Shorter than we thought: The duration of venture creation processes

Jaehu Shim*, Per Davidsson

Australian Centre for Entrepreneurship Research, QUT Business School (Management), Queensland University of Technology, Brisbane, QLD 4000, Australia

ARTICLE INFO

ABSTRACT

This study aims to discern the duration of venture creation processes that end in successful venture emergence as well as those ending with abandonment. Discerning duration is essential to understand a process since every process unfolds over time. However, identifying correct venture creation duration is tricky because available panel data sets in entrepreneurship are length biased toward long duration due to left truncation. In order to adjust this problem, we applied Inverse Probability Weighting (IPW) and analyzed the venture creation duration of 1673 nascent ventures in the Harmonized PSED (Panel Study of Entrepreneurial Dynamics). The weighting analyses for the duration show that the emergence chance is maximized around three months after inception of the process, implying that venture creation in the typical case is faster and simpler than previously believed. Past three months the emergence chance decreases, whereas the abandonment hazard constantly increases over time and surpasses the emergence chance after seven months. The results show that weighing strongly influences duration time estimates and also can influence other estimates based on samples using the same type of sampling mechanism. Thus we conclude that IPW should be considered for all analyses using this type of data. Our study offers a possible interpretation of the different mechanisms of emergence and abandonment and provides guidance and insights for future studies.

1. Introduction

There is increasing consensus that entrepreneurship is a process (Bakker and Shepherd, 2017; McMullen and Dimov, 2013; Zahra and Wright, 2011). With this insight follows an interest in temporal issues, because processes unfold over time. One fundamental characteristic of venture creation processes is their duration – i.e., the amount of time it takes from inception of the process to its resolution either as successful emergence of the venture or abandonment of the effort before it has reached successful emergence. Although a few previous entrepreneurship studies address the duration issue (e.g., Kim et al., 2015; Reynolds, 2007, 2016), our results suggest that left truncation and the resultant over-sampling of cases of long duration (so-called length bias or length-time bias) may have led to considerable misrepresentation of what the venture creation process entails in most cases.

The purpose of this study is to provide more accurate stylized facts about venture creation duration by employing Inverse Probability Weighting (IPW) and to suggest plausible mechanisms behind these empirical facts. By identifying and explaining the venture creation duration, we can help answering these essential questions in entrepreneurship: "How long does it take to successfully launch a venture on average?", "Is there a peak duration that maximizes the chance of venture emergence?" and, similarly,

"Is there a critical period during which the risk of abandonment is particularly high?"

Several panel data sets such as PSED (Panel Study of Entrepreneurial Dynamics) have been established, which trace thousands of venture creation processes over time (Reynolds, 2009). These panel data sets provide information about the duration of venture creation processes. For instance, a descriptive analysis of duration of venture creation processes in the original, US-based PSED study showed that around 10% of nascent ventures had successfully emerged within 12 months, and this rate barely reached 30% after 60 months (Reynolds, 2007). This conveys an image of the start-up process being long and complex in most cases, which seems at odds with findings suggesting that proceeding at high pace increases the success chances of nascent ventures (Lichtenstein et al., 2007) and that often relatively few actions are essential for a nascent venture to start generating profits (Arenius et al., 2017). The previously reported long duration may be associated with length-time bias due to left truncation (Yang and Aldrich, 2012). In other words, nascent ventures that have longer duration have higher chances to be selected into the sample at any one point in time because abandoned as well as already successfully emerged ventures from the same cohort are no longer eligible for inclusion in the sample (Davidsson, 2016: 125). It is therefore not sufficient to analyze the duration of venture creation processes "as they are" in the data sets.

Some prior research has noted this potential problem and tried to mitigate it by including only cases that were initiated more recently (see, e.g., Delmar and Shane, 2003). This reduces the left truncation issue but does not completely eliminate it. The practice also means that many cases are excluded from the analysis, and it has also been criticized on other grounds (Honig and Samuelsson, 2014). Alternatively, the Conditional Likelihood Approach has been suggested as a statistical solution to length-biased samples (Yang and Aldrich, 2012). This approach is applicable to the estimation of the hazard rate of a focal event in survival [or event history] analysis, but not applicable to other analyses. We instead employed IPW as a suitable adjustment for the data sets that have left truncation issues. Applying IPW, we analyzed the five-cohort, outcomes-harmonized data set ("Harmonized PSED"; Reynolds et al., 2016) in order to establish credible empirical facts about the duration of venture creation processes.

