

## Thesis Changes Log

**Name of Candidate:** Maksim Malyy

**PhD Program:** Engineering Systems

**Title of Thesis:** The data-driven model of technology-based new ventures growth

**Supervisor:** Associate Professor Zeljko Tekic, HSE University

*The thesis document includes the following changes in answer to the external review process.*

### **Report of Prof. Clement Fortin**

#### **The reviewer's comment:**

“The methods of analysis are well adapted to the topic; in the first part, the methodology is presented in a condensed format which could be further developed for better precision of the content. As an example, the following statement on page 53 is not clear: “I selected the type of correlation measure, which provides the most reliable results in such uncertain scenarios (Allen, 2017).” This is the most important statement for the thesis, but the reason why it is considered the “most reliable” must be explained in detail. The same applies to the relationship between Kendall’s tau coefficient and Pearson’s rho coefficient, where a formula is given but is not explained. More precise explanations must be included in this part of the thesis.”

#### **Response to the reviewer:**

- Thank you for the comment and suggestion. The text was corrected to include more precise explanations of the relationship between Kendall’s tau coefficient and Pearson’s rho coefficient. Please, see part 3.2.4.3, paragraph 4.
- Additionally, the following sentence was corrected in Chapter 1, paragraph 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.”

### **Report of Prof. Alexey Nikolaev**

#### **The reviewer's comment:**

“Namely, even though the thesis includes considerations on the limitations for the applicability of the proposed research and its conclusions, they require additional elaboration. The breadth and variety of technology-based new ventures (varying in technology and its development lifecycle, business/partners ecosystems readiness and maturity, target markets, product adoption lifecycle specifics, exit strategy the company founders pursue, business model features, company publicity strategy, etc.) may not fit into correlations revealed by the thesis study and result in wrong estimations of the company growth and its OLC stage. Further clustering and classification of technology based new ventures with respect to their specifics and applicability of the proposed models can be suggested as the direction for the future research.”

#### **Response to the reviewer:**

Thank you for the comment. The text was corrected to better elaborate limitations of the results as suggested by the reviewer. Please, see Chapter 8, paragraph 2.

**The reviewer's comment:**

“The second block of suggestions deals with the consideration of the opportunities and limitations for the predictions to be made basing on Google Trends data. This also can be related to the directions of further study and include, for example, consideration of what amount and character of data (e.g., Google Trends observations time window, variability of data with time, etc.) can be considered as the basis for further dynamics prediction and with what accuracy probability.”

**Response to the reviewer:**

Thank you for the comment. The text was corrected to include this valuable suggestion. Please, see Chapter 8, paragraph 3.

**Report of Prof. Dmitry Koroteev****The reviewer's comment:**

“I would like to see a much wider discussion on why the author selected S-curve to approximate the data. It is rather clear that the S-curve is a robust and straightforward way of fitting the nonhuge dataset. Still, it seems like there could be some conventional machine learning (ML) techniques that can handle the dataset more precisely. I am not talking about Deep Learning here, just simple regression models, random forests, support vector machines, and similar. Moreover, some classical ML methods allow the non-trivial analysis of the feature importance for the final prediction. So, if the data on the companies is a bit richer than what google trends provide, there is an option for detailed analysis on what makes the companies successful. I encourage Maksim to comment on this.”

**Response to the reviewer:**

Thank you for the comment and the “invitation” to discuss the topic in more details. I selected models with the S-curve growth part for several reasons. First, these models were employed in the previous topic-related academic works but in a conceptual and illustrative manner (I review these studies in Chapter 4, Section 4.2). Therefore, I wanted to check if these (very popular) conceptual models can be backed by empirics. Second, since the goal of the study was to identify the basic and general TBNVs growth trajectory, I started from the basic mathematical models and searched for those, which were able to propose a mechanism for a tipping point identification since it is known as one of the most influential points within TBNVs' growth. Knowing the coordinates of this point makes it possible to use it as a reference for accurate alignment of growth-related managerial practices with actual TBNVs' conditions and to enhance them while studying companies' historical decisions with great precision. Third, since it is the pioneering study applying this source of data for analysing TBNVs, I decided to start from the simpler methods backed with literature, adding complexity where it was required. For instance, some measures used to assess curve-fitting quality, like dynamic time warping, cross-correlation, and cross-validation, are also known as the instruments applied for solving machine and deep learning tasks.

