used within an integrated research design: A case study in outshopping using semi-structured and non-directed interviewing techniques. Mark. Intell. Plan. 14, 6–15. https://doi.org/10.1108/02634509610131108

Järvinen, J., Tollinen, A., Karjaluoto, H., Jayawardhena, C., 2012. Digital and social media marketing usage in B2B industrial section. Mark. Manag. J. 22, 102–117.

Jirásek, M., Bílek, J., 2018. The organizational life cycle: Review and future Agenda. Qual. Innov. Prosper. 22, 1–18. https://doi.org/10.12776/QIP.V22I3.1177

Jun, S.P., Park, D.H., 2016. Consumer information search behavior and purchasing decisions: empirical evidence from Korea. Technol. Forecast. Soc. Change 107, 97–111. https://doi.org/10.1016/j.techfore.2016.03.021

Jun, S.P., Park, D.H., Yeom, J., 2014a. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. Technol. Forecast. Soc. Change 86, 237–253. https://doi.org/10.1016/j.techfore.2013.10.021

Jun, S.P., Yeom, J., Son, J.K., 2014b. A study of the method using search traffic to analyze new technology adoption. Technol. Forecast. Soc. Change 81, 82–95. https://doi.org/10.1016/j.techfore.2013.02.007

Jun, S.P., Yoo, H.S., Choi, S., 2018. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. Technol. Forecast. Soc. Change 130, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Kane, T.J., 2010. The importance of startups in job creation and job destruction. SSRN Electron. J.

https://doi.org/10.2139/ssrn.1646934

Kaplan, S.N., Strömberg, P., 2004. Characteristics, contracts, and actions: Evidence from venture capitalist analyses. J. Finance LIX, 2173–2206.

Katz, D., Kahn, R. l., 1978. The Social Psychology of Organizations, 1st ed. Wiley.

Kaufmann, K.W., 1981. Fitting and using growth curves. Oecologia 49, 293–299. https://doi.org/10.1007/BF00347588

Kazanjian, R.K., 1988. Relation of dominant problems to stages of growth in technology-based new ventures. Acad. Manag. J. 31, 257–279. https://doi.org/10.5465/256548

Kazanjian, R.K., 1983. The organizational evolution of high technology new ventures: The impact of stage of growth on the nature and structure of planning process. University of Pennsylvania.

Kazanjian, R.K., Drazin, R., 1990. A stage-contingent model of design and growth for technology based new ventures. J. Bus. Ventur. 5, 137–150. https://doi.org/10.1016/0883-9026(90)90028-R

Kazanjian, R.K., Drazin, R., 1989. An Empirical Test of a Stage of Growth Progression Model. Manage. Sci. 35, 1489–1503. https://doi.org/10.1287/mnsc.35.12.1489

Kendall, M.G., 1948. Rank correlation methods, 1st ed. Griffin, London.

Ketchen, D.J., Combs, J.G., Russell, C.J., Shook, C., Dean, M.A., Runge, J., Lohrke, F.T., Naumann, S.E., Haptonstahl, D.E., Baker, R., Beckstein, B.A., Handler, C., Honig, H., Lamoureux, S., 1997. Organizational configurations and performance: a meta-analysis. Acad. Manag. J. 40, 223–240. https://doi.org/https://doi.org/10.2307/257028

King, A.A., Lenox, M.J., Terlaak, A., 2005. The strategic use of decentralized institutions: Exploring certification with the ISO 14001 management standard. Acad. Manag. J. https://doi.org/10.5465/AMJ.2005.19573111

King, W.R., He, J., 2005. External validity in IS survey r esearch. Commun. Assoc. Inf. Syst. 16, 17. https://doi.org/10.17705/1cais.01645

Kirk, C.P., Ray, I., Wilson, B., 2013. The impact of brand value on firm valuation: the moderating influence of firm type. J. Brand Manag. 20, 488–500. https://doi.org/10.1057/bm.2012.55

Klein, L.R., 1998. Evaluating the potential of interactive media through a new lens: search versus experience goods. J. Bus. Res. 41, 195–203. https://doi.org/10.1016/S0148-2963(97)00062-3

Kobzar, A.I., 2006. Applied mathematical statistics. For engineers and scientists., 1st ed. Fizmatlit, Moscow.

Köhn, A., 2018. The determinants of startup valuation in the venture capital context: a systematic review and avenues for future research. Manag. Rev. Q. 68, 3–36. https://doi.org/10.1007/s11301-017-0131-5

Kollmann, T., Kuckertz, A., 2010. Evaluation uncertainty of venture capitalists' investment criteria. J. Bus. Res. https://doi.org/10.1016/j.jbusres.2009.06.004

Kollmann, T., Stoeckmann, C., Hensellek, S., Kensbock, J., 2016. Deutscher startup monitor 2016. Berlin. https://doi.org/10.1017/CBO9781107415324.004

Kucharavy, D., De Guio, R., 2011. Application of S-shaped curves. Procedia Eng. 9, 559–572. https://doi.org/10.1016/j.proeng.2011.03.142

Kucharavy, D., Guio, R. De, 2007. Application of S-shaped curves, in: TRIZ-Future Conf. 2007 Curr. Sci. Ind. Real. Frankfurt, Germany, pp. 81–88.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3

Land, G.T.L., 1973. Grow or Die. J. Creat. Behav. 7, 77–132. https://doi.org/10.1002/j.2162-

6057.1973.tb00951.x

Last, D., 2016. Consensus forecasting of zero-day vulnerabilities for network security, in: 2016 IEEE International Carnahan Conference on Security Technology (ICCST). IEEE, pp. 1–8. https://doi.org/10.1109/CCST.2016.7815718

LaViola, J.J., 2003. Double exponential smoothing: an alternative to Kalman filter-based predictive tracking, in: EGVE ’03: Proceedings of the Workshop on Virtual Environments 2003. pp. 199–206. https://doi.org/10.1145/769953.769976

Le Nghiem, T.P., Papworth, S.K., Lim, F.K.S., Carrasco, L.R., 2016. Analysis of the capacity of google trends to measure interest in conservation topics and the role of online news. PLoS One 11, 1–12. https://doi.org/10.1371/journal.pone.0152802

Lee, C., Lee, K., Pennings, J.M., 2001. Internal capabilities, external networks, and performance: a study on technology-based ventures. Strateg. Manag. J. 22, 615–640. https://doi.org/10.1002/smj.181

Lee, J., Lee, S., 2006. Failure factors of new technology-based ventures according to the growth stages, in: Babson Kauffman Entrepreneurship Research Conference (BKERC). Babson College, pp. 951–972.

Lee, T.Y., Bradlow, E.T., 2011. Automated Marketing Research Using Online Customer Reviews. J. Mark. Res. 48, 881–894. https://doi.org/10.1509/jmkr.48.5.881

Leskovec, J., Adamic, L.A., Huberman, B.A., 2007. The dynamics of viral marketing. ACM Trans. Web 1. https://doi.org/10.1145/1232722.1232727

Lester, D.L., Parnell, J.A., Carraher, S., 2003. Organizational life cycle: a five-stage empirical scale. Int. J. Organ. Anal. 11, 339–354. https://doi.org/10.1108/eb028979

Lester, D.L., Parnell, J.A., “Rick” Crandall, W., Menefee, M.L., 2008. Organizational life cycle and performance among SMEs. Int. J. Commer. Manag. 18, 313–330. https://doi.org/10.1108/10569210810921942

Levie, J., Lichtenstein, B.B., 2010. A Terminal Assessment of Stages Theory: Introducing a Dynamic States Approach to Entrepreneurship. Entrep. Theory Pract. 34, 317–350. https://doi.org/10.1111/j.1540-6520.2010.00377.x

Lewis-Beck, M.S., Bryman, A., Futing Liao, T., 2004. Cross-Sectional Design, in: The SAGE Encyclopedia of Social Science Research Methods. Sage Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America. https://doi.org/10.4135/9781412950589.n204

Liang, T.P., Huang, J.S., 1998. An empirical study on consumer acceptance of products in electronic markets: a transaction cost model. Decis. Support Syst. 24, 29–43. https://doi.org/10.1016/S0167-9236(98)00061-X

Lippitt, G.L., Schmidt, W.H., 1967. Crises in a Developing Organization. Harv. Bus. Rev. 45, 102–112.

Lyden, F.J., 1975. Using Parsons’ Functional Analysis in the Study of Public Organizations. Adm. Sci. Q. https://doi.org/10.2307/2392123

Maignan, I., Lukas, B.A., 1997. The nature and social uses of the internet: a qualitative investigation. J. Consum. Aff. 31, 346–371. https://doi.org/10.1111/j.1745-6606.1997.tb00395.x

Malyy, M., Tekic, Z., 2018. Role of Startup’s Intellectual Property in VCs Investment Decision-Making: Evidence from Russia, in: 2018 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE). IEEE, Beijing, pp. 1–7. https://doi.org/10.1109/TEMS-ISIE.2018.8478525

Malyy, M., Tekic, Z., Golkar, A., 2019. What drives technology innovation in new space? A preliminary analysis of venture capital investments in earth observation start-ups. IEEE Geosci. Remote Sens. Mag. 7, 59–73. https://doi.org/10.1109/MGRS.2018.2886999

Malyy, M., Tekic, Z., Podladchikova, T., 2021. The value of big data for analyzing growth dynamics of technology-based new ventures. Technol. Forecast. Soc. Change 169, 120794. https://doi.org/10.1016/j.techfore.2021.120794

Marshall, A., 1920. Principles of Economics, 8th ed. Macmillan and Co., London.

Mason, C., Brown, R., 2014. Entrepreneurial ecosystems and growth oriented entrepreneurship, Final Report to OECD. Paris.

May, J.O., Looney, S.W., 2020. Sample Size Charts for Spearman and Kendall Coefficients. J. Biom. Biostat. 11, 1–7. https://doi.org/10.37421/jbmbs.2020.11.440

McCann, J.E., 1991. Patterns of growth, competitive technology, and financial strategies in young ventures. J. Bus. Ventur. 6, 189–208. https://doi.org/10.1016/0883-9026(91)90009-3

McDougall, P., 2003. A Comparison of International and Domestic New Ventures. J. Int. Entrep. 1, 59–82. https://doi.org/10.1023/A:1023246622972

McDougall, P.P., Covin, J.G., Robinson, R.B., Herron, L., 1994. The effects of industry growth and strategic breadth on new venture performance and strategy content. Strateg. Manag. J. 15, 537–554. https://doi.org/10.1002/smj.4250150704

McKelvie, A., Wiklund, J., 2010. Advancing firm growth research: A focus on growth mode instead of growth rate. Entrep. Theory Pract. https://doi.org/10.1111/j.1540-6520.2010.00375.x

McKelvie, A., Wiklund, J., Davidsson, P., 2006. A Resource-Based View on Organic and Acquired Growth. Adv. Entrep. Firm Emerg. Growth. https://doi.org/10.1016/S1074-7540(06)09007-6

Meade, N., 1985. Forecasting Using Growth Curves—An Adaptive Approach. J. Oper. Res. Soc. 36, 1103–1115. https://doi.org/10.1057/jors.1985.200

Mikalef, P., Boura, M., Lekakos, G., Krogstie, J., 2019. Big Data Analytics Capabilities and Innovation: The Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment. Br. J. Manag. 30, 272–298. https://doi.org/10.1111/1467-8551.12343

Miller, D., 1987. The Genesis of Configuration. Acad. Manag. Rev. https://doi.org/10.5465/amr.1987.4306720

Miller, D., 1986. Configurations of strategy and structure: towards a synthesis. Strateg. Manag. J. 7, 233–249. https://doi.org/10.1002/smj.4250070305

Miller, D., Friesen, P.H., 1984. A Longitudinal Study of the Corporate Life Cycle. Manage. Sci. 30, 1161–1183. https://doi.org/10.1287/mnsc.30.10.1161

Miloud, T., Aspelund, A., Cabrol, M., 2012. Startup valuation by venture capitalists: an empirical study. Ventur. Cap. 14, 151–174. https://doi.org/10.1080/13691066.2012.667907

Misangyi, V.F., Greckhamer, T., Furnari, S., Fiss, P.C., Crilly, D., Aguilera, R., 2016. Embracing Causal Complexity: The Emergence of a Neo-Configurational Perspective. J. Manage. 43, 255–282. https://doi.org/10.1177/0149206316679252

Mitchell, V. -W., Boustani, P., 1994. A preliminary investigation into pre- and post-purchase risk perception and reduction. Eur. J. Mark. 28, 56–71. https://doi.org/10.1108/03090569410049181

Modis, T., 2007. Strengths and weaknesses of S-curves. Technol. Forecast. Soc. Change 74, 866–872. https://doi.org/10.1016/j.techfore.2007.04.005

Modis, T., 2006. Discussion: R. Kurzweil, The Singularity is Near. Technol. Forecast. Soc. Change 73, 104–112. https://doi.org/10.1016/j.techfore.2005.12.004

Montgomery, D.C., Peck, E.A., Vining, G.G., 2012. Introduction to linear regression analysis, 5th ed. Wiley.

Moore, D.S., McCable, G.P., Craig, B.A., 2009. Introduction to the practice of statistics, New York.

Moore, D.S., Notz, W.I., Fligner, M.A., 2012. The basic practice of statistics, 6th ed. Macmillan Learning, New York, New York, USA.

Morwitz, V.G., Steckel, J.H., Gupta, A., 2007. When do purchase intentions predict sales? Int. J. Forecast. 23, 347–364. https://doi.org/10.1016/j.ijforecast.2007.05.015

Mosca, L., Gianecchini, M., Campagnolo, D., 2021. Organizational life cycle models: a design perspective. J. Organ. Des. 10, 3–18. https://doi.org/10.1186/s41469-021-00090-7

Mudambi, Schuff, 2010. Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. MIS Q. 34, 185. https://doi.org/10.2307/20721420

Muhos, M., Kess, P., Phusavat, K., Sanpanich, S., 2010. Business growth models: review of past 60 years. Int. J. Manag. Enterp. Dev. 8, 296. https://doi.org/10.1504/IJMED.2010.034865

Müller, M., 2007. Dynamic Time Warping, in: Information Retrieval for Music and Motion. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 69–84. https://doi.org/10.1007/978-3-540-74048-3_4

Nambisan, S., Siegel, D., Kenney, M., 2018. On open innovation, platforms, and entrepreneurship. Strateg. Entrep. J. 12, 354–368. https://doi.org/10.1002/sej.1300

Niesert, R.F., Oorschot, J.A., Veldhuisen, C.P., Brons, K., Lange, R.-J., 2020. Can Google search data help predict macroeconomic series? Int. J. Forecast. 36, 1163–1172. https://doi.org/10.1016/j.ijforecast.2018.12.006

Oberholzer-Gee, F., Strumpf, K., 2016. The effect of file sharing on record sales, revisited. Inf. Econ. Policy 37, 61–66. https://doi.org/10.1016/j.infoecopol.2016.11.001

Orbach, Y., 2016. Parametric analysis of the Bass model. Innov. Mark. 12, 29–40. https://doi.org/10.21511/im.12(1).2016.03

Ozmel, U., Reuer, J.J., Gulati, R., 2013. Signals across multiple networks: How venture capital and alliance networks affect interorganizational collaboration. Acad. Manag. J. https://doi.org/10.5465/amj.2009.0549

Papavlasopoulos, S., 2019. Scientometrics Analysis in Google Trends. J. Scientometr. Res. 8, 27–37. https://doi.org/10.5530/jscires.8.1.5

Parfenov, A., Tekic, Z., Malyy, M., 2022. Analyzing and Forecasting Entrepreneurial Activity in Russia: Big Data Approach. Higher School of Economics - Graduate School of Business (MSc thesis, ongoing).

