New Venture Creation and Growth

07/12/2022

Saul Estrin[a], Andrea Herrmann[b], Moran Levesque[c], Tomasz Mickiewicz[d], Mark Sanders[e]

[a] London School of Economics, London, UK, S.Estrin@lse.ac.uk
[b] Radboud University, Netherlands, Andrea.Herrmann@ru.nl
[c] Schulich School of Business, Canada, mlevesque@schulich.yorku.ca
[d] Aston University, Birmingham, UK, t.mickiewicz@aston.ac.uk
[e] Maastricht University, Netherlands, m.sanders@maastrichtuniversity.nl

The authors acknowledge the time to develop these ideas at the 2020 Oxford Residence Week for Entrepreneurship Scholars, as well as comments during an early presentation from numerous participants. The paper was also presented at Utrecht University and at Aston University and we thank participants there for useful suggestions. Per Davidsson also made extensive helpful comments. This project received funding from the FIRES-project (http://www.projectfires.eu/) under the European Union's Horizon 2020 Research and Innovation Program, under grant agreement number 649378, for collecting the data.

New Venture Creation and Growth

Abstract

This study investigates the role of new venture creation in innovation. We model new venture creation to derive some testable hypotheses and test these for a sample of new ventures in information and communication technology (ICT) sector, using a novel and unique dataset. In the earliest venture creation stage, new firms face fundamental trade-offs between introducing innovation, getting to break-even point quickly, and reaching high levels of revenue. We first present an analytical model that illustrates why such trade-offs arise. We then empirically analyze the earliest stage venture creation process applying a multi-input, multi-output stochastic frontier model to account for unobserved heterogeneity across new ICT ventures in four different countries. This approach allows us to identify a strong trade-off between innovativeness on the one hand and speed to break-even point (profitability) and revenue on the other. We find that proprietary resources, founders' equity and labor (also known as "sweat equity"), are the main drivers of new ventures' multidimensional success. Our paper thus proposes that new venture creation is a key bottleneck in the process of innovation and opens an important line of research to explain the (large) unexplained variation in efficiency across new ventures in ICT.

JEL Codes: D22, L26, L29, O31

Keywords: Entrepreneurship; new venture creation; stochastic frontier analysis.

I. Introduction

New venture creation (NVC) is modeled rather simplistically in the theory of economic growth. Although Schumpeter's work stressed NVC as the source of change and sustainable growth in a market economy (Schumpeter, 1934; Koellinger and Thurik, 2012; Acemoglu and Robinson, 2012; Haltiwanger et al., 2013), macroeconomic growth models have not problematized the process of new venture creation. Instead, they focus on either its drivers (new knowledge creation) or its consequences (innovation and growth). The neoclassical Solow-Swan model ignored innovation. Modern ideas-based endogenous growth models (Romer, 1990; 1994; Jones, 1995; 2005) put innovation at their cores. But even the subclass of "Schumpeterian" growth models (Segerstrom et al., 1990; Aghion and Howitt, 1992; Dinopoulos and Thompson, 1998; Aghion et al., 2014) have viewed the process of new venture creation as a trivial step in the process (Acs and Sanders, 2012; 2013). In all these models, profitable opportunities to innovate automatically find an entrepreneur who creates a new organization, in the same way as all vacancies in the labor market get filled and all investment opportunities find an investor in equilibrium.

In this paper we want to zoom in on the venture creation process. To bring venture creation into the mainstream models of innovation and growth, we take inspiration from the literature on entrepreneurship. That literature studies how entrepreneurs create new ventures to introduce new goods and services to the market (e.g., Lucas, 1978; Evans and Jovanovic, 1989; Baumol, 1990; Levine and Rubinstein, 2016; Parker, 2018) and there is ample evidence that this process can be an important bottleneck in turning promising ideas into innovations. If NVC is an important bottleneck in the innovation process, a micro-economic understanding of what resources make entrepreneurs successful in NVC will complement our macro-economic understanding of economic change and growth.

In this paper we therefore start by outlining a theoretical framework to understand the process of new venture creation in the form of a very simple mathematical model. We assume that, due to asymmetries in information, only founder equity can finance the activities before break-even and founder labor is necessary to bring down the uncertainty over a new venture's value proposition (innovativeness). This idea builds on the work by Evans and Jovanovic (1989) who argued that entrepreneurial venturing can be understood as the founding team learning about the productivity and profitability of the venture as they engage in the activity. We then link this idea to a standard idea-based growth model set-up (e.g. Romer, 1990; Aghion and Howitt, 1992), where post-innovation monopoly rents incentivize the innovators. Under those empirically reasonable assumptions, we show that founders will choose to hire labor during the NVC to maximize the (externally verifiable) value of their venture at break-even. This simple model generates some straightforward and intuitive hypotheses on the trade-offs that founders face between the time to break-even, the revenue when they do and the innovativeness of their idea. We then proceed to build an empirical model to test these hypotheses in the data.