For sure, machine learning techniques can also be applied to TBNVs' GT data. For example, random forests, support vector machines, and logistic regression models can be employed for identifying the TBNVs' features, which influence the correlation level between GT data related to them and VC valuations. From my perspective, due to the many possible independent variables (i.e., TBNVs' features), the sample should contain more cases to solve this task with high statistical significance. Otherwise, it may be hard to derive meaningful conclusions. Moreover, the TBNVs' related Bass models' parameters (i.e.,  $p$ ,  $q$ ,  $m$ , and timing of the critical points) can be used in this case as the dependent variables. These parameters can be accurately identified for each TBNV with available GT data and, thus, used as its unique growth describing characteristics. Therefore, adding these parameters into regressions and other ML techniques is likely to provide more insightful outcomes backed with rigorous theoretical knowledge. Another application of ML algorithms on TBNVs GT data may help to build precise short-term forecasts of their evolution. Despite the fact that forecasting power was not in the focus of the current study, based on the previous literature, I can assume that TBNVs-related GT data can be of particular value for implementing this kind of analysis (please, see paragraphs 4 and 5 of part 3.2.6 and paragraph 3 of Chapter 8). These techniques are planned to be examined in future studies (please, see Chapter 9, points 10 and 11).

**The reviewer's comment:**

“It is absolutely clear and is confirmed by the author that the developed model is applicable for US only. I would like to ask Maksim to provide some ideas on what to do with the evaluation of the growth in the EU, MENA, Russia, China, India, Singapore, Australia, and other tech hubs.”

**Response to the reviewer:**

Thank you for this comment. Considering the case of Google is the main search engine in the region, the following text was added to Chapter 8, paragraph 1: “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, 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.”

Considering the case that Google is not a leading search engine, the following text was corrected in Chapter 8, paragraph 5: “In addition, Google Trends may have limited power for some countries, which have their own dominating web search engines, i.e., China and North Korea with Baidu, Russia with Yandex, and South Korea with Naver. 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 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.”

**Report of Prof. Fred Phillips****The reviewer's comment:**

“p3. Dynamics correlated with search traffic. Technically, only variables can be “correlated.” Dynamics is not a variable. It's clear you're phrasing this informally, however, so it may not be essential to correct it on p.3. You do use the same phrasing on p.35 and in H2 on p.44; there in the body of the paper, more rigorous expression is needed.”

**Response to the reviewer:**

The following sentence was corrected in the Abstract: “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.” In a similar vein were corrected the other text of this kind in hypotheses and other parts of the text (pp. 36, 42, 44, 45, 47, 48, 78, 79).

**The reviewer's comment:**

“p3, bottom. “due to their features.” More specifics here would help the reader.”

**Response to the reviewer:**

Thank you for carefully reading the thesis and paying attention to important details that I missed to clarify in the previous version. The following sentence was corrected in the Abstract: “Next, bearing in mind that the common sigmoid equations have limited applicability due to their intrinsic autocatalicity leading to the inability of growth starting point determination...”

**The reviewer's comment:**

“p15, 1st sentence, “...builds is operations...” This states an ideal. Most organizations' biggest problem is ‘strategic alignment,’ i.e., making sure ‘managerial decisions’ (as you mention at the bottom of the page) all serve to further the stated strategy.”

**Response to the reviewer:**

The following sentence was corrected in Chapter 1, paragraph 1: “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.”