Patel, J.C., Khurana, P., Sharma, Y.K., Kumar, B., Ragumani, S., 2018. Chronic lifestyle diseases display seasonal sensitive comorbid trend in human population evidence from Google Trends. PLoS One 13, e0207359. https://doi.org/10.1371/journal.pone.0207359

Penrose, E.T., 1952. Biological analogies in the theory of the firm. Am. Econ. Rev. 42, 804–819.

Phelps, R., Adams, R.J., Bessant, J., 2007. Life cycles of growing organizations: A review with implications for knowledge and learning. Int. J. Manag. Rev. 9, 1–30. https://doi.org/10.1111/j.1468-2370.2007.00200.x

Phillips, F., 2007. On S-curves and tipping points. Technol. Forecast. Soc. Change 74, 715–730. https://doi.org/10.1016/j.techfore.2006.11.006

Phillips, F., Hwang, G.G., Limprayoon, P., 2016. Inflection points and industry change: Was Andy Grove right after all? J. Technol. Manag. Grow. Econ. 7, 7–26. https://doi.org/10.15415/jtmge.2016.71001

Picken, J.C., 2017. From startup to scalable enterprise: laying the foundation. Bus. Horiz. 60, 587–595. https://doi.org/10.1016/j.bushor.2017.05.002

Polgreen, P.M., Chen, Y., Pennock, D.M., Nelson, F.D., 2008. Using Internet Searches for Influenza Surveillance. Clin. Infect. Dis. 47, 1443–1448. https://doi.org/10.1086/593098

Preis, T., Moat, H.S., Eugene Stanley, H., 2013. Quantifying trading behavior in financial markets using google trends. Sci. Rep. 3, 1–6. https://doi.org/10.1038/srep01684

Puka, L., 2011. Kendall's Tau, in: International Encyclopedia of Statistical Science. Springer, pp. 713–715. https://doi.org/10.1007/978-3-642-04898-2_324

Puri, M., Zarutskie, R., 2012. On the Life Cycle Dynamics of Venture Capital and Non Venture-Capital Financed Firms. J. Finance 67, 2247–2293. https://doi.org/10.2307/23324373

Qu, S.Q., Dumay, J., 2011. The qualitative research interview. Qual. Res. Account. Manag. 8, 238–264. https://doi.org/10.1108/11766091111162070

Quinn, R.E., Cameron, K., 1983. Organizational Life Cycles and Shifting Criteria of Effectiveness: Some Preliminary Evidence. Manage. Sci. 29, 33–51. https://doi.org/10.1287/mnsc.29.1.33

Rachleff, A., 2019. How to Know If You've Got Product Market Fit [WWW Document]. Start. Gt. with Mike Maples. URL https://greatness.floodgate.com/episodes/andy-rachleff-on-how-to-know-if-youve-got-product-market-fit-XxGvX8DH/transcript (accessed 7.28.21).

Ragin, C.C., 2008. Qualitative comparative analysis using fuzzy sets (fsQCA). Config. Comp. Anal. 87–121. https://doi.org/http://dx.doi.org/10.4135/9781452226569.n5 Print

Ragin, C.C., 2000. Fuzzy-set social science. University of Chicago Press, Chicago.

Ramadan, A., Lochhead, C., Peterson, D., Maney, K., 2015. Time To Market Cap: the new metric that matters.

Ratzinger, D., Amess, K., Greenman, A., Mosey, S., 2018. The impact of digital start-up founders' higher education on reaching equity investment milestones. J. Technol. Transf. 43, 760–778. https://doi.org/10.1007/s10961-017-9627-3

Rawlings, J.O., Pantula, S.G., Dickey, D.A., 1998. Applied Regression Analysis : A Research Tool , Second Edition Springer Texts in Statistics.

Rėklaitis, K., Pilelienė, L., 2019. Principle Differences between B2B and B2C Marketing Communication Processes. Manag. Organ. Syst. Res. 81, 73–86. https://doi.org/10.1515/mosr-2019-0005

Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., Almeida, C.M.V.B., 2019. A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions. J. Clean. Prod. 210, 1343–1365. https://doi.org/10.1016/j.jclepro.2018.11.025

Reymen, I., Berends, H., Oudehand, R., Stultiëns, R., 2017. Decision making for business model development: a process study of effectuation and causation in new technology-based ventures. R D Manag. 47, 595–606. https://doi.org/10.1111/radm.12249

Ries, E., 2011. The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. Currency.

Rogers, E.M., 1962. Diffusion of innovations, 1st ed. Free Press of Glencoe, New York.

Rogers, E.M., Singhal, A., Quinlan, M.M., 2019. Diffusion of innovations, in: An Integrated Approach to Communication Theory and Research, Third Edition. https://doi.org/10.4324/9780203710753-35

Roure, J.B., Keeley, R.H., 1990. Predictors of success in new technology based ventures. J. Bus. Ventur. 5, 201–220. https://doi.org/10.1016/0883-9026(90)90017-N

Rumble, R., Mangematin, V., 2015. Business Model Implementation: The Antecedents of Multi-Sidedness, in: The Eletronic Library. pp. 97–131. https://doi.org/10.1108/S0742-332220150000033021

Rutherford, M.W., 2001. The organizational life cycle and the small firm: toward the identification of thresholds. Aubum University.

Rutherford, M.W., Buller, P.F., McMullen, P.R., 2003. Human resource management problems over the life cycle of small to medium-sized firms. Hum. Resour. Manage. 42, 321–335. https://doi.org/10.1002/hrm.10093

Saha, S.K., Islam, A., Rodela, R.S., 2014. A comparative study on B2B vs. B2C based on Asia pacific region. Int. J. Sci. Technol. Res. 3, 294–298.

Salamzadeh, A., 2015. New venture creation: controversial perspectives and theories new venture creation: controversial perspectives and theories. Econ. Anal. 48, 101–109.

Salvador, S., Chan, P., 2007. Toward accurate dynamic time warping in linear time and space. Intell. Data Anal. 11, 561–580. https://doi.org/10.3233/ida-2007-11508

Santos, F.M., Eisenhardt, K.M., 2005. Organizational boundaries and theories of organization. Organ. Sci. 16, 491–508. https://doi.org/10.1287/orsc.1050.0152

Schneider, C.Q., Wagemann, C., 2012. Set-theoretic methods for the social sciences: a guide to qualitative comparative analysis. Cambridge University Press, Cambridge, Massachusetts.

SciPy.org — SciPy.org [WWW Document], n.d. URL https://www.scipy.org/ (accessed 9.27.21).

Scott, B.R., 1973. The Industrial State: Old Myths and New Realities. Harv. Bus. Rev. 51, 133–148.

Scott, B.R., 1970. Stages of Corporate Development (Part I). Harvard Bus. Sch. Backgr. Note 371-294.

Scott, M., Bruce, R., 1987. Five stages of growth in small business. Long Range Plann. 20, 45–52. https://doi.org/10.1016/0024-6301(87)90071-9

Seip, M., Castaldi, C., Flikkema, M., De Man, A.P., 2018. The timing of trademark application in innovation processes. Technovation 72–73, 34–45. https://doi.org/10.1016/j.technovation.2018.02.001

Shane, S., Venkataraman, S., 2000. The promise of enterpreneurship as a field of research. Acad. Manag. Rev. 25, 217. https://doi.org/10.2307/259271

Sheng, J., Amankwah-Amoah, J., Wang, X., Khan, Z., 2019. Managerial Responses to Online Reviews: A Text Analytics Approach. Br. J. Manag. 30, 315–327. https://doi.org/10.1111/1467-8551.12329

Shepherd, D., Wiklund, J., 2009. Are we comparing apples wiith apples or apples with oranges? Appropriateness of knowledge accumulation across growth studies. Entrep. Theory Pract. https://doi.org/10.1111/j.1540-6520.2008.00282.x

Shilling, M.A., 2013. Strategic Management of Technological Innovation, 4th ed. McGraw-Hill Companies, Inc., New York.

Shim, S., Eastlick, M.A., Lotz, S., 2000. Examination of US Hispanic-owned, small retail and service businesses: an organizational life cycle approach. J. Retail. Consum. Serv. 7, 19–32. https://doi.org/10.1016/S0969-6989(99)00015-6

Shim, S., Eastlick, M.A., Lotz, S.L., Warrington, P., 2001. An online prepurchase intentions model: the role of intention to search. J. Retail. 77, 397–416. https://doi.org/10.1016/S0022-4359(01)00051-3

Shrader, R., Siegel, D.S., 2007. Assessing the relationship between human capital and firm performance: evidence from technology-based new ventures. Entrep. Theory Pract. 893–909.

Shumway, R.H., Stoffer, D.S., 2017. Time Series Analysis and Its Applications, Springer Texts in Statistics. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-52452-8

Sievers, S., Mokwa, C.F., Keienburg, G., 2013. The Relevance of Financial versus Non-Financial Information for the Valuation of Venture Capital-Backed Firms. Eur. Account. Rev. https://doi.org/10.1080/09638180.2012.741051

Singh, S.K., El-Kassar, A.-N., 2019. Role of big data analytics in developing sustainable capabilities. J. Clean. Prod. 213, 1264–1273. https://doi.org/10.1016/j.jclepro.2018.12.199

Smith, K.G., Mitchell, T.R., Summer, C.E., 1985. Top Level Management Priorities in Different Stages of the Organizational Life Cycle. Acad. Manag. J. 28, 799–820. https://doi.org/10.5465/256238

Söderling, R.A., 1998. The Dynamics of The Firm: In Search for a General Model, in: 43rd ICBS World Conference, Singapore, June 8-10, 1998.

Spence, M., Orser, B., Riding, A., 2011. A Comparative Study of International and Domestic New Ventures. Manag. Int. Rev. 51, 3–21. https://doi.org/10.1007/s11575-010-0065-9

Strehle, F., Katzy, B.R., Davila, T., 2010. Learning capabilities and the growth of technology-based new ventures. Int. J. Technol. Manag. 52, 26–45. https://doi.org/10.1504/IJTM.2010.035854

Stuart, T.E., Hoang, H., Hybels, R.C., 1999. Interorganizational Endorsements and the Performance of Entrepreneurial Ventures. Adm. Sci. Q. 44, 315. https://doi.org/10.2307/2666998

Symeonidou, N., Bruneel, J., Autio, E., 2017. Commercialization strategy and internationalization outcomes in technology-based new ventures. J. Bus. Ventur. 32, 302–317. https://doi.org/10.1016/j.jbusvent.2017.02.004

Tam, S., Gray, D.E., 2016. What Can We Learn from the Organizational Life Cycle Theory? A

Conceptualization for the Practice of Workplace Learning. J. Manag. Res. 8, 18. https://doi.org/10.5296/jmr.v8i2.9093

Tambe, P., 2014. Big Data Investment, Skills, and Firm Value. Manage. Sci. 60, 1452–1469. https://doi.org/10.1287/mnsc.2014.1899

Tchernichovski, O., Parra, L.C., Fimiarz, D., Lotem, A., Conley, D., 2019. Crowd wisdom enhanced by costly signaling in a virtual rating system. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1817392116

Teece, D.J., 1986. Profiling from technological innovation: implications for integration, collaboration, licencing and public policy. Res. Policy 15, 285–305. https://doi.org/10.1016/0048-7333(86)90027-2

Tekic, Z., Koroteev, D., 2019. From disruptively digital to proudly analog: a holistic typology of digital transformation strategies. Bus. Horiz. 62, 683–693. https://doi.org/10.1016/j.bushor.2019.07.002

Terpstra, D.E., Olson, P.D., 1993. Entrepreneurial Start-up and Growth: A Classification of Problems. Entrep. Theory Pract. 17, 5–20. https://doi.org/10.1177/104225879301700301

Tidd, J., Bessant, J., Pavitt, K., 2005. MANAGING INNOVATION. Integrating Technological, Market and Organizational Change, 3rd ed. John Wiley & Sons, Ltd, New York.

To, P.L., Liao, C., Lin, T.H., 2007. Shopping motivations on internet: a study based on utilitarian and hedonic value. Technovation 27, 774–787. https://doi.org/10.1016/j.technovation.2007.01.001

Torbert, W.R., 1975. Pre bureaucratic and post bureaucratic stages of organization development. Interpers. Dev.

Tzabbar, D., Margolis, J., 2017. Beyond the startup stage: the founding team's human capital, new venture's stage of life, founder-CEO duality, and breakthrough innovation. Organ. Sci. 1–16.

Van De Ven, A.H., Poole, M.S., 1995. Explaining Development and Change in Organizations. Acad. Manag. Rev. 20, 510–540. https://doi.org/10.5465/amr.1995.9508080329

Verly Lopes, D.J., Bobadilha, G.D.S., Peres Vieira Bedette, A., 2021. Analysis of Lumber Prices Time Series Using Long Short-Term Memory Artificial Neural Networks. Forests 12, 428. https://doi.org/10.3390/f12040428

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: Survey-based indicators vs. Google trends. J. Forecast. 30, 565–578. https://doi.org/10.1002/for.1213

Waerden, B.L., 1969. Mathematical Statistics, International Edition. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-22137-2

Wang, S., Zhou, H., 2004. Staged financing in venture capital: moral hazard and risks. J. Corp. Financ. 10, 131–155. https://doi.org/10.1016/S0929-1199(02)00045-7

Weaver, L., Beebe, T.J., Rockwood, T., 2019. The impact of survey mode on the response rate in a survey of the factors that influence Minnesota physicians' disclosure practices. BMC Med. Res. Methodol. 19, 73. https://doi.org/10.1186/s12874-019-0719-7

Weinzimmer, L.G., Nystrom, P.C., Freeman, S.J., 1998. Measuring organizational growth: Issues, consequences and guidelines. J. Manage. https://doi.org/10.1177/014920639802400205

Wetherill, G.B., Weiss, L., 1962. Statistical Decision Theory. J. R. Stat. Soc. Ser. A 125, 635. https://doi.org/10.2307/2982628

Witt, P., 2004. Entrepreneurs' networks and the success of start-ups. Entrep. Reg. Dev. 16, 391–412. https://doi.org/10.1080/0898562042000188423

Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.-J., 2017. Big Data in Smart Farming – A review.