We thus conceptualize NVC as the transformation process by which entrepreneurs acquire and organize resources to achieve the objective of establishing a viable new enterprise (Davidsson, 2016). NVC is therefore not a costless activity and inefficiencies in the process cause scarce resources to be wasted; a point noted by de Meza and Southey (1996). The rapidly expanding management literature on NVC (e.g., Shane and Venkataraman, 2000; Mcmullen and Dimov, 2013; Shepherd et al, 2019; Davidsson and Gruenhagen, 2020) has studied the process by introducing different outcome measures (e.g., survival, profit, employment growth, investment, innovation) and has added a plethora of control variables, such as founder (team) characteristics, resource inputs, and environmental variables in different permutations. Entrepreneurship scholars have also analyzed the motivations and characteristics of entrepreneurs, including their cognitive and social abilities (e.g, Lazear, 2004; Hartog, van Praag, van der Sluis, 2010). Davidson and Gruenhagen (2020) provide an overview of this literature. The 24 key papers in their list show how the field that has focused on NVC processes as a sequence of steps over time. Notably, the empirical contributions by Reynolds and Miller (1992), Carter et al. (1996), Davidson and Honig (2003), Delmar and Shane (2004), Foo et al. (2009), Held et al. (2018) show a clear trajectory of collecting time stamped data and of analyzing venture creation as a process that takes time. But this approach remains limited in terms of theory framing: the empirical models remain ad hoc and do not link outputs to inputs utilizing a quantifiable production function approach. Moreover, by applying standard statistical methods, the estimates of the impact of (measured) factors on NVC outcomes are necessarily the *average* effects for the *average* venture at the *average* levels and intensities of input use. These papers thus (implicitly) attribute the (very large) deviations from the average to random noise and symmetric measurement error, while we know that a large part of these errors arises from (unobserved) heterogeneity across NVC processes.

We concentrate our attention on the earliest stage of the venture creation, operationalized here as the time between the registration of the new company and attaining break-even; an approach which is close to Reynolds (2018). Furthermore, we propose that Stochastic Frontier Analysis (SFA), developed to study productivity performance across heterogeneous firms, can provide new insights into NVC process. Traditional methods (implicitly) assume all observations come from the same data generation process and that randomly distributed measurement error explains deviations from the true model. SFA replaces that assumption with a less restrictive assumption on the distribution of unobserved heterogeneity. We then illustrate our method using time stamped data on new venture creation processes in the information and communication technology (ICT) sector, taken from the Perfect Timing Database (Herrmann, 2019; Held et al., 2018; Herrmann et al. 2020). The data allow for the analysis of multiple dimensions of both NVC objectives and of resources and we identify the trade-offs among three objectives and alternatively two or four inputs. Our method could be extended on both the dimensionality of the objectives/outputs and the resources/inputs, thus allowing future researchers to build on our model and unpack the NVC-process in a stepwise manner.

Using this approach, we find that the trade-offs in outputs between venture innovativeness and revenue size, as well as between innovativeness and speed-to-break-even are strong, whereas the trade-off between revenue size and ventures' speed-to-break-even is much weaker. Moreover, and in line with the resource-based view of the firm (Barney, 1991; Alvarez and Barney, 2005), we find that, of the labour employed and capital invested, it is especially the "founders' sweat" (Bhandari and McGrattan, 2021) and equity that contributes to better outcomes.

We thus make four contributions in this paper. First, we show in an analytical model that, at the efficient frontier, NVC should reveal trade-offs between time to break-even, revenue at break even and innovativeness. Moreover, our model shows that these trade-offs are less stringent when founder equity and labor are more abundant. We then confirm the existence of these trade-offs between entrepreneurial objectives, especially between those related to long-term benefits (associated with innovation) as against the short-term financial gains. The proposed framework also allows us to estimate quantitatively sensible substitution and output elasticities for the inputs considered. Third, we apply an empirical method for analyzing the transformation process of multidimensional inputs into multidimensional outputs across heterogeneous units during the earliest stage of NVC and demonstrate that it produces valuable insights. Fourth, we present and