**The reviewer's comment:**

“p18. A more important comment: A ‘disruptive’ innovation creates a new market. ‘Market share growth’ is thus meaningless as a measure of the disruptive new firm's growth. One could say that the innovative firm initially has 100% of its newly created market and as imitators jump in, the first firm may grow in revenue while shrinking in share. So, a shrinking share in a newly created market could be a correlate of success! The disruptive firm/innovation offers a new customer benefit (or benefits similar to products already on offer) at significant (at least 10x) improved cost/benefit ratio. So it's necessary to distinguish

between share of product market, and share of benefit offerings.”

**Response to the reviewer:**

Thank you for this comment. In general, I can agree with this statement. However, I could not find rigorous studies employing a shrinking share as a measure of success. In this part of the text, I tried to discuss the widely employed measures of TBNVs dynamics and explain their drawbacks. I think the topic of studying a shrinking share as a measure of success may result in an interesting future study.

**The reviewer’s comment:**

“p.19, top. In-person interviews are a qualitative technique, best only used as ways to structure subsequent quantitative market research. They are usually not sufficient in themselves to make marketing decisions.”

**Response to the reviewer:**

Thank you for this comment. The following sentence was corrected in Chapter 1, paragraph 7: “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) and are commonly employed for qualitative research design while aiming to identify research directions for a subsequent quantitative study (Jarratt, 1996; Qu and Dumay, 2011).”

**The reviewer’s comment:**

“p.21. Comment on whether you think 241 is a sufficient sample size for your purpose.”

**Response to the reviewer:**

Thank you for this comment. Yes, I think that 241 TBNVs is a sufficient size for testing the main and secondary hypotheses with the selected methods. The following sentence was corrected in Chapter 1, paragraph 9: “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.” More corrections on this comment are presented further in part 3.2.4.1: “Applying the method based on 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.”

**The reviewer’s comment:**

“p.22. Is autocatalyticity a real word? Don’t imply that I coined it!”

**Response to the reviewer:**

Thank you for the comment. First, I could not find any other reference to this phrase and, thus, decided that this term was developed by the reviewer. However, later I found that other similar terms exist and applied, e.g., “intrinsically autocatalytic.” The following sentence was corrected in Chapter 1, paragraph 12: “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).”

**The reviewer’s comment:**

“p.23, “Since my model is quite...” From here to the bottom of the page, the achievements appear overstated. In particular, “objective” may not be correct, as you later state the GT algorithms are opaque. You don’t know how objective the GT data are.”

**Response to the reviewer:**

Thank you for the comment. The following text in Chapter 1, paragraph 14, was toned down: “Third, by proposing and describing this model, I introduce a solution to the existing ambiguity, which exists 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.”

**The reviewer's comment:**

“p.24 and elsewhere. I'm delighted that you think the q value solves the product-market fit problem. I'm going to consult with Dr. Steve Walsh, who's familiar with the product-market fit idea, and send you further feedback on this. Keep in mind though that the p-m fit problem might equally be solved by repeat-buying rate, or by still something else.”

**Response to the reviewer:**

Thank you for the comment. I will be happy to hear more on the product-market fit idea from you and Dr. Steve Walsh. This term is known as the “holy grail” of the current new venture development practice and would be very interesting to study it further.

**The reviewer's comment:**

“p.26 is very interesting.”

**Response to the reviewer:**

Thank you for this comment.

**The reviewer's comment:**

“p.28. Why are SMEs relevant in this dissertation on TBNVs? Unlike TBNVs, which aim at growing fast, SMEs may persist while never growing at all. And if a hypothesis tested on 4600 SMEs only proved out on 5 of them (!), why is it worth mentioning at all?”

**Response to the reviewer:**

Thank you for this comment. The following footnote was added to Chapter 2, paragraph 1: “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.”

Also, I have corrected the following sentence in Chapter 2, paragraph 4: “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).”

**The reviewer's comment:**

“p.30, 1st ¶ re validity and non-response bias. Good point! Also a very nice summary on p.32.”

**Response to the reviewer:**

Thank you for this comment.