Agric. Syst. 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Wu, L., Brynjolfsson, E., 2009. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. SSRN Electron. J. https://doi.org/10.2139/ssrn.2022293

Xue, M., Hein, G.R., Harker, P.T., 2005. Consumer and co-producer roles in e-service: analysing efficiency and effectiveness of e-service designs. Int. J. Electron. Bus. 3, 174. https://doi.org/10.1504/ijeb.2005.006909

Yin, R.K., 2009. Case study research: design and methods, 4th ed.

Yu, W., Jacobs, M.A., Chavez, R., Feng, M., 2019. Data-Driven Supply Chain Orientation and Financial Performance: The Moderating Effect of Innovation-Focused Complementary Assets. Br. J. Manag. 30, 299–314. https://doi.org/10.1111/1467-8551.12328

Zhou, H., Sandner, P.G., Martinelli, S.L., Block, J.H., 2016. Patents, trademarks, and their complementarity in venture capital funding. Technovation 47, 14–22. https://doi.org/10.1016/j.technovation.2015.11.005

Zhu, F., Zhang, X. (Michael), 2010. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. J. Mark. 74, 133–148. https://doi.org/10.1509/jmkg.74.2.133

Zider, B., 1998. How venture capital works. Harv. Bus. Rev. 76, 131–9.

Zupic, I., Giudici, A., 2017. New Venture Growth: Current Findings and Future Challenges, in: The SAGE Handbook of Small Business and Entrepreneurship. SAGE Publications Ltd, 1 Oliver’s Yard, 55 City Road London EC1Y 1SP, pp. 191–219. https://doi.org/10.4135/9781473984080.n11

Appendix A – The list of TBNVs and GT codes for Chapter 3

Identifier	Name	# of val points	GT code	Founded	# of funding rounds
5	Uber	19	/m/0gx0wlr	1-Mar-09	22
233	SolarCity	8	/m/03bzk3g	4-Jul-06	20
577	Zero Motorcycles Inc.	9	/m/04y5m1g	1-Jan-06	19
64	Lyft	15	/m/0wdpqnj	22-May-12	18
901	Tigo Energy	6	/m/0t5392c	1-Jan-07	18
14	Kabbage	6	/m/0gx1j64	1-Jan-08	16
225	Pinterest	11	/m/0h3tm0f	28-Oct-08	16
251	Prosper Marketplace	10	/m/0bts_2	22-Mar-05	16
289	Biodesix	6	/g/12hm05116	23-Dec-05	16
40	Airbnb	9	/m/0svqyn7	11-Aug-08	15
228	Sunrun	6	/m/04gr0b4	1-Jan-07	15
487	Ambiq Micro	7	Ambiq Micro	1-Jan-09	15
72	Lending Club	10	/m/03dzvm6	10-Feb-06	14
142	OnDeck	7	/g/11cn94s0zz	4-May-06	14
386	The We Company	9	/m/01306xmp	1-Jan-10	14
819	Compass	7	/g/11gf2dj58c	4-Oct-12	14
1387	Next Step Living	9	/g/1td9fs3_	1-Mar-08	14
1	Twitter	9	/m/0hn1vcg	21-Mar-06	13
3	Facebook	12	/m/0hmyfsv	2-Apr-04	13
20	Postmates	7	/m/012j2p62	1-Mar-11	12
32	MongoDB	9	/m/0gyv7_f	26-Nov-07	12
48	Cloudera	8	/m/0brz3t5	13-Oct-08	12
99	AppNexus	7	/m/04ybndj	13-Sep-07	12
104	Sonendo	6	/g/1tnpl99k	1-Jun-06	12
136	BlueVine	6	/g/11bbx10zfc	1-Jul-13	12
175	Chegg	7	/m/06zq99y	29-Jul-05	12
219	SoFi	7	/m/0z89d4p	26-Apr-11	12
290	Kareo	8	/m/0_v6pl7	17-Feb-04	12
396	Vidyo	7	/m/05zn6bq	15-Feb-05	12
403	23andMe	9	/m/02r_jll	28-Apr-06	12
408	Sphero	7	/m/0_l6fbb	1-Feb-10	12
484	Nexenta Systems	7	/m/0_frh5f	19-Sep-05	12
517	Snap Inc.	11	/g/11c1s2xz66	1-May-11	12
540	Box	11	/m/026lcm7	1-Jan-05	12
648	Quantenna Communications	10	/m/02_y__4	28-Nov-05	12
826	Illumitex	8	/g/11bwjbvy4w	1-Jan-05	12
9	Remitly	6	/g/12qfsrgtc	23-May-11	11
16	Instacart	8	/m/01371452	29-May-12	11
53	Redfin	8	/m/0c2gbc	1-Oct-04	11
126	AppDynamics	8	/m/0h55dn3	1-Apr-08	11

191	One Medical	10	/g/11b7cr9d7g	1-Jan-07	11
195	Bill.com	9	/m/0120zgh3	1-Aug-06	11
211	ConforMIS	6	/m/06btkk_	26-Mar-04	11
245	Smule	8	/m/09k6st5	1-Jan-08	11
326	GoodData	6	/m/0105tfp9	4-Apr-07	11
429	iRhythm Technologies	6	/m/0k3q4k7	14-Sep-06	11
451	Fab	6	/m/0rzp2pp	1-Feb-09	11
645	Obalon Therapeutics	6	/m/04kyxyr	1-Jan-08	11
13	Rent the Runway	9	/m/0hqzp81	3-Mar-09	10
25	Slack Technologies	9	/m/0dd97nr	25-Feb-09	10
26	Twilio	7	/m/0h1bs6j	30-Jun-08	10
30	Matterport	6	/g/11c2k1ph_j	17-Jun-11	10
45	Nutanix	6	/m/0k3lpf4	22-Sep-09	10
47	Acorns	6	Acorns App	29-Feb-12	10
49	Eventbrite	9	/m/02_b5b7	1-Jan-06	10
54	Turo	8	/m/0bwgplz	12-Aug-09	10
68	ChargePoint	8	/m/05t10kb	13-Sep-07	10
85	Stripe	9	/m/0h3qnb8	13-Apr-09	10
94	Anaplan	8	/m/0gg54rn	1-Jan-06	10
100	Coinbase	6	/m/0wr3qjq	14-May-12	10
122	Namely	8	/g/11b8ym57f9	31-Oct-11	10
131	Tanium	11	/g/11c0rpr039	31-May-07	10
156	Welltok	11	Welltok	1-Jan-09	10
158	SeatGeek	7	/m/0b76vk2	1-Apr-09	10
174	DoorDash	7	/g/11b7xlbf4l	21-May-13	10
204	Rover	7	/m/011pyj5r	13-Jun-11	10
210	Avalara	7	/m/02w3zy5	1-Jan-04	10
213	Moderna Therapeutics	8	/m/0w2ztxm	1-Jan-10	10
215	HotelTonight	7	/m/0w19dls	1-Dec-10	10
232	Glassdoor	7	/m/0t53np8	7-Jun-07	10
243	Avant	6	/m/011f4rzr	2-Nov-12	10
430	Apigee	10	/g/11c5wmmh5k	3-Jun-04	10
445	RichRelevance	8	/m/05f4kqj	1-Feb-06	10
470	Kabam	7	/m/0crf19k	29-Sep-06	10
746	Adesto Technologies	7	/g/11bwd8c8hz	1-Jan-06	10
760	BrightSource Energy	6	/m/047sr75	5-Apr-04	10
768	Zuora	7	/m/0hrfb59	12-Sep-06	10
773	SOASTA	8	/m/026ltkj	26-Jan-06	10
814	NxThera	7	/g/1263lm_29	1-Jan-08	10
1165	Practice Fusion	9	/m/06w1f0y	3-Nov-05	10
6	The RealReal	7	/g/11gcx_yx3w	29-Mar-11	9
28	Docker	7	/m/0wkcjgj	28-Apr-10	9
56	Etsy	6	/m/0byvh_	18-Jun-05	9
66	Coursera	7	/m/0j9kbbz	7-Oct-11	9
79	Health Catalyst	8	/g/12610wrhn	1-Jan-08	9

80	Wish	6	Wish app	4-Jul-10	9
90	Zynga	7	/m/05p4w4b	1-Apr-07	9
106	DoubleDutch	6	/g/1v5k37gv	3-Dec-07	9
148	Acquia	7	/m/07904fx	25-Jun-07	9
183	SnapLogic	7	/m/0g9v849	13-Jul-06	9
189	Tealium	7	/m/0126_x3_	1-Mar-08	9
212	Lookout	7	/m/0h_9m0q	13-Jan-05	9
324	Knewton	7	/m/05f6mvz	1-Jan-08	9
325	LevelUp	6	/m/0h96rw1	1-Jan-08	9
362	Turn Inc	6	/m/03dzwfz	1-Jan-04	9
373	Rethink Robotics	7	/g/11c3w9hjp0	1-Jan-08	9
382	Counsyl	6	/g/1td76tht	1-Jan-08	9
446	LivingSocial	8	/m/0s923kl	6-Jul-07	9
461	Ticketfly	6	/m/010fdl6m	1-Feb-08	9
495	Appcelerator	6	/m/0_gw0ml	1-Sep-06	9
558	Personal Capital	7	/g/11bwycg2gc	1-Jul-09	9
839	Carta	6	/g/11c2lbsyz1	5-Jul-12	9
1006	Blekko	6	/m/0cr74xm	1-Jun-07	9
1257	Inkling Systems	6	/m/0wy4tpz	24-Aug-09	9
2	Impossible Foods	6	/g/11bv3x0v3x	23-Mar-11	8
15	Square	8	/m/0by16yq	1-Feb-09	8
37	BuzzFeed	7	/m/076vlrm	1-Nov-06	8
41	Oscar Health	7	/g/11bwk_9yk_	1-Nov-12	8
59	Outreach.io	7	Outreach.io	1-Jan-13	8
60	Gusto	7	/m/012npkpv	1-Nov-11	8
93	Pure Storage	7	/m/0gky6v_	1-Sep-09	8
107	Actifio	6	/m/0g5r_9l	14-Apr-09	8
138	Apptio	8	/m/0100lkxx	2-Nov-07	8
151	MobileIron	8	/m/0ds2j91	23-Jul-07	8
171	SendGrid	7	/m/0wr97kd	20-Jul-09	8
177	ezCater	8	/g/11c55v8jdx	1-Aug-07	8
184	Tradeshift	7	/m/0bmgxrp	1-Jul-09	8
201	Outbrain	7	/m/0pb5nkb	11-Aug-06	8
202	Vroom	6	/g/11clw82p68	1-Jan-12	8
237	MapR Technologies	8	/m/0gvvkny	17-Jun-09	8
259	Guardant Health	6	/m/0yf8dwl	9-Dec-11	8
261	ThreatMetrix	6	/m/010f70yf	22-Jan-08	8
265	PowerReviews	6	/m/03dzv_x	1-Jan-05	8
310	Scale Computing	8	/m/0_yvfgj	1-Jan-06	8
319	Carbonite	7	/m/03brymd	10-Feb-05	
335	Ionic Security	6	/g/11btm3l0r2	4-May-11	8
345	Twist Bioscience	9	/g/11h9q1sjcw	4-Feb-13	8
351	Kony	6	Kony platform	1-Jan-07	8
412	RedSeal	6	/g/11cn6h6rm9	1-Jan-04	8
420	Gigya	8	/m/03hd64n	17-Oct-06	8
459	Xactly	7	/m/010h6tq9	1-Mar-05	8

473	OOYALA	6	/m/0c3z2hp	5-Apr-07	8
521	Marin Software	8	/m/0gx11ws	16-Mar-06	8
529	Phononic	6	/g/1z44b4xm_	1-Jan-08	8
531	Joyent	6	/m/0kqmtt	1-Jan-04	8
559	Flurry	6	/m/03dzv9c	1-Jan-05	8
569	Tintri	8	/m/0n498zx	1-Jan-08	8
583	Coupa Software	8	/m/03cp_t5	17-Feb-06	8
644	Aquantia	10	/m/02tbl03	27-Jan-04	8
807	3-V Biosciences	6	/m/06_z7w3	19-Dec-06	8
867	Scrollmotion	6	/g/12hl2l31t	1-Jan-08	8
955	Liquidia Technologies	7	/m/02s9lyn	8-Jun-04	8
1051	Virident Systems	7	/m/0h54_jq	7-Jul-06	8
1105	Zoove	7	Zoove	1-Mar-04	8
1106	Demandbase	6	/m/0y512d7	16-May-05	8
1107	Jumptap	8	/m/02qqlbz	18-Jan-05	8
1175	Xobni	6	/m/0d7d3c	1-Mar-06	8
1385	ProspX	6	ProspX	1-Jan-05	8
8	Business Insider	8	/m/09gd5v1	1-May-07	7
11	Groupon	7	/m/0rzpt9x	11-Nov-08	7
12	Looker	6	Looker Data Sciences	1-Jan-11	7
27	Credit Karma	6	/m/04ybk2y	8-Mar-07	7
34	Roblox	6	/m/0kytjpp	23-Mar-04	7
36	Dollar Shave Club	7	/m/0j64jfv	13-Apr-11	7
38	Discord	7	/g/11cjpb9d52	22-Mar-12	7
46	FireEye	6	/m/05b_8h7	18-Feb-04	7
65	The Honest Company	6	/m/012968k0	23-May-12	7
70	Nextdoor	6	/m/0hgp88s	4-Dec-07	7
74	Okta	7	/m/0cqdqrq	1-Jan-09	7
78	Marqeta	6	/g/11c54c_40k	4-Nov-10	7
82	Yext	6	/m/0nhhw58	28-Nov-06	7
86	Handy	7	/m/011v6kw_	4-May-12	7
112	New Relic	7	/m/0hhsfhv	1-Feb-08	7
113	Phreesia	7	/m/02_9ssy	1-Feb-05	7
134	FanDuel	6	/m/0j6638f	1-Jul-09	7
137	Bonobos	8	/m/064k2n7	1-Jun-07	7
153	HomeAway	7	/m/02zd585	1-Jan-04	7
162	Betterment	6	/m/0gj9220	25-Aug-08	7
167	Fastly	6	/g/11gg_8kk1x	2-Mar-11	7
173	Sprinklr	6	/m/01270tg8	1-Sep-09	7
188	Brightcove	6	/m/02h_92l	24-Aug-04	7
196	Veracode	8	/m/0bhcd0j	6-Sep-05	7
221	ServiceMax	7	/m/0ynsgfs	1-Jan-07	7
231	Gilt Groupe	8	/m/080cfp9	1-Jan-07	7
267	TrendKite	6	/g/1pv3gt1kn	1-Jan-12	7