Attention to duration is also important for other methods-related reasons. For example, measures of early-stage entrepreneurial activity like the Total Early-Stage Entrepreneurial Activity (TEA) index of the Global Entrepreneurship Monitor (GEM) (Kelly et al., 2016) can be misinterpreted if differences in duration are not considered (Bergmann and Stephan, 2013). If the average venture creation duration in a country is relatively long, the country will have a high TEA without having a high inflow of successfully emerging ventures. Further, for addressing research questions about venture creation attempts rather than about individuals qualifying as "nascent entrepreneurs" at a particular point in time, the GEM survey is also subject to length-time bias due to left truncation. Additionally, with the understanding of distributions of the venture creation duration, researchers can apply a correct duration model, since each parametric duration model assumes a specific distribution (Box-Steffensmeier and Jones, 2004).

Gaining insights into process duration is also of theoretical interest. For theories with a strong temporal element, duration is a central issue. Examples include theories about real options reasoning (McGrath, 1999), decision speed (Baum and Wally, 2003), first-mover advantages (Lieberman and Montgomery, 1988), and time compression diseconomies (Dierickx and Cool, 1989). The importance of correcting for sample bias may also extend to research questions that do not directly concern duration. If the characteristics and/or causal relationships pertaining to cases ending up having long duration differ from those of short-duration cases, *all* sample descriptions and variable relationships will be affected. If so, the descriptions and coefficient estimates from an unweighted analysis will not correctly represent the underlying population of ventures.

Our study contributes to entrepreneurship research in several ways. First, we introduce IPW to adjust for length-time bias due to left truncation in data sets using the PSED-GEM sampling mechanism (Reynolds, 2009). We thereby establish credible empirical facts about the duration of the venture creation processes which alter received views and highlight a need and solution for correction that might extend to any descriptions and relational analysis based on the used type of nascent venture data. Second, we delve into the peak duration of successful venture emergence as well as the change of abandonment hazard over time. Third, we suggest possible mechanisms behind venture creation duration, providing theoretical and practical insights into the temporality of this phenomenon. Based on these contributions, our study can serve as methodological guidance and theoretical inspiration for future studies.

2. Inverse Probability Weighting (IPW)

IPW can be utilized for a sampling procedure if each unit's sampling probability is not constant, but there is a good predictor for the selection probability of each unit. General random sampling procedures assume "missing at random" (MAR). In contrast, IPW extends the assumption and assumes that the sampling probability for each unit can be predicted from one of the unit's variable values. Thus, for the application of IPW, it is critical to select a reasonable predictor of a unit that is systemically associated with the unit's sampling probability. Specifically, IPW has four assumptions: (i) The weights of units denoted by vector w_i can be predicted whenever the units are included in the sample, (ii) there is another vector z_i that denotes inclusion probabilities of the units, (iii) each z_i is greater than 0, and (iv) each z_i is observed whenever the corresponding w_i value is included (Wooldridge, 2007). Practically, IPW weights each sampling unit by the inverse of its probability of being included in the sample, i.e. $w_i = 1/z_i$ (Seaman and White, 2013). As an example of IPW in entrepreneurship research, the inverse of venture team size (e.g. 1/2 for two-member teams) has been assigned for estimating the frequency of ventures that has a specific team size based on a random survey of individuals such as PSED (Reynolds et al., 2016; Yang and Aldrich, 2012).

In this study, we suggest that a case's duration in the process is a good predictor for the selection probability of the case if the cases are randomly sampled at the same time. Fig. 1 illustrates five hypothetical nascent ventures' varying duration of venture creation processes. In this figure, if sampling is conducted between t_0 and t_1, all five cases will be included in the data set, whereas sampling is conducted after t_4, only *NV5* will be included in the data set. This demonstrates that ventures like *NV5* have a sampling probability