**The reviewer's comment:**

“p.36: It's possible, actually likely, that potential unicorns are more heavily hyped by the VC investors. This will affect subsequent searches by journalists, analysts, small investors, etc. ”

**Response to the reviewer:**

Thank you for this comment. I would say this is an interesting direction for future research. From the first glance, I can fully agree with this position: yes, VCs may influence the unicorns' popularity growth with their heavy marketing activities. However, if the rise in search queries would be highly influenced by VCs only, probably, we would have seen mainly peaky data without the main growth trend because when the marketing hype ends, users slowly stop searching for a company. So, I assume the number of search queries, in this case, would oscillate around its average, and growth in valuations would not be correlated with the growth in GT data.

**The reviewer's comment:**

“p.36: “The study makes two contributions” – This statement appears to be out of place. Put it in an intro section or in the later discussion section.”

“I add to the extant literature...” Only if the dissertation is published! And, you are hardly likely to add to a non-extant literature... And you repeat this statement and the one above, on p.86.”

“contribution is... primarily methodological.” Too modest, you are making bold theoretical assertions in this dissertation.”

“p.37. “organized as follows.” Repetitious – you already said this in earlier pages. Page 38 continues repeating earlier material.”

**Response to the reviewer:**

One more time, thank you for carefully reading the thesis and paying attention to important details that I missed to clarify in the previous version. The related text is excluded from part 3.1, after paragraph 6 and until the end of the part.

**The reviewer’s comment:**

“p.39, bottom: “previously evidenced” Is it now different?”

**Response to the reviewer:**

Thank you for this comment. The following sentence was corrected in part 3.2.1. “It was evidenced 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).”

**The reviewer’s comment:**

“p.46, footnote: Are “digital platforms” and “two-sided networks” really synonymous? Can you think of platforms that are not 2-sided markets, or vice versa?”

**Response to the reviewer:**

Thank you for this comment. Some studies and other information sources employ the notion of "a digital platform" in the meaning of the products, which consist of the functional core and the amount of possible third-party modules (for instance, please, see here: <https://www.equalexperts.com/blog/our-thinking/so-what-is-a-digital-platform-anyway/>). The footnote you mention is intended to distinguish one definition from another, which is applied in the research.

**The reviewer’s comment:**

“p.60ff. As I said, you were very careful about sample selection. 241 companies met all your conditions. Are so many statistical tests appropriate on what is a relatively small (though necessarily small, as you explain) sample? Perhaps you over-analyzed, which can lead to false positives or false negatives. Why did you smooth the data before fitting? What value did this add?”

**Response to the reviewer:**

Thank you for the comment. Text added to part 3.2.4.1, paragraph 4, describes (backed with references) the calculation of the minimum needed sample size considering the hypotheses set, power and confidence level taken, and the method applied. The threshold size of the sample is 17, so it can be concluded that in the current research with 241 subjects in the sample, type 1 and type 2 errors are avoided. I assume that there could be many statistical methods appropriate for this size of the sample, but I selected those which were more suitable for solving the stated research tasks. Considering the data smoothing, usually, real-time time-series data contains three components: main trend, noise, and seasonal component (*Introduction to Time Series and Forecasting*, 2016 by Brockwell, P. J. and Davis, R. A.). In the current study, I was interested in analysing the link between the main growth trend in GT data and VC valuations. Therefore, it was needed to filter out noise from the main trend (also, I was not interested in seasonal variations and, thus, did not consider this possible component as well). Otherwise, I would have measured the correlation between noise+main trend and VC valuation that might have resulted in erroneous outcomes.

**The reviewer’s comment:**

“p.89, ¶2: “important...contribution” – It is up to the doctoral committee and later journal reviewers/editors to decide whether it is important. Just state straightforwardly what you did.”

**Response to the reviewer:**

Thank you for the comment. The related sentence is excluded from part 3.2.6, paragraph 4: “Despite the mentioned limitations, the analysis has been quite promising, offering an important methodological contribution.”

**The reviewer's comment:**

“p.94, last sentence of 2nd¶: Can you restate this in a way that does not insult the Jindal School faculty? Or insults them more subtly ;) ?”