322	Urban Airship	6	/m/0hhrcm4	1-Jun-09	7
323	Care.com	6	/m/03ggs3x	1-Oct-06	7
357	3D Robotics	6	/m/012blrfm	1-Jan-09	7
368	Freshworks	7	/g/1tggb7c9	18-Aug-10	7
371	Demandware	7	/m/02yxm9z	18-Feb-04	7
372	InsideView	6	/m/03cnkfp	20-Jan-05	7
399	Natera	6	/m/03bvqp5	1-Jan-04	7
409	Ring	6	/m/0r4sjp1	1-Jan-12	7
422	LifeLock	8	/m/02ryw2h	12-Apr-05	7
453	Boku, Inc.	6	/m/0dgsc5l	1-Jan-08	7
491	Guavus	8	/m/010f7mc2	23-Mar-06	7
518	HyTrust	6	/m/0rytj__	1-Jan-07	7
621	MuleSoft	9	/m/07s5jcv	12-Apr-06	7
641	TabbedOut	6	/g/1hc3bqqlg	26-Jun-09	7
878	Damballa	7	/m/027kmkh	1-Apr-06	7
914	MyoScience	6	/m/0ykxsc3	1-Jan-05	7
1124	Linkable Networks	6	/m/010qc02k	1-Jan-10	7
1126	Balihoo	7	Balihoo	1-Jan-04	7
1304	TeamSnap	6	/m/0w7q7nl	1-May-09	7
1324	Qliance Medical Management	6	/m/0nfwymp	9-Mar-06	7
1514	AtheroMed	6	/g/1tdks0ql	1-Jan-06	7
1538	Jingle Networks	6	/m/0kqsf5	1-Sep-05	7
2001	CrowdStrike	6	/g/11bz0yw54s	29-Aug-11	6
2007	Livongo Health	6	/g/11b7_1l_nd	16-Oct-08	6
2011	PagerDuty	6	/g/11f3xjjwhn	1-Feb-09	6
2020	Sumo Logic	6	/m/0j63nb_	29-Mar-10	6
2025	HubSpot	6	/m/02yxmbc	4-Apr-05	6
2031	Aerohive Networks	8	/m/02_20h8	15-Mar-06	6
2057	Jet.com	6	/g/11bytmql23	4-Apr-14	6
2063	Duo Security	7	/g/11c1tf6dw9	23-Dec-09	6
2065	RetailMeNot	7	/m/0w1fv6q	27-Oct-06	6
2085	thredUP	6	/m/05zsl_7	7-Jan-09	6
2090	Marketo	9	/m/02_rf5d	2-Oct-06	6
2109	Planet Labs	6	/m/0zdmg9r	28-Dec-10	6
2116	Quanergy Systems	6	/g/11gc2dy1gd	1-Nov-12	6
2129	BetterCloud	6	/m/0tkjlgr	1-Nov-11	6
2145	SoundHound Inc.	7	/m/0n3ycdv	1-Sep-05	6
2167	C2FO	7	/m/0qfsrdc	1-Feb-08	6
2264	One Kings Lane	6	/m/010qk09z	2-Dec-09	6
2276	Lotame	6	/m/02qdsjy	6-Mar-06	6
2301	Blue Apron	6	/g/11cs01znsb	1-Aug-12	6
2304	Appirio	6	/m/0gg8nrl	9-Sep-06	6
2324	Rocket Fuel	7	/m/0zn01_j	1-Mar-08	6
2343	Extole	6	/g/11c1x98m9g	1-Jan-07	6
2377	Yelp	6	/m/0bngkg	1-Jul-04	6

2401	Loggly	6	/m/0129c3z2	1-Aug-09	6
2402	Digg	6	/m/07fydk	11-Oct-04	6
2425	Datadog	6	/m/0wfvvyq	4-Jun-10	6
2428	Neotract	7	/g/1tffdfgy	15-Dec-04	6
2431	Mavenir Systems	6	/m/0_gzwrt	30-Mar-06	6
2435	Yodle	6	/m/03gnm7p	15-Mar-05	6
2453	Kala Pharmaceuticals	6	/g/11dxq1jm3r	7-Jul-09	6
2487	SI-BONE	6	SI-BONE	1-Apr-08	6
2492	Thumbtack	8	Thumbtack.com	1-Jan-08	6
2550	ThoughtSpot	6	/g/11cp7fhhtj	21-May-12	6
2565	Tarveda Therapeutics	6	/g/1hc7_gx8q	1-Jan-11	6
2676	Slacker	7	/m/02x0d74	1-Jan-06	6
2683	Elcelyx Therapeutics	6	/g/1tfkwlx7	27-Feb-09	6
2700	Glooko	6	/m/0131zhjz	1-Nov-10	6
2744	Versartis	6	Versartis	1-Dec-08	6
2803	BIND Therapeutics	8	/m/049yr_9	1-Jan-06	6
2810	WePay	6	/m/0g9szr_	1-Jan-08	6
2860	Akros Silicon	9	/m/03bt4x0	19-Jan-05	6
2919	Peermeta, Inc.	6	/m/03bqhng	1-Jan-07	6
2970	PogoPlug	6	/g/1215clc8	1-Jun-07	6
2983	ANTERIOS	7	ANTERIOS	30-Mar-06	6

Appendix B – The list of TBNVs and GT codes for Chapter 4

Identifier	Name	GT code	Founded	# of funding rounds
1166	Allakos	/g/11f770m0w1	1-Jan-12	4
324	MoEngage	/g/11fmstcdq7	1-Jul-14	7
1397	Finally Light Bulb	Finally Light Bulb	1-May-10	5
299	Blend	/g/11g7z02h62	1-Jan-12	9
630	Yerdle	/m/012hdn2b	1-Jan-12	4
1345	Cleave Therapeutics	Cleave Biosciences	1-Jan-10	4
1098	Tresata	/g/11dxq4hjwv	1-Jan-11	3
1134	UNUM Therapeutics	/g/11c73xj0f9	1-Jan-14	5
511	FloQast	/g/11h2p5rvb8	1-Jan-13	7
434	FLEXE	/g/11g6_zy113	1-Jan-13	6
1171	Genprex	/g/11f019y943	1-Jan-09	7
1527	Buoy Health	/g/11g7z0739m	1-Jan-14	6
1476	iHear Medical	/g/11fy271y2m	1-Jan-10	6
864	X4 Pharmaceuticals	/g/11f2b6m7nd	1-Jan-14	5
703	Helion Energy	/m/011l4384	1-Jan-13	6
1449	Edge Therapeutics	/m/0gytdwt	1-Jan-09	9
692	Human API	/g/11f01k1x5r	1-Apr-13	5
379	FinancialForce	/m/0gvt4f_	1-Oct-09	5
334	Turo	/m/0bwgplz	1-Jan-09	12
59	Headspace	/m/01027p7b	1-Jan-10	9
2317	Pasturebird, Inc.	/g/11cs0s46g9	1-Aug-15	3
1985	Liberation Behavioral Health	/g/11b7jfjp0_	1-Jan-15	3
148	Happify	/g/11b77f5yvp	2-Jan-12	9
1041	Capriza	/g/11c740vw6w	14-Jun-11	4
1159	Ygrene Energy Fund	/g/11h465k_jk	1-Jan-10	3
1663	Bigtincan	/g/11dx9390r4	1-Jan-11	3
335	Lever	/g/11clgdlwlv	1-Jan-12	7
969	Zero Hash	/g/11f3sv26pd	1-Jan-15	9
1144	Prothena	/g/11_p54qgg	1-Jan-12	3
1536	Channel Medsystems	/g/1hc38wtp3	1-Jan-09	6
1049	Relationship Science	RelSci	1-Jan-10	6
1596	EndoStim	/m/0zn24n0	1-Jan-09	8
651	Quanergy Systems	/g/11gc2dy1gd	1-Nov-12	6
581	Revel Systems	/m/0h52cl4	1-Jan-10	10
477	LendingHome	/g/11g70ccs3c	1-Jan-13	7
1673	Surna	/g/11dx9jn4kl	1-Jan-09	3
1031	Magenta Therapeutics	/g/11dxpm75b_	1-Jan-15	3
1095	Qwilt	/g/11dxpy4wtv	1-Jan-10	4
1853	FlowBelow Aero	FlowBelow	1-Jan-12	5
971	HG Insights	/g/11g7098t16	1-Oct-10	7
98	Chorus.ai	Chorus.ai	1-Jan-15	4
294	Welltok	/g/11dym8c0mb	1-Jan-09	12
489	Quartet Health	/g/11g70jdv92	1-Jan-14	5
710	Modernizing Medicine	/m/010r8r1c	1-Jan-10	9
874	Urjanet	/m/0113lcrw	1-Jan-10	5
1586	Scaled Agile	/g/11clw4p_mx	1-Jan-11	3

90	Lyft	/m/0wdpqnj	22-May-12	23
1083	AutoGravity	/g/11g70j6x97	20-Oct-15	4
1154	TabbedOut	/g/1hc3bqqlg	24-Jun-09	7
1108	Wildflower Health	/g/11f_j3vqr5	1-Jan-12	4
1814	MeeGenius	MeeGenius	1-Apr-09	5
165	Tile	/m/0wdsm5f	1-Dec-12	5
1246	KeyedIn Solutions	/m/0pdc3tk	1-Jan-11	7
962	Morphic Therapeutic	/g/11f01cdg32	1-Jan-15	3
1089	Cypress Creek Renewables	/g/11c73lfqsl	1-Jan-14	4
1326	C3Nano	/m/0lp1c68	1-Jan-10	6
442	Evelo Biosciences	/g/11c73y7qvf	1-Jan-15	5
506	UJET	/g/11g88nb180	1-Jan-15	4
1032	Imperative Care	/g/11gk7dbm6c	1-Jan-15	3
1581	4th & Heart	/g/11c6vffb1v	1-Jan-14	3
673	Kenna Security	Kenna Security	1-Jan-10	6
1217	Magnolia Medical Technologies	Magnolia Medical Technologies	1-Jan-11	7
1411	Palvella Therapeutics	Palvella Therapeutics	1-Jan-15	4
1374	Fluxergy	/g/11g6__3zyd	1-Jan-13	4
1324	Loadsmart	/g/11dxq0swmm	1-Jan-14	6
764	Bellhops	/g/11b8252d47	1-Jan-11	8
39	MoneyLion	/g/11dxpr24sf	1-Jan-13	6
1415	Tout	/m/0k2fyf9	1-Jan-10	5
222	SmileDirectClub	/m/0130_fqn	1-Dec-13	3
1866	Zindigo	Zindigo	1-Jan-11	6
538	HealthTap	/m/0jzq1bd	1-Jan-10	6
808	ClearDATA	/g/11b6ntbt3b	1-Jan-11	6
396	Everlaw	/g/11dxpyqsh8	28-Oct-10	5
1404	Beyond Limits	/g/11gdg4_s9_	1-Jan-14	3
1150	Harpoon Therapeutics	/g/11c73hlryh	1-Jan-15	3
377	Fundbox	/m/0138t8n4	1-Jan-13	7
939	MacStadium	/g/11g6_xxy9f	1-Jan-11	4
1579	Kinetic Social	/m/0r4388v	1-Oct-11	5
801	LiveIntent	/g/11g7z06m6t	23-Apr-09	4
432	LeafLink	/g/11f01j16rr	1-Jan-15	6
18	Yotpo	/g/11ffv6kt1k	1-Jan-11	10
1683	First Aid Shot Therapy	First Aid Shot Therapy	3-Feb-10	3
827	Wellframe	/g/11c73gtvb2	1-Jan-11	6
571	Fieldwire	/m/0134d8yj	7-Jan-13	6
532	Leanplum	/g/11dxph94bh	1-Jan-12	9
253	BetterCloud	/m/0tkjlgr	1-Nov-11	8
316	Frame.io	/g/1tf3_8tq	7-Feb-14	4
1544	Airway Therapeutics	/g/11c73klwyq	1-Jan-11	7
1076	World View Enterprises	/m/011v626m	1-Jan-13	5
210	Blue Apron	/g/11cs01znsb	1-Jan-12	6
381	Rigetti Computing	/g/11d_tfzg9t	1-Jan-13	9
1656	TeamSnap	/m/0w7q7nl	1-May-09	8
523	Digital Asset	/g/11bw637fzy	1-Jan-14	5
1434	Vividion Therapeutics	/g/11dxpnrq4h	1-Jan-13	5
657	Denali Therapeutics	/g/11c73r92jl	14-May-15	4