**Response to the reviewer:**

Thank you for the comment. I apologise for choosing inattentive wording. It was not my intention to insult anyone. The following sentence was corrected in part 4.2, paragraph 4: “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.”

**The reviewer's comment:**

“p.95. I believe Everett Rogers pioneered the sigmoid curve idea for diffusion of innovation. You are citing a late edition of his book. I think the 1st edition came out in about 1961. Of course epidemiologists used these curves earlier.”

**Response to the reviewer:**

Thank you for this comment. The following text was added to part 4.2, paragraph 6: “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. In the work of another well-known scholar, Clayton Christensen, the concept of innovation value also utilizes a sigmoid model (Christensen, 2011).”

**The reviewer's comment:**

“p.104: Explain “dynamic time warping,” or be suspected of authoring science fiction!”

**Response to the reviewer:**

Thank you for this comment. Please, note that the dynamic time warping (DTW) method is explained in part 4.3.2, paragraph 4. I also changed the wording from *dynamic time warping* to *DTW* in the section 4.3.3, paragraph 1.

**The reviewer's comment:**

“p.110: The idea that the logistic curve has no tipping points, nor could have tipping points, is really due to Modis, though he didn't say that in exactly those words. It was his work on this that led me to look at the Bass curve as a way to uncover tipping points in new product diffusion.”

**Response to the reviewer:**

Thank you for this comment. The related text was corrected in the section 4.4, paragraph 2: “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 Phillips' 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).”

**The reviewer's comment:**

“p119, you mention hype. I'm surprised that you did not find (more?) companies that suffered through the hype cycle and then recovered and grew. The Gartner Group has now branded the hype curve as its own – mainly because it is indeed so common – but marketing researchers knew of it long before Gartner. See Fig. 2.7 of the attached chapter from my (now near-obsolete) 2001 textbook F. Phillips, Market-Oriented

Technology Management: Innovating for Profit in Entrepreneurial Times. Springer, Heidelberg, 2001.”

**Response to the reviewer:**

Thank you for the comment and the provided materials. Here, the hype is mentioned in a speculative manner in order to suggest an explanation to the phenomenon of  $q$ -points with zero  $x$ -coordinates, which was observed in 87 cases. However, the idea proposed by you sounds very interesting and may be useful for our future research. So far, I've decided to put the examination of the reasons causing this effect aside as they are outside of the scope of my research.

**The reviewer's comment:**

“Appendix: In the body of the dissertation, explain why you have moved further methodological details to the Appendix. Then eliminate the significant amount of Appendix material that repeats (in Discussion, Conclusions, Limitations....) what was already said in the body text.”

**Response to the reviewer:**

Thank you for this comment. The following text was added to Chapter 1, last paragraph: “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.”

Also, the mentioned parts were revised and excluded if needed. In particular, were excluded parts in Appendix D, part 2, paragraph 6; part 9, paragraph 9; and the first three sentences in part 10, paragraph 2. In addition, please, note that the text in Appendix D, part 9, paragraph 8 explains the contribution made by the preliminary analysis described in this appendix from the signalling theory perspective, which was not discussed in the main body of the thesis.

**Report of Prof. Vincent Mangematin**

**The reviewer's comment:**

“Why Organization Life Cycle so central?”

The thesis aims at contributing on the organizational life cycle. This theory is an important one. However, it brings several questions:

- The reason why for choosing this approach remains unexplored. Why this approach rather than another one to explain the growth of TBNV? Which could have been the other alternatives?”

**Response to the reviewer:**

Thank you for the comment. The major reason for choosing OLC theory is the sequential manner of TBNVs valuation growth. As it is demonstrated in parts 3.1 and 3.2, venture capital companies seek for TBNVs with high growth potential and ultimately focuses on this growth. This fact may be explained by the mechanism of VC investment, due to which VCs return their investments when they exit the company and sell its shares to a next round investor, public market during IPO, or to a bigger corporation during M&A. Therefore, their primary task is to support the high pace of the portfolio TBNVs' growth while targeting the maximum returns, which, obviously, happen when a company reaches its maximum valuation. VCs also invest in serial principle, when each series reflect a particular state of a company's evolution. Since OLC theory also understands organizations' evolution process as a set of particular states, I decided to consider it as a core / underling theory for my research. I also, to some extent, agree with Mosca et al. (2021, p. 15), who argue that “organizational configurations are equifinal instead of universal either because organizational dimensions depend on one another or because organizational configurations must solve contrasting needs” while adding that some configurations still may be universal (like ones proposed by the extremely popular practice-oriented Lean Startup methodology).