559	Trumid	/g/11bw5mdtpk	4-Jun-14	6
361	InfluxData	/g/11crxrq_2g	1-Jan-12	5
265	Heap	/g/11g7y_22mw	1-Jan-13	7
829	Signpost	/m/011lggxy	1-Jan-10	8
1881	Solaris Power Cells	/g/11c6qtzvyq	1-Jan-13	3
688	Usermind	/g/11g6_zgqy7	12-Jun-13	5
99	Cynet	/g/11h_cczzkz	1-Jan-15	4
981	Wibbitz	/g/11h15xzwnk	1-Jan-11	6
76	Pear Therapeutics	/g/11g70b0kg_	1-Jan-13	6
2120	Inova Payroll	/g/11f2b0wmps	1-Jan-10	4
1040	Ubicquia	/g/11g_zlx0d2	1-Jan-14	3
160	Biohaven Pharmaceutical	/g/11c7szg4rf	25-Sep-13	4
544	Decibel Therapeutics	/g/11fy1sskzs	1-Jan-15	4
273	Songtradr	/g/11h_3t7qjx	1-Apr-14	8
421	WP Engine	/m/012vmgq9	1-Mar-10	6
470	CrowdStreet	/g/11c5bjt3_0	1-Mar-13	4
798	Aceable	/g/11dxppc7xt	1-Oct-12	5
177	Benchling	/g/11bwh5v9q6	1-Jan-12	7
13	Airtable	/g/11c3ypc1q3	1-Jan-13	8
576	REX - Real Estate Exchange	/g/11f01gy3pz	1-Sep-14	6
806	ScyllaDB	/g/11g6__c_3_	1-Dec-12	6
1094	Ventec Life Systems	/g/11gt__lj94	1-Jan-13	6
125	Cybereason	/g/11dyzf9js9	1-Jan-12	5
16	SecurityScorecard	/g/11f57szl5t	1-Jan-13	7
478	Fivestars	/g/1q6skph_h	1-Jan-10	7
566	LogDNA	/g/11hcdz1fbl	1-Jan-15	8
963	Whistle	/m/012ngqnm	1-Jan-12	8
699	Human Longevity	/g/11bt_4y5yt	1-Jan-13	4
820	3D Robotics	/m/012blrfm	1-Jan-09	7
693	VideoAmp	/g/11c73zfvpq	1-Jun-14	4
374	Swift Navigation	/g/11hz6tpn5n	1-Jan-12	5
244	Artsy	/m/0n50vbf	1-Jan-09	9
1231	Farcast	Farcast Biosciences	1-Jan-10	5
662	EverQuote	/g/11csqjqxb5	1-Jan-10	4
411	GOQii	/m/012csznx	1-Jan-14	12
214	Collective Health	/g/11b5pjhy26	31-Oct-13	5
1242	Aptinyx	/g/11dymcc2mj	1-Jan-15	3
975	Juniper Square	/g/11dxq2xcl4	1-Jan-14	4
356	HackerRank	/m/0136_l27	1-Jan-09	4
415	FiscalNote	/g/11gg9d69vk	1-Apr-13	12
973	Inspirato	/g/11h15y0670	1-Jan-11	5
1169	MORE Health	/g/11dxpn1r4c	1-Jan-10	6
535	Catalant	/g/11c4b39zgk	1-Jan-13	7
741	Wickr	/m/0zm_y72	1-Jan-11	4
791	Peel	/m/011l2lkw	1-Jan-09	4
228	Dremio	/g/11c73srbzj	9-Jun-15	6
923	High Fidelity	/m/0tkg4bz	1-Apr-13	5
110	VidMob	/g/11dxq1pcl1	1-Jan-14	7
895	Revinate	/g/11dxppbnfy	1-Jan-09	6
730	Cortexyme	/g/11dym8l1l7	1-Jan-12	6
327	Narvar	/g/11dxq02mqw	1-Jan-12	4

1088	Kindred Biosciences	/g/1yl3h8d51	1-Jan-12	4
498	JOOR	/g/11bzxy308w	1-Mar-10	4
1047	PagerDuty	/g/11f3xjjwhn	1-Feb-09	6
139	Flexport	/g/11clw1xl2m	1-Jan-13	8
507	Branch International	/g/11dxpw21gb	1-Jan-15	9
928	EX.CO	/m/012w3q3v	7-Jul-12	5
1377	TableSafe	/g/11dxpp57lw	1-Jan-10	7
82	Ginger	/g/11dxpw38c5	1-Jan-11	10
1269	Kairos Aerospace	/g/11c73z1x5_	1-Jan-14	4
1205	vXchnge	/m/0gty261	1-Jun-13	3
1021	Syros Pharmaceuticals	/g/11bzzydlz2	1-Jan-12	8
1176	Front	FrontApp	1-Oct-13	8
1343	Allena Pharmaceuticals	/g/11f1rkjbl_	1-Jan-11	4
1128	Biocytogen	/g/11cmch3x04	1-Nov-09	4
1433	IOU Financial	/g/11cs69mpl4	1-Jan-11	12
922	ThreatQuotient	/g/11f015z2m0	1-Mar-13	6
563	PeerStreet	/g/11dymvkvh9	1-Jan-13	8
479	Daily Harvest	/g/11c742j2pj	1-Jan-15	5
443	TraceLink	/g/11c73g93qt	1-Jan-09	7
825	OpenSesame	/g/11cn3jqbhm	1-Jul-11	4
61	Stitch Fix	/m/010ql3n8	1-Jan-11	5
999	PerimeterX	/g/11c73p2lp2	1-Nov-14	6
622	Siemplify	/g/11dxq35vnz	1-Jan-15	6
691	Sight Machine	/m/0ncr5lh	1-Jan-12	6
533	Ridecell	/g/11dxpy0ngb	1-Jan-09	8
1261	High Brew Coffee	/g/11dxd2rfhq	1-Jan-13	3
465	Porch	/m/0wfcsr7	1-Jan-13	5
1839	OHR Pharmaceutical	OHR Pharmaceutical	1-Jan-09	3
179	SalesLoft	/m/0k0xg2z	1-Jan-11	8
1063	Nok Nok Labs	/g/11c73fx6nn	1-Nov-11	4
1196	Aerpio Pharmaceuticals	/g/11f0124xcf	1-Jan-11	6
390	mParticle	/g/11g7y_8pgg	1-Jan-13	11
831	Farmer's Fridge	Farmer's Fridge	1-Jan-13	3
269	Boxed	/g/11c73wxj2y	1-Jun-13	5
616	SOPHiA GENETICS	/g/11h4lhq1z7	1-Jan-11	7
283	Uplift	/g/11g700g926	1-Jan-14	7
315	Quora	/m/0bm8t1r	1-Jun-09	4
255	Adaptive Biotechnologies	/g/11byd42rq7	1-Jan-09	9
932	OM1	/g/11fx7wz9xb	1-Jan-15	3
433	CleverTap	/g/11gdw6__yh	24-May-13	5
551	Elation Health	/g/11dxq3zf4d	1-Jan-10	5
1448	Harry's	/g/11bt_dv1c3	1-Jan-13	8
812	Federated Wireless	/g/11gjhfrf37	1-Jan-12	5
333	Gravie	/g/11c6sr6rpq	1-Jan-13	7
155	Stem	/m/0zp1dmh	1-Jan-09	15
297	Fattmerchant	/g/11dxpstf5n	1-Jan-14	4
1513	Ipsidy	/g/11c73jjtht	1-Jan-09	5
132	Strava	/m/0w51322	1-Jan-09	7
189	Medisafe	Medisafe	1-Jan-12	5
230	Copper	Prosperworks	1-Jan-11	5
663	TrustRadius	/g/11g70b39gz	1-Jun-12	4

1004	DiCE Molecules	DiCE Molecules	1-Jan-13	4
1102	REGENXBIO	/g/11ckr6wjmd	1-Jan-09	5
595	Moven	/m/012ng_9s	1-Apr-11	5
593	Apollo	Apollo GraphQL	1-Jan-11	5
775	AiCure	/g/11dxplv4bt	1-Dec-09	3
804	Tastemade	/g/11c54_5l_6	1-Jan-12	5
1984	VitaPath Genetics	/m/07cm552	1-Jan-09	4
295	Illumio	/m/012vyx5r	1-Jan-13	5
242	Alto Pharmacy	/g/11f281s5dq	1-Jun-15	6
1485	Surefire Medical	Surefire Medical	1-Jan-09	4
574	Housecall Pro	/g/11c73k_bq5	1-Jun-13	7
729	Moda Operandi	/m/0vpsc9c	1-Aug-10	8
376	Sun Basket	/g/11bxc7bbnv	1-Apr-14	9
553	Reonomy	/g/1hm26656y	1-Mar-13	7
305	Zenefits	/m/012nw0vk	1-Jan-13	5
783	Valimail, Inc.	/g/11c73t4g0x	1-Mar-15	5
1072	Unchained Labs	/g/11f01d7ybw	1-Dec-14	5
224	ForgeRock	/m/0ng_0sh	1-Feb-10	5
172	Tipalti	/g/11cjj5fjwp	1-Jan-10	7
961	Simulmedia	/g/11cn6b84bg	1-Jan-09	7
1198	Ryan Specialty Group	/g/11fy1xybgs	1-Jan-10	6
1487	MindMixer	MindMixer	10-Jan-10	4
608	Jounce Therapeutics	/g/11c6cyymtq	1-Jan-13	4
240	Solid Biosciences	/g/11c73k7ltn	1-Jan-14	6
1653	HealthVerity	/g/11dxpl3tff	1-Jan-14	3
1111	EcoSense Lighting	/g/11fy25b4fv	1-Jan-09	6
286	Sourcegraph	/g/11c3s8_dkt	1-Jan-13	5
348	Eaze	/g/11c75pq98t	29-Jul-14	10
1190	NowThis	/m/0swljcf	1-Jan-12	4
1311	Eyenovia	/g/11f3t_nx0m	1-Jan-14	3
1529	Sensus Healthcare	/g/11cmsx601z	1-Jan-10	4
51	Clari	/g/11dxp_3nww	1-Jan-12	5
1054	AppDome	/g/11c73k9pwv	1-Jan-11	5
56	Evidation Health	/g/11dxplfxm3	1-Jan-12	10
207	Elastic	/m/0h64sgb	1-Feb-12	5
1420	INmuneBIO	/g/11fjy6xncr	1-Jan-15	5
1327	Aziyo Biologics	/g/11g7y_9jg2	1-Jan-15	4
1029	Ultragenyx Pharmaceutical	/g/1ywtx6gly	1-Jan-10	3
282	Indigo	/g/11c606mtlv	1-Jan-14	11
1045	Theatro	/g/11c73p3hjw	1-Jan-11	7
685	Poshmark	/m/0_x77ck	1-Feb-11	6
1222	Funding U	/g/11fct8m47y	1-Mar-15	6
1059	Royole Corporation	/g/11hbt3hn_y	1-Jan-12	7
1085	Rhythm Pharmaceuticals	/g/11f11kf9pm	1-Feb-10	4
1495	Helius Medical Technologies	/g/11c6qtm7y3	1-Jan-14	8
584	MakeSpace	/g/11dxpqnksm	1-Jan-13	7
702	PrecisionHawk	/g/11c1wxhfdc	1-Jan-10	7
760	Cala Health	/g/11dxq2hxg4	1-Jan-14	3
941	Threat Stack	/g/11bz_zx7cn	1-Nov-12	6

Appendix C – Rules of Google Trends data quality assessment

The first thing I applied to was the GT search tag for a particular term. For instance, Google Trends may provide various search tags for a name of a company, e.g., "Company," "Corporation," "Solar energy company," "Software," "Website," etc., which influence the quality of statistical data. The company name may also reflect the common term (e.g., "Stripe") that influences the quality of GT data as well, so I considered it too. I have also developed the measure of noise in search query data following the simple logic: the closer search data to the beginning of the analyzed period, the lower should be the GT values. The reason behind this logic is connected to the naturally low public fame of a TBNV during the first period of its lifecycle and its future increase with company marketing. To measure this value, I calculated the ratio between the mean at the beginning of GT data and the overall mean – the lower the rate between means, the less systematic error in the case. The last measure I used reflects the amount of the related search queries: the more related queries have the search term, the better are its GT data. All rules and assessment principles are exhibited below.

1. Selection of the company brand name. The brand name of the company may be corrected during the search code selection process by adding the identifiers, e.g., "Inc.," "Corp.," ".com."
2. Selection of the starting point of a data time range as the company foundation date
3. Selection of the category related to the company title in the following priority:
    a. **Group A:**
        i. "Company," "Corporation"
        ii. Category describing a particular type of company, for example, "Solar energy company," "Transportation company," "Photovoltaics company."
        iii. "Software," "Website"

b. **Group B:**

   i. Related to the particular geographical location, for example, “Corporate campus in Lexington, Massachusetts,” “Health in Holladay, Utah,” “Software company in San Francisco, California.”

   ii. “Topic”

   iii. Without categorical relation

4. Assessment of the results according to the following empirically developed criteria:

   a. Brand name uniqueness. By a unique brand name, I understand used-for-it words or a combination of words written together (without blanks) that rarely could be met in the normal language, e.g., “Lyft,” “Airbnb,” “Twitter.” Assessment criteria:

      i. **Good case, 1 point** – **unique** name and identified category is from the **Group A**

      ii. **Fine case, 0.7 points** – **not unique** name and identified category is from the **Group A**

      iii. **Suspicious case, 0.3 points** – **unique** name and identified category is from the **Group B**

      iv. **Bad case, 0 points** – **not unique** name and identified category is from the **Group B**

   b. Systematic noise in the time series. The level of noise is calculated as the ratio of means: the mean GT index of the first year is divided by the overall mean. Assessment principle:

      i. **Good case, 1 point** – the ratio of means ≤ 0.5

      ii. **Suspicious case, 0.5 points** – 0.5 < ratio of means ≤ 0.85

      iii. **Bad case, 0 points** – ratio of means > 0.85

c. Fast noise in the time series. This type of noise is described by sudden outbreaks and fast drops to zero. It is assessed by the level of the overall mean:

   i. **Good case, 1 point** – overall mean ≥ 4

   ii. **Suspicious case, 0.5 points** – 2 ≤ overall mean ≤ 4

   iii. **Bad case, 0 points** – overall mean < 2.

d. Amount of the related search queries. According to the GT mechanism, “users searching for your term also searched for these queries.” Assessment principle:

   i. **Good case, 1 point** – related search queries ≥ 10

   ii. **Suspicious case, 0.5 points** – 5 ≤ related search queries < 10

   iii. **Bad case, 0 points** – related search queries <5

The total assessment criterion is calculated in two steps.

1. First, I calculated the average point between the points for *brand name uniqueness*, *related search queries*, and the *level of fast noise*.
2. Second, I calculated the average between the obtained during the previous step result and the *level of systematic noise*. Since the cases with high systematic noise do not follow the common-sense growth from lower to higher, I treated the significance of this value as equal to the average of all other criteria.
3. Finally, I marked each company according to the obtained grade:
   a. Cases with the grade **lower than 0.6** are marked as **bad**
   b. Cases with the grade **higher than and equal to 0.6** are marked as **good**

Bad cases are also re-checked with other related brand names, even with the lower priority category. If the assessment criterion demonstrates improvement, it is selected for the case. If improvement was not reached, the bad case is excluded from the sample.

Appendix D – The preliminary analysis of GT data utilization for valuations estimation

1. Data selection and preparation

As a research sample, I selected technology-based new ventures (TBNVs) from two utterly different contexts (defined by industry and market sector): b2c food-delivery and b2b money-lending. Based on the previously evidenced marketing differences in b2c/b2b market sectors (Järvinen et al., 2012; Malyy et al., 2021; Rėklaitis and Pilelienė, 2019), I aim to study them separately and, thus, find any similarities or dissimilarities in the developed models.

I aimed to find market segments with specific qualities: (1) highly competitive environment with more than five active players and similar key value propositions; (2) sum of the available VC valuation points should be higher than 20, so at least four valuations per company on average; and (3) companies should be founded not earlier than in 2004. The first quality was developed to make us possible to assume that the segment has a relatively high amount of ventures 'fighting' for the interest of the same audience that will be reflected in GT search query data. The second feature is aimed to protect the models from underfitting, while the third one is the built-in limitation of Google Trends since the instrument does not provide any data for periods earlier than January 1, 2004[43].

In more detail, the b2c sector of the food-delivery industry is characterized by variability of products. For instance, a company could deliver its food of specific types (e.g., *Farm Hill*[55]), meals cooked by freelance professionals (e.g., *Munchery*[56]), or just aggregate proposals from local restaurants (e.g., *DoorDash*[57]). In addition, it is characterized by relatively low entry barriers, which

---

[55] Farm Hill - Overview | Crunchbase [WWW Document], n.d. URL https://www.crunchbase.com/organization/farm-hill (accessed 2.8.20).

[56] Start Eating Better - Delivering Fresh, Tasty Meals at Affordable Everyday Prices [WWW Document], n.d. URL https://www.munchery.com/ (accessed 2.8.20).

[57] DoorDash Food Delivery [WWW Document], n.d. URL https://www.doordash.com/ (accessed 2.8.20).

result in a substantial amount of new ventures working or used to work in this segment. Compared to most other segments, such variability provides more potential cases and adds robustness to the results. In addition, the modern food-delivery industry is based on web technologies and, thus, heavily utilizes internet-marketing channels, which condition potentially high popularity of the related search terms.

Conversely, companies from the b2b segment of the money-lending industry are specialized in providing various SMEs with affordable credit products. Some of them (e.g., *Kabbage*[58], *OnDeck*[59]) utilize proprietary algorithms for scoring companies' business performance, assessing their creditworthiness, and providing the most suitable lending options from partner banks. Other companies (like *Lendio*[60], *SmartBiz Loans*[61]) work as marketplaces linking borrowers to lenders and leaving the decision-making process to the latter. Although both solutions have their customers, the first one seems to be more suitable for SMEs, according to the found valuation data. Despite the fact that TBNVs from this segment operate with significant capital, all of it is provided by third-party banks (for instance, by *Celtic Bank*[62]) what makes them more technology than financial ventures. Together with a clear business model of debt financing, this fact decreases entrance barriers in this segment and leads to the appearance of many players.

---

[58] Small Business Funding Options Up To $250,000 | Kabbage [WWW Document], n.d. URL https://www.kabbage.com/ (accessed 2.13.20).

[59] Small Business Loans Up to $250,000, Simple, Quick, Easy | OnDeck [WWW Document], n.d. URL https://www.ondeck.com/ (accessed 9.2.20).

[60] Simple Small Business Loans | Lendio [WWW Document], n.d. URL https://www.lendio.com/ (accessed 9.2.20).

[61] Top Small Business Financing Online | SmartBiz Loans [WWW Document], n.d. URL https://www.smartbizloans.com/ (accessed 9.2.20).

[62] Celtic Bank - What Can Celtic Bank Do For You Today? [WWW Document], n.d. URL https://www.celticbank.com/ (accessed 9.2.20).

Through analysis of market reports from *CB Insights*[63,64,65,66] and other sources (Eckenrode and Friedman, 2017), I built the initial sample of 17 TBNVs working in the b2c food delivery segment and 31 companies developing b2b products in the money lending industry. Next, I collected information on their valuations during various series of investments and dates of foundation from the *CB Insights* companies' database[67]. After that, I filtered out those, which do not have any valuation information (nine companies from the b2c food-delivery segment, 13 from the b2b money-lending) and are known as subsidiaries of more prominent brands (two companies from the b2c food delivery, one from the b2b money lending). The absence of valuation data makes it impossible to analyze these cases, while marketing campaigns of subsidiaries, as well as their starting conditions, could be influenced by mother companies that may increase the error of the analysis. In addition, I excluded two companies from the b2b money-lending segment (*CircleUp*[68] and *Gemino*[69]), as they reported their operations to be focused on particular sub-industries. These ventures' narrow target audience makes it incorrect to compare their valuation dynamics with the broader-scope companies' evolution.

Finally, to collect Google Trends data of the remaining ventures, I first aggregated the GT codes proposed by the engine related to each sample subject. These GT codes typically have tags, like "Company," "Business," or industry-specific "Online food ordering company," "Caterer," and

---

[63] Celtic Bank - What Can Celtic Bank Do For You Today? [WWW Document], n.d. URL https://www.celticbank.com/ (accessed 9.2.20).

[64] The Fintech 250: The Top Fintech Companies Of 2020 - CB Insights Research [WWW Document], n.d. URL https://www.cbinsights.com/research/report/fintech-250-startups-most-promising/ (accessed 9.2.20).

[65] The Food Startup Pyramid [WWW Document], n.d. URL https://www.cbinsights.com/research/food-startup-pyramid/ (accessed 2.8.20).

[66] The Future of Dining: 89+ Startups Reinventing The Restaurant In One Infographic [WWW Document], n.d. URL https://www.cbinsights.com/research/restaurant-tech-market-map-company-list/ (accessed 2.8.20).

[67] CB Insights Search Company [WWW Document], n.d. URL https://www.cbinsights.com/search/company (accessed 9.2.20).

[68] CircleUp | Creating a transparent and efficient market to drive innovation for consumer brands [WWW Document], n.d. URL https://circleup.com/ (accessed 9.2.20).

[69] Gemino - Healthcare ABL [WWW Document], n.d. URL http://www.gemino.com/ (accessed 9.2.20).

"Financial Technology Company." After this step, I excluded one company whose GT data could not be detached from the noise (*Sprig*[70]) and, thus, undoubtedly related to this company. After all, I obtained two sub-samples: six ventures from the b2c food-delivery segment with 29 valuation points and nine ventures from the b2b money-lending segment with 28 points (Table 1).

Name	**Google Trends code**	**Number of valuation points**	**Date of foundation**
b2c food-delivery segment			
ChowNow	/g/11d_wy162k	1	1/1/2011
DoorDash	/g/11b7xlbf4l	7	5/21/2013
EatStreet	/m/0_x6dd0	3	1/1/2009
Grubhub	/m/03hc1f9	5	1/1/2004
Munchery	/g/11bv303h_r	5	1/1/2010
Postmates	/m/012j2p62	8	5/1/2011
b2b money-lending segment			
Kabbage	/m/0gx1j64	6	1/1/08
OnDeck	/g/11cn94s0zz	7	5/4/06
BlueVine	BlueVine	6	7/1/13
Fundbox	/m/0138t8n4	2	1/1/12
Dealstruck	Dealstruck	1	1/1/12
Street Shares	/g/11bz0g59k7	1	12/1/13
Snap Advances	Snap Advances	2	1/1/09
SmartBiz	/g/11ckky8962	2	1/1/08
Lendio	lendio	1	1/1/11

Table 1. The sample overview

[70] Sprig - CB Insights [WWW Document], n.d. URL https://app.cbinsights.com/profiles/c/qzOdK (accessed 9.2.20).

2. Methodology

During the first step, I automatically collected GT data of the sample companies in a weekly dimension and applied forward and backward (double) exponential filtering (LaViola, 2003) with a weak smoothing ($1 - \alpha = 0.01$) in order to avoid zero drops in the GT time series. Since Google Trends provides a weekly dimension of data maximum for two hundred points, I developed an algorithm, which "sews" the parts of a one time series with simultaneous rescaling. Taking into account that the "official" dates of companies' foundations may not be correct due to various reasons, I collected company data starting a year in advance of the given date. For one case, *Grubhub*[71], which was founded in 2004, for the beginning date of analysis, I took the lowest possible date for GT, i.e., January 1, 2004.

After the "sewing" of GT data pieces was completed, I applied the second stage of smoothing with a stronger coefficient ($1 - \alpha = 0.8$) to avoid fast noise and significant outbreaks. As a result, I obtained time series data for every sample company and scaled between 0 and 100 points. However, for the research goals, I needed to have all series of data on the same scale relative to each other. Therefore, I manually collected from Google Trends the relative maximum values of all ventures. For example, putting three companies into one plot in GT (*Postmates*[72], *DoorDash*[73], and *Grubhub*[71]) will provide us with the understanding that the global maximum has *DoorDash*, and relatively to it, the maximums of the rest are 0.74 (*Grubhub*) and 0.42 (*Postmates*) times lower. After applying these scaling coefficients to the whole time series of the mentioned companies and normalizing between 0

---

[71] Grubhub, Inc. - About Us - Company Timeline [WWW Document], n.d. URL https://about.grubhub.com/about-us/company-timeline/default.aspx (accessed 2.8.20).

[72] Postmates: Food Delivery, Groceries, Alcohol - Anything from Anywhere [WWW Document], n.d. URL https://postmates.com/ (accessed 2.20.20).

[73] DoorDash Food Delivery [WWW Document], n.d. URL https://www.doordash.com/ (accessed 2.8.20).

and 1, I obtained the rescaled curves, which will be analyzed further (Fig. 1). I have repeated this procedure for all sample companies from both analyzed segments.

I put all valuation points to the corresponding period of the GT time series of a particular sample company during the next step. Since, on average, companies from the sample have only three valuation points, I decided to utilize all of them in order to build a single regression model. Worth noting, no valuation intersections were observed: none of the companies had an investment event on a similar date. I have also considered the rates of yearly inflation and recalculated all valuations relatively the smallest one.

After that, I applied the regression analysis for given pairs of GT data points and corresponding valuations to find the regression model with high explanatory power (Eisenhauer, 2009) and demonstrate the undeniable link between the valuations of companies from one *particular* market segment and their rescaled Google Trends data. Therefore, bearing in mind that the valuation of companies lies in the same numerical space (i.e., millions of dollars), I will be able to conclude that GT statistics of one-industry and one-sector-focused new ventures are also located in the same scale (i.e., from 0 to 100) and, thus, can be directly compared.

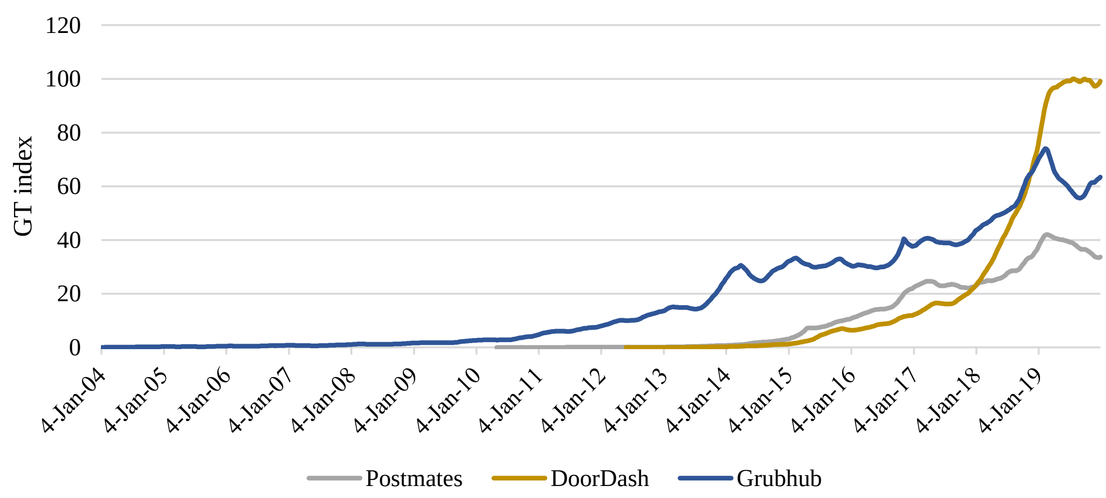


Figure 1. The processed GT data of three “top” companies from the b2c food-delivery sample

For the model of the regression analysis, I applied the polynomial equation (Eq. 1). This model is linear in the fitting parameters but, at the same time, explains the curvilinear relationship between variables through the increase of the model's order. Since so far, for the specific industry, it is unknown whether the relationship between companies' valuations and their GT data is linear, I implement a regression analysis by increasing the fitting equation order until its statistics will not be optimal.

$$\widehat{Val} = a + b_1 GT + b_2 GT^2 + \cdots + b_n GT^n, \tag{1}$$

where $\widehat{Val}$ – is a valuation predicted by model, GT – Google Trends index of the related valuation point, a – an intercept of the Val axis, and $b_1, b_2 \ldots b_n$ – regression coefficients determined from the ordinary least-square method (Rawlings et al., 1998).

For the model quality statistical measures, I selected two core parameters: the multiplier describing the difference between the models by the *Akaike Information Criterion weight (AIC weight)* and the *predictive R^2*, which is based on the *PRESS* statistic. In this research, I utilize the built-in function "Compare models" of the OriginLab[74] statistical software. When a higher-order model is added to the analysis, I calculate AIC between the new and the most accurate models in the research to compare their accuracy. The model with higher accuracy is utilized next for further calculations. The *predictive R^2*, is a "derivative" of another well-known statistical measure: *Prediction Sum of Squares (PRESS)* statistic. It was introduced in 1971 and is now used to assess the regression models' two features: predictive power and fitting quality (Allen, 1974; Montgomery et al., 2012). This measure is a special case of the *Leave-one-out cross-validation (LOOCV)* technique (Kuhn and Johnson, 2013), which estimates how the model-fitting analysis results will generalize to an out-of-

---

[74] Help Online - Origin Help - Comparing Two Models (OriginPro Only) [WWW Document], n.d. URL https://www.originlab.com/doc/Origin-help/PostFit-CompareFitFunc-Dialog (accessed 9.2.20).

sample dataset. PRESS is calculated as a sum of squared errors between the excluded "real" values and the values predicted by the models without it (Allen, 1974). In other words, each value in the model is consequently excluded from the analysis, the model is re-fitted, the excluded value is recalculated, and the residual is measured, squared, and summed. The lower the PRESS, the better is the model for predicting and less probable that it is overfitted (Montgomery et al., 2012). In order to assess its size, it is compared to the *Total Sum of Squares (SST)* produced during any regression analysis implemented by the *Ordinary least squares (OLS)* method. The result of the comparison is known as the *predictive* R^2 measure, which is utilized in the current study. The closer its value to the adj.R^2, the higher its prediction power and the less likely the overfitting of the model. This measure is calculated by the code written by authors in the Python[75] programming language.

By applying the regression analysis, I also identified the outliers of the each-order model. To implement that, I utilized the *Studentized deleted residuals* (or *externally studentized residuals*) technique (Graybill and Iyer, 1994), which is known to be more effective in identifying the outliers than *internally studentized residuals* (Graybill and Iyer, 1994). This technique principle is similar to the calculation of the predictive R^2: each value is excluded from the model, the residual is calculated, and then studentized (Graybill and Iyer, 1994). The absolute residuals higher than a chosen threshold value are proposed to be considered as the outliers. By the threshold value, some scholars utilize the value of *two*, associated with the alpha level of 95%. Since the companies' valuation data *may* contain a significant error, I selected a more strict value of *three* by the threshold (an approximation of 3.29) that makes us 99.9% sure that the detected point is an outlier (Field, 2013).