At the same time, alternative theories, providing frameworks for explaining organizational development, may also be applicable. From the well-known concepts proposed by Van de Ven & Poole (1995), *teleological* understanding can benefit first from the results obtained in the current study. Teleology sees organizational evolution as driven by the actions taken to reach the goal or a particular state. In this way, a TBNV's maximum valuation may be considered as this particular state, to which an organization should grow while taking various and not predefined (as proposed by OLC theory) actions. I can assume that other organizational frameworks (e.g., those discussed by Van de Ven & Poole (1995), evolution and



dialectic, a mix of all, and others) can be considered from the perspective proposed in the actual thesis research. Overall, the key idea is that (a) to reach success, TBNVs should grow (at least VC-backed); (b) this growth is reflected by GT data; (c) and has an S-shape with three precisely defined parts. So, existing OLC concepts, as well as the other organizational evolution theories, both discussed here and not, can be reviewed from the proposed growth character perspective and accordingly advanced.

**The reviewer's comment:**

“- Emerging TBNVs are usually digital and based on multisided markets. The organization life cycle has been developed far before the emergence of digital platforms. To what extent is the approach adapted to describe the evolution of platform new venture.”

**Response to the reviewer:**

Thank you for the comment. In this research, the OLC theory is used only as a concept to explain the staged growth of TBNVs' valuations, while the question on companies' configurations, which should be predetermined for each stage according to OLC models, was put aside. So, the applicability of the OLC theory for studying the exact features configurations of digital platform and other cases is planned to be addressed in future research. Additionally, since in both Study 1 and Study 2 the obtained results were also considered from the product type dimension (digital platform vs traditional product), it can be concluded that the use of the developed growth logic is legitimate for digital platforms as well (or even more legitimate comparing to the traditional products, since the average correlation between digital platforms' valuations and GT data related to them is higher).

**The reviewer's comment:**

“- One of the hypotheses behind the OLC is that the evolution is not chaotic. Considering the actual trajectory of TBNV, this hypothesis needs to be addressed”

**Response to the reviewer:**

Thank you for the comment. Looking from the perspective of the obtained results to the non-chaotic character of TBNVs growth hypothesized by the OLC theory, it can be argued that the research outcomes, to some extent, support this hypothesis. Indeed, under the TBNVs' growth-driven evolution perspective, the process of movement from zero to the maximum possible company valuation looks predetermined. Moreover, the serial character of investment implies the similarity of conditions and states needed to reach the next investment round. However, at the same time, each company most likely will need to take specific actions and demonstrate the case-related results to overcome the barriers for further investment rounds. Also, I cannot exclude the situation when a TBNV fails to grow and, thus, reach the forthcoming investment stages. In this case, the obtained results will have limited usefulness since the proposed model is able to provide TBNVs' evolution curve based only on the previous dynamics and under the perspective that a company will continue to take managerial actions in a similar to the previous period manner. So, to underline, from the macro perspective, the evolution of a TBNV while moving from zero to maximum valuation is not chaotic and has particular VC-driven stages. At the same time, it may have a chaotic character on the micro-level when moving from stage to stage.

**The reviewer's comment:**

“- Boundaries of organization. One of the central hypotheses of the OLC model is that the boundaries of organization are well-defined and given. Recent evolutions show that the boundaries are fuzzy and questioned. The notion of ecosystem blurs the boundaries of the firm and weakens the interest of OLC as it becomes less relevant to analyse the firm per se compared to the whole development of eco-system.”