Altogether, the regression analysis contained the following steps:

1. To derive coefficients of the initial, first-order model;

---

[75] Welcome to Python.org [WWW Document], n.d. URL https://www.python.org/ (accessed 9.2.20).

2. To detect the probable outliers and exclude them from the data;
3. To develop the second-order model;
4. To compare its accuracy with the initial model by the AIC test;
    a. If the AIC demonstrates that the second-order model is more accurate, to accept it as the final model and repeat the outliers' analysis;
    b. In the opposite case, to calculate the predictive R^2 of both models and compare them. The model with the higher predictive R^2 is then accepted as the final;
5. To repeat the previous steps until the model with the best statistic measures is not detected;
6. To compare the best-statistics model with the same-order regression but without intercept;
7. To conclude on the final regression model, which has the highest explanatory and predictive power.

I apply this algorithm separately for two samples and report on the results in the next section.

8. Results

First, I focus on TBNVs from the b2c food-delivery segment. In order to assess the character of the relation visually, I built the scatter plot of valuation points and the related Google Trends scores. Then, applying the OriginLab[76] software, I calculated the first-order linear regression model's coefficients and put its graphical representation to the plot. This initial model had a relatively high adjusted coefficient of determination $adj.R^2 = 0.88$ what makes it possible to conclude the existence of a positive relationship between b2c food-delivery companies' valuations and related GT scores. Next, using a method of *Studentized deleted residuals*, I excluded the probable outlier point (98.33, 12600) and increased the $adj.R^2$ from 0.88 to 0.94.

[76] OriginLab - Origin and OriginPro [WWW Document], n.d. URL https://www.originlab.com/Origin (accessed 9.2.20).

To increase the explanatory power of the model, I derived coefficients of the second-order linear regression and compared it to the initial model by the *AIC test*. Comparing the two models resulted in the fact that the second-order model by 1 / 0.007 = 143.08 times more likely to be correct. Therefore, the *AIC test* demonstrated strong evidence that the second-order linear regression model approximated the data correctly. Next, I repeated the outlier analysis for the second-order regression and put the previous point (98.33, 12600) back to the model with the exclusion of another one whose Studentized deleted residual was higher than a chosen threshold. This step increased the $adj.R^2$ from 0.94 to 0.99.

After that, I calculated the third-order linear regression model but did not evidence the increase in the AIC test. It demonstrated that the second-order regression was by 1 / 0.26 = 3.85 times more accurate than the third-order model. Repeated analysis of outliers did not result in any change in regression data; therefore, I did not accept the third-order model.

Further, considering that the intercept point confidence interval crosses the x-axis, I recalculated the model without intercept, i.e., intercept equals zero. The difference in $adj.\ R^2$ between the model with and without the intercept point is less than 0.001, while the Akaike weight of the model without intercept is by 1 / 0.83 = 1.23 times higher. Since the difference between the two models in the AIC test was not high enough, I calculated the *PRESS* statistic and its *predictive* R^2. The results demonstrated that the model without intercept provides bigger $pred.R^2$; therefore, I selected it as the result of the regression analysis (Table 2).

Comparing the $pred.R^2$ with $adj.R^2$ of the resulting model, I can conclude that it can be used for predicting companies' valuations: the difference equals 0.005. Thus, it can be inferred that the analysis demonstrates a steady relationship between b2c food-delivery market companies' valuations and related to them GT scores. It can be described by the quadratic regression model without intercept and with a Standard Error of \$262.25M, $adj.R^2 = 0.991$, and $pred.R^2$ = 0.985 (Fig. 2).

Coefficient	Parameter	Model 1 – 1st order	Model 2 – 2nd order	Model 3 – 3rd order	Model 4 – 2nd order, no intercept
intercept	Value	-96.57	93.38	111.72	-
	Standard Error	85.25	63.49	74.23	-
	95% LCL	-271.8	-37.39	-41.49	-
	95% UCL	78.67	224.14	264.92	-
b_1	Value	71.93	17.7	8.85	22.63
	Standard Error	3.48	5.71	18.68	4.73
	95% LCL	64.77	5.94	-29.7	12.92
	95% UCL	79.1	29.46	47.41	32.35
b_2	Value	-	1.12	1.44	1.07
	Standard Error	-	0.07	0.66	0.06
	95% LCL	-	0.98	0.082	0.95
	95% UCL	-	1.26	2.81	1.2
b_3	Value	-	-	-0.003	-
	Standard Error	-	-	0.005	-
	95% LCL	-	-	-0.013	-
	95% UCL	-	-	0.008	-
Statistics	R-Square (COD)	0.943	0.990	0.990	0.991
	Adj. R-Square	0.940	0.989	0.989	0.990
	SE	$379.3M	$256.57M	$260.52M	$262.25M
	# of outliers	1	1	1	1
	List of outliers	(98.33, 12600)	(91.56, 7100)	(91.56, 7100)	(91.56, 7100)
	Pred. R-Square	0.912	0.967	-0.93	0.985
	AIC multiplieer*	Model-1 / Model-2 0.007	Model-2 / Model-4 0.81	Model-3 / Model-2 0.26	1

**The AIC multipliers are calculated relatively to the model with best statistics and demonstrate by how many times its accuracy is higher (if it is bigger than one) or lower (if it is smaller than one).*

Table 2. Summarized analysis results for b2c food-delivery sector. The final model parameters are provided in the column marked by green color.

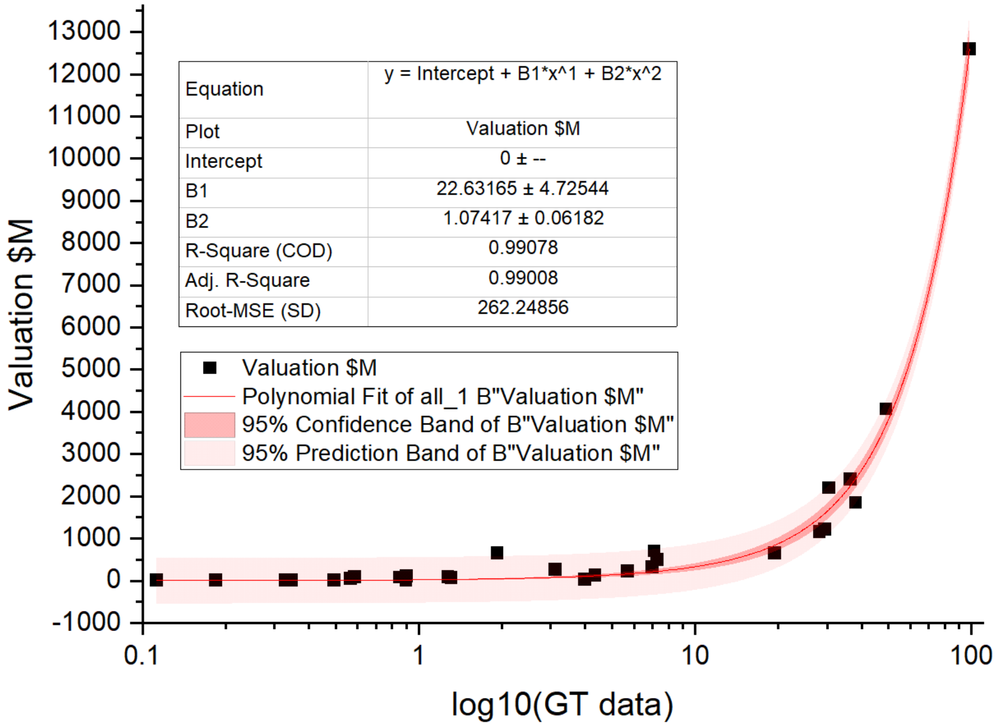


Figure 2. Model-4 regression results (b2c food-delivery segment)

Next, I implemented the same analysis algorithm for TBNVs from the b2b money-lending segment. As in the previous case, a scatter plot demonstrated the existence of the positive link between companies' valuation points and GT data, which can be explained by the initial linear model with $adj.R^2 = 0.86$. I calculated the coefficients of the second-order linear regression and did not achieve any significant increase in $adj.R^2$. The AIC test also demonstrated that the first-order model was 1.87 times more likely to be correct. However, after analyzing the Studentized deleted residuals of the first-order model and excluding two possible outlying points, the repeated AIC test showed that the second-order linear regression provided by 1 / 0.86 = 1.17 times better approximation results. It was also supported by the small increase in the $adj.R^2$ from 0.94 in the first-order linear regression to 0.95 in the second-order model. Therefore, I concluded that the second-order linear regression better explains the sample data. After that, I repeated the analysis of the Studentized deleted residuals and excluded two more outlying points that increased the *adj.*R^2 to 0.98 and decreased the standard error to $46.47 million.

Further, I calculated the third-order linear regression. The AIC test results proposed that the third-order model is 3.27 times more accurate than the second-order, but I detected slight overfitting by checking the *pred.*R^2: *pred.*R^2 of the second-order model is six points less than *pred.* R^2 of the third-order. Since I aim to obtain the optimal statistics model, I selected the second-order regression for further analysis. Next, I compared models with and without the intercept point. The AIC test demonstrated that the model with an intercept point by 1 / 0.17 = 5.88 times more likely to be correct. Therefore, I accepted the second-order linear regression with an intercept as the final model representing the relationship between valuation and GT data of b2b money-lending ventures (Table 3).

Following the methodology of analysis, I tested the obtained model with the *PRESS* statistic and *pred.*R^2. The results showed that the model has good predictive power with the difference between *pred.*R^2 and *adj.*R^2 of 0.02. The standard error of the final model is ten times lower than for the b2c food-delivery case, but this difference may be explained by the lower average valuation of the b2b money-lending ventures. Overall, I can conclude that the relationship between b2b money-lending companies exists and is also quadratic, with the intercept point at $33.76 million (Fig. 3).

Coefficient	Parameter	Model 1 – 1st order	Model 2 – 2nd order	Model 3 – 3rd order	Model 4 – 2nd order, no intercept
intercept	Value	-15.78	33.76	48.01	-
	Standard Error	22.8	13.2	13.55	-
	95% LCL	-62.85	6.3	19.74	-
	95% UCL	31.28	61.21	76.27	-
b_1	Value	14.47	4.33	-0.36	6.68
	Standard Error	0.71	1.42	2.42	1.21
	95% LCL	13.01	1.38	-5.41	4.17
	95% UCL	15.93	7.28	4.69	9.18
b_2	Value	-	0.1	0.27	0.08
	Standard Error	-	0.02	0.076	0.01
	95% LCL	-	0.07	0.11	0.05
	95% UCL	-	0.13	0.43	0.11
b_3	Value	-	-	-0.0013	-
	Standard Error	-	-	5.66E-4	-
	95% LCL	-	-	-0.0025	-
	95% UCL	-	-	-1.16E-4	-
Statistics	R-Square (COD)	0.95	0.983	0.987	0.983
	Adj. R-Square	0.94	0.982	0.985	0.981
	SE	$96.72M	$46.47M	$42.38M	$51.99M
	# of outliers	2	4	4	4
	List of outliers	(20.98, 750) (45.46, 252.24)	(20.98, 750) (45.46, 252.24) (46.52, 848.57) (73.8, 1211.21)	(20.98, 750) (45.46, 252.24) (46.52, 848.57) (73.8, 1211.21)	(20.98, 750) (45.46, 252.24) (46.52, 848.57) (73.8, 1211.21)
	Pred. R-Square	0.94	0.96	0.90	0.97
	AIC multiplieer*	Model-1 / Model-2 0.86	1	Model-3 / Model-2 3.27	Model-4 / Model-2 0.17

**The AIC multipliers are calculated relatively to the model with best statistics and demonstrate by how many times its accuracy is higher (if it is bigger than one) or lower (if it is smaller than one).*

Table 3. Summarized analysis results for b2b money-lending sector. The final model parameters are provided in the column marked by green color.

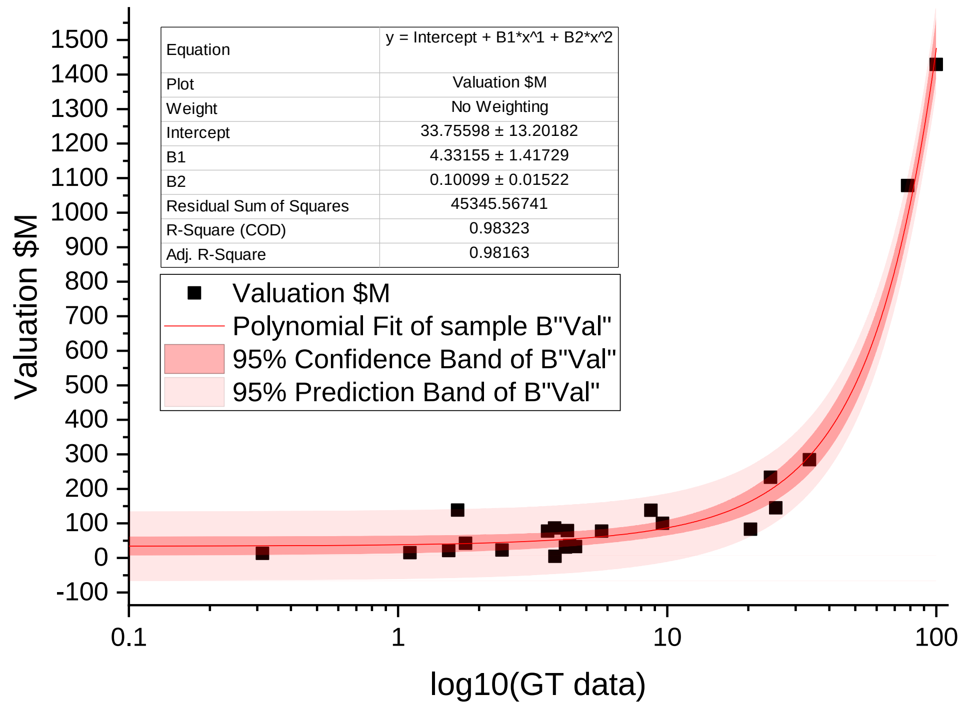


Figure 3. Model-2 regression results (b2b money-lending segment)

9. Discussion

Our study demonstrates that statistics on web search queries of companies can be beneficially utilized for building trustworthy mathematical models that link startups' GT data with their valuations. With the future enhancement, these models can be applied as a complementary tool, helping to increase the accuracy of market value estimations of a particular startup based on its GT data and on GT and (some) investment data of other startups from the same industry and market sector (the competitors). The discovered relation between companies' GT data and their valuations is clear and follows a second-order regression model in both selected contexts: b2c food delivery and b2b money lending. Both developed models demonstrate high determination power (*adj.*R^2 (b2c food-delivery) = 0.99; *adj.*R^2 (b2b money-lending) = 0.98) and high predictive potential (*pred.*R^2 (b2c food-delivery) = 0.985; *pred.*R^2 (b2b money-lending) = 0.96). The models have only the second-order and small differences between the adjusted and predicted r-squared; hence, I can conclude that they are not overfitted.