**Response to the reviewer:**

Thank you for the comment. According to the OLC theory, the boundaries of an organization are well-driven and given, while the recent studies related to new ventures 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). In my understanding, the proposed evolution curve 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 that is mainly influenced by their sales-related dynamics but also by partnerships, participation in the technological clusters, and other network-related measures, it can be concluded that any – given or blurred – boundaries may be present in a particular organization and related to its particular stage of evolution. Thus, various phenomena (like organizational boundaries in the ecosystem context and predetermined character of the development) related to the previously developed OLC concepts may be revised from the perspective of the proposed data-driven TBNVs' growth model and the analytical division

on stages.

**The reviewer's comment:**

“How can we analyse the correlation with Google trends?”

The finding of the correlation is interesting. Considering the methodology with 6 rounds of founding and the proportion of platform and B2C, this is not surprising. What does web search traffic mean and what does it represent? When a start-ups is becoming unicorn, it drives more visibility and traction. How to interpret the correlation and to what extent can we infer causality?”

**Response to the reviewer:**

Thank you for this comment. In the context of business and management studies, Google Trends search query statistics are argued to reflect the interest of a public audience toward a product or a company, either a new or existing one. It was also demonstrated previously that these GT data are directly connected to future sales of a product (for reference, please, see part 3.2.2, or 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> and 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>). Considering TBNVs, it is also well-known that sales (or customer adoption) are positively and directly connected with a company valuation (for reference, please, see part 3.2.1 or 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> and other). So, higher the public interest – higher the sales or sales related measures like amount of the unique users of the website (i.e., the rate of adoption).

I can hypothesize that when search queries growth speed slows and changes from growth to decline, there can be two reasons for that: (1) a company is not interesting due to the problems with a product, or (2) a company is not interesting because the majority of the potential users know about it. So, the first case reflects the negative scenario, which may lead to a failure of a TBNV, while the second presents the ideal case of the maximum market penetration. However, since no previous studies have examined this behavior of the search query dynamics under the TBNVs' context, this logic should be treated consciously.

Yes, you are correct that for unicorns, the link between TBNVs' dynamics expressed by VC valuations and connected to them GT data is stronger than for non-unicorns. However, it was also evidenced that for non-unicorns, this correlation is high as well: mean Kendall's tau = 0.72 (for unicorns = 0.8), median Kendall's tau = 0.73 (for unicorns = 0.82). I can assume that everything depends on the product's popularity and the character of its diffusion (at the end, unicorns became unicorns because people skilled in the art (i.e., VCs) invested in them based on evidence that their products show strong traction among users and have high sales prospects). Indeed, the public audience, which is significantly bigger than VC investors and, thus, adds a greater proportion to search queries statistics, most likely does not know anything about the status of a company, whether it is a unicorn or “decacorn,” or “centaur,” or a “pony” in terms of valuation. I think for the end user, it is more interesting if a product of this TBNV fits her or his need and expectations. And I assume that if a product reaches this fit, the interest of users will demonstrate a continuous growth trend (which is of the interest for this thesis), while the peak rise caused by VCs will disappear as fast as it emerged. Thus, I believe a deeper examination of this question may be an interesting direction for future studies.

As it is known and was applied in the current research, correlation means that two measures evolve together in the same manner. Following the logic described above, both measures (VC valuations and TBNVs' related GT data) are, thus, most likely connected to sales or sales related-measures (the rate of adoption). In other words, we can assume that there should be a positive causal link between the *rate of adoption and search queries* and between *the rate of adoption and valuations*. The full causal link may look as follows: diffusion of a product leads to higher adoption expressed in sales or sales-related measures, which is indicated by the increase of search queries, and influences the TBNV's valuation. However, more complex situations are possible when interest in a company, as well as its valuation, grows with awaited future sales. This situation is mainly a reflection of a brand (of a founder or of a company) acceptance. Anyhow, I assume that the direct causal link between search query data and VC valuations is less likely. In the current study I do not imply any causality but aim to work on this question in future studies (please, see Chapter 9, point 2).