If to consider a standard error – the characteristic of the regression models, which is directly connected to the variance in data – I can see that it is relatively big in both models: for the b2c food-delivery segment, SE is equal to $262.25 million, and for b2b money-lending segment, $46.47 million. In their turn, these errors lead to relatively wide predictive intervals and, thus, lower accuracy of the calculated data points. For instance, if I would like to reach a 95% precision, I should consider predicted values to lie in ±1.96*SE = ±$514 million bounds for the b2c food-delivery segment and in ±$91.08 million for b2b money-lending segment. Bearing in mind that the majority of valuation data points are related to the ventures, which are valued lower than $100 million**,** such predictive intervals significantly decrease the usefulness of the obtained models. Nevertheless, if I accept the imprecise character of the VC valuation process (Frei and Leleux, 2004; Gompers et al., 2016), I can consider precision levels as low as 70% or even 50% that enhances prediction intervals to ±1.04*SE and ±0.67*SE, respectively (Moore et al., 2009). The recalculated prediction intervals can then be three times narrower if to compare 50% and 95% levels of precision: in the case of b2c food delivery, $PI_{50\%}$ = ±$175.7M and $PI_{95\%}$ = ±$514M. Such an allowance makes sense, especially for the early-stage companies, which have very limited sales results and, thus, are valuated very imprecisely even by the existing methods (Gompers et al., 2016; Miloud et al., 2012). In other words, while the current startup valuation process is more intuitive than objective (Gompers et al., 2016), the models have the potential to make it more “scientific” and data-driven by adding *at least some* assessable precision.

If I compare the results of the b2c and b2b sectors, I can observe several crucial differences. Firstly, the number of detected outliers is significantly greater for the b2b sector (one vs. four). This difference may be explained by the assumption that b2b companies have a weaker connection between their valuation and related GT data (Malyy et al., 2021). New ventures, which develop products for businesses, do not target only web channels for marketing (Rėklaitis and Pilelienė, 2019) and, therefore, may grow in valuation without the corresponding increase in web search queries. This explanation is also backed by the fact that three out of four outlying points of the b2b sample lie above the regression curve, i.e., higher valuation for the related GT score than predicted by the model. The same reason may explain the second difference between the two sectors: the model for b2b sample has an intercept while the b2c does not. An intercept in the obtained model means that b2b companies demonstrate some value when they do not have any Google Trends search statistics. This valuation can be reached only by applying the marketing channels different to the web.

The third difference is connected to the regression curves' forms: the slope of the curve for the b2c market sector is significantly steeper than for the b2b. It is needed to be mentioned that forms of the obtained curves should not be directly compared since their difference may be caused by two independent factors: market sector (b2c vs. b2b) and industrial area (food delivery vs. money lending). However, it can be observed that the steepness of the b2c sector curve is mostly caused by the maximum valuation in the b2c food-delivery sample, from which I can conclude that the b2b money-lending ventures obtain significantly lower valuations than b2c food-delivery. At least two factors may explain this difference: (1) the b2b money-lending market is considerably smaller than b2c food-delivery; (2) b2b money-lending market is considerably younger. Analysis of these factors, as well as their premises and consequences, can be implemented in future studies. Differences in maximum valuations of b2c food-delivery and b2b money-lending companies also lead to another fourth difference: a standard error for the latter is more than five times smaller than for the former. However, due to the different valuation axis scales, it is not correct to compare SEs directly. I assume that both models' accuracy can be compared better by the use of *relative standard errors (RSE)* calculated for the sample means (Montgomery et al., 2012). For the b2c food-delivery sub-sample, RSE_{mean} is equal: 262.25 / 1261.65 * 100 = 20.8%; for the b2b money-lending sub-sample RSE_{mean} = 46.47 / 260.8 *

100 = 17.8%. Thus, I can conclude that both cases have very similar accuracy when related to their data.

I have also noticed that the obtained regression models have a parabolic character. That signals the existence of valuation growth acceleration while reaching the industry-top position in terms of search query statistics. In other words, the more public audience is interested in the product, the faster the company grows in valuation. At the same time, the highest GT score among the companies from the sample, in my opinion, signals about sector leadership: the more people searching the particular brand, the more popular it is and the higher its sales are (Jun et al., 2014b). Hence, I may conclude that the closer the TBNV to market leadership positions, the faster growth in valuation it will face. Although this conclusion follows common sense, I would like to point out that the dependency of the startup's valuation growth speed from its market position was not evidenced in the literature so far.

Our research makes three contributions. First, I contribute to the literature on VC valuation of firms (Gompers et al., 2016; Kaplan and Strömberg, 2004; Puri and Zarutskie, 2012), and especially on usage of non-financial information for the valuation of VC-backed firms (Amir and Lev, 1996; Hsu, 2007; Hsu and Ziedonis, 2013, 2008; Sievers et al., 2013). I present, test, and discuss – to the best of my knowledge for the first time – a method to build sector-specific mathematical models, linking information on public interest and new ventures' valuations. With future enhancements in terms of accuracy, these models may help to assess the unknown valuations of the "hidden" ventures from the same context (industry and market sector), relying on the public, free, and easy to access GT data. Bearing in mind that startups and TBNVs prefer not to publicly disclose their valuations, the proposed methodology may become an extremely valuable instrument for theoretical analysis of new ventures' (value) evolution.

Next to opening the door to more "scientific" valuations of new ventures and reduction in their over and under-valuations, the developed approach offers one more (practice-oriented) opportunity.

Parameters of the regression curves for various contexts can be compared to determine the most promising one. Such kind of insights could be of value for venture capitalists seeking new investment opportunities, startup owners facing problems with sales in the current market, government decision-makers aiming to sustain the startups' ecosystem growth, and other players in the entrepreneurial market. At the same time, it opens the door to democratizing the practice of venture valuation. Namely, being reasonably simple and based on open data, the proposed approach helps in leveling the playing field in the valuation efforts of different parties – investors, founders, competitors, and observers – and provides ground for decreasing the level of information asymmetry among parties interested in valuation.

Secondly, I add new insights to a recent stream of research on the role of third-party signals in venture capital financing (Chen and Xie, 2005; Courtney et al., 2017; Gulati and Higgins, 2003; King et al., 2005; Mudambi and Schuff, 2010; Ozmel et al., 2013; Stuart et al., 1999; Tchernichovski et al., 2019; Zhu and Zhang, 2010). The findings show that information contained in aggregated Internet search query data about a specific startup, once put into the context of a particular industry and market sector (through aggregated Internet search query data about other similar startups), can be used as a signal about customers' interest into a product or service, and thus, of the startup's market potential (underlying quality). According to the signaling theory, to be reliable and to reduce information asymmetry, a signal has to have two main characteristics – to be observable by outsiders and costly to produce (Connelly et al., 2011). Google Trends data related to a startup satisfies both requirements. It is a costly signal since it directly depends on the level of public interest towards the product and the happening of the meeting market demand that is, by itself, an infrequent and costly situation. Moreover, not to lose the obtained level of public interest, the company needs to spend budgets to sustain the accepted level of quality and service. Furthermore, Google Trends data related to a startup (and all startups from the context defined by the industry and market sector) are observable for external parties through public, real-time data, which is available without any fees and payments on a specialized website. It does not depend on anyone's (including the startup) willingness to disclose this information. Information is free, but the signal is costly to emit. Thus, Google Trends data related to startups seems to be valid Spencian signals of the startup's market potential – in a similar vein as patents filed by startups signal the underlying quality of their technology (Hsu and Ziedonis, 2008). This finding is highly interesting as it brings the promise of better assessment of market acceptance of startup products, which is still a source of extreme uncertainty for VCs (Kollmann and Kuckertz, 2010), and calls for further analysis.

10. Conclusions, limitations, and further studies

New ventures are hard to valuate objectively. And the "newer" they are, the harder the task is. Due to their immaturity, they can barely provide enough financial data commonly used in corporate investment decision-making. Thus, the valuation process heavily depends on future expectations of the "valuator" and becomes more intuitive and, hence, risky. To solve this issue, I propose a new, easily accessible, objective, and boundless source of data (*Google Trends*[6]) and test a method that can be beneficially utilized to build the sector-specific models, describing the dependency of companies valuations from search query statistics related to them. Up to this moment, GT demonstrated capabilities in various management applications (for example, Choi & Varian, 2012; Goel et al., 2010; Jun, Park, et al., 2014), which made it possible to hypothesize the existence of the particular link between the company's web traffic data and its value (in line with the recent study of Malyy et al., 2021). In the current research, I proved this hypothesis for the US technology-based new ventures from two different contexts: b2c food-delivery and b2b money-lending. I obtained almost the same type of model for both sub-samples: second-order (i.e., quadratic) linear regression with intercept for the b2b money-lending segment and without it for the b2c food-delivery. This similarity and the possibility of identifying the models with high explanatory and predictive power make it possible to conclude the particular generalizability of the research outcome. Overall, the results suggest that *Google Trends* can be beneficially utilized for building the sector-specific valuation models, which have the potential to help VCs in assessing the companies of interest market value.

Nevertheless, the research has particular limitations. Firstly, the results would be more comparable if I took new ventures from one industry and two different market sectors. However, it is extremely difficult to find a narrow technology area with enough new companies separately working in both b2b and b2c sectors and providing detailed information on their funding rounds.

Secondly, during the regression analysis, in both models, I obtained significant standard errors, critically decreasing the practical utility of the models. One possible explanation of these errors may be a "spy" character of information on companies' valuations during funding rounds that leads to an unpredictable variance in data. According to the references in the selected data source CB Insights, information on valuation may be provided by companies' representatives like CEOs[77] and some abstract people "familiar with the matter[78]" or "spokesmen[79]." In some cases, valuation data may not be even referenced (and, therefore, presented as an insight of the database) or referenced by some SEC's indirect information. Moreover, the provided valuation data can be changed and deleted over time. For instance, funding information of b2b money-lending company *Kabbage* was changed presumably after its acquisition by *American Express*[80] (Fig.4).

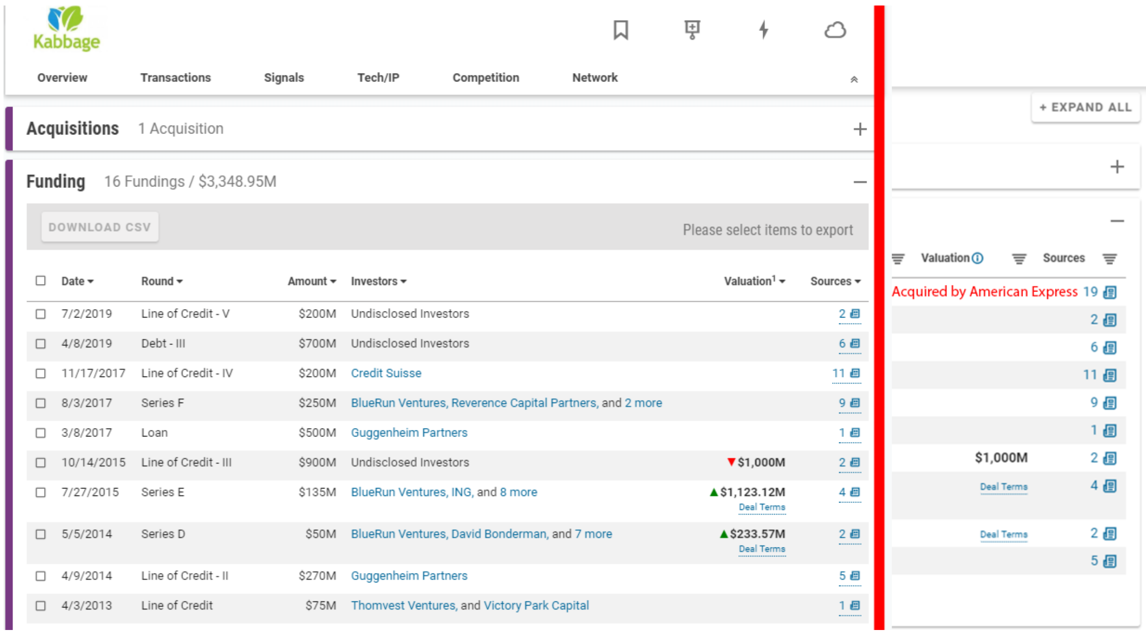
---

[77] DoorDash Is Now Worth Nearly As Much As Grubhub After $400 Million Funding Infusion [WWW Document], n.d. URL https://www.forbes.com/sites/bizcarson/2019/02/21/doordash-funding-400-million-grubhub-7-billion-valuation/#54e0ad547e10 (accessed 9.2.20).

[78] DoorDash Is Raising at Least $500 Million in Funding [WWW Document], n.d. URL https://www.bloombergquint.com/business/2019/05/17/doordash-is-said-to-be-raising-at-least-500-million-in-funding (accessed 9.2.20).

[79] Fintech Fundbox Raises $176 Million to Lend to Business Using AI - BNN Bloomberg [WWW Document], n.d. URL https://www.bnnbloomberg.ca/fintech-fundbox-raises-176-million-to-lend-to-business-using-ai-1.1320877 (accessed 9.2.20).

[80] Kabbage - CB Insights [WWW Document], n.d. URL https://app.cbinsights.com/profiles/c/jDqO (accessed 9.2.20).



(a) (b)

Figure 4. Change in the *Kabbage* profile in the CB Insights database before (a) and after (b) it was acquired by *American Express*

Altogether, these limitations are caused by a relatively low number of regression points and their unequal distribution across the valuation scale. I can assume that these issues may be solved in future studies by (1) finding more market segments with enough players; (2) obtaining more detailed information on companies' valuations during VC funding rounds; and (3) applying the advanced Data Science techniques for data preparation and analysis. I aim to solve these issues in future studies and, thus, increase the practical applicability of the proposed instrument and source of data.

Generally, I can infer that the link observed and discussed in the study opens new perspectives in the area of new venture analysis. Despite the mentioned limitations, the wide accessibility of the employed data source (*Google Trends*) may lead to more accurate assessments of the companies-of-interest valuations after additional research and methods elaboration. Since the problem of valuing new ventures is widely recognized and accepted, I believe that the proposed technique may become a starting point towards turning the startup community from being subjective and intuitive to more data-driven and scientific.