**The reviewer's comment:**

“Data analysis

Chap 3. It is surprising that data analysis is sequential rather than building an integrated model. In addition

you have longitudinal data and you analyse it as cross section. Without being a specialist in econometrics, it is surprising.

Chap 4. Your sample selection process p 99 may introduces a bias as you selection first data available in Google trends and with enough points. You introduce after different dimensions as Unicorn etc. controlling how the sample is reduced is important.”

**Response to the reviewer:**

Thank you for this comment. Since the goal of Study 1 was to analyze the validity of TBNVs’ related Google Trends data as a source of information able to reflect the growth dynamics of a particular venture (which is a temporal process due to its nature), the longitudinal design was understood as the logical and natural approach built in the studied TBNVs’ growth process. At the same time, I targeted to reach a higher external validity of possible results and therefore decided to employ cross-sectional techniques. Yes, there could be separate clusters of TBNVs, for which the correlation between companies’ related GT data and their valuation will deviate from the average values, and I plan to implement this kind of study further. In other words, to get the most generalizable outcome, here I selected the deductive approach with the mix of longitudinal techniques built in the nature of the growth-related data and cross-sectional methods to observe the wider picture. The sequential research, at the same time, was driven by the goal to understand the link between two sets of data and analyze its validity. An integrated model could provide more insights on causal links, but the character of the studied complex phenomenon and possible variance in the outcome would also decrease the external validity of the results making the research more a case study than a quantitative one. The creation of the integrated model is also possible in future studies.

As for the mentioned issue related to the selection process in Study 2, I can comment that it was limited by the research design and limitations of the GT data. First, I built the initial sample of companies founded in the taken boundaries, evidenced to experience the late-stage funding events, and attracted at least three series of VC investments. The first two selection criteria are intended to select companies with a higher probability of finishing the growth, while the last one targets to separate TBNVs from SMEs, which rarely attract external funding. Next, since the analysis required much time-consuming handwork and could hardly be automated at that phase, I decided to randomly select 500 companies while controlling the distribution of funding rounds by means of the Pearson’s chi-square test. The amount of funding rounds and its distribution across the sample looks like a well-fitting measure to control the valuation-related effects in the population. Since VC targets to finance the TBNVs’ growth, a faster-growing company will attract more series of investments than a population average. Therefore, controlling the distribution of this parameter provided an ability to avoid the possible bias related to the overpresence of successful or unsuccessful companies in the sample. One may be interested in repeating the results of the study on a full initial sample.

Next, since I targeted to employ GT data to analyze TBNVs’ growth dynamics, each company in the randomly built sample was tested for GT data availability and quality by the algorithm described in Appendix C of the thesis. Based on the various reasons (error caused by the similar term, absence of enough search query statistics, etc.), the quality of GT data was assessed and graded, and some cases were excluded from the study. In future research, one may be interested in examining the causes of the low data quality in some cases and assessing the limitations of the GT data application more precisely. That is to add, at this step, I’ve also controlled for the distribution of the number of funding rounds by the Pearson chi-square test: the probability that the randomly selected 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 (I added text explaining this in the section 4.3.1, paragraph 4). Therefore, it can be concluded that characteristics of the sample, which might lead to a bias in results, were controlled.

**The reviewer’s comment:**

“These are two important contributions. However, basic questions remain open: is OLC the relevant conceptual tool to describe TBNV’s growth? Is it relevant when ecosystems are becoming more important? Which are the competing concepts to describe growth? what is the explanatory power of OLC?”

**Response to the reviewer:**

Thank you for this comment. With all potential limitations, I assume that answer to your question is yes. In newly written paragraph 4, Chapter 5, I argue why (some parts of this text are already presented in the responses to your previous comments).

“However, is the OLC theory the most suitable organizational development theory for backing the proposed TBNVs’ growth model? I assume that 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. Thus, I believe, OLC theory is the most suitable framework for explaining the research results. Although 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 takes into account 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."

And to Chapter 6, paragraph 5: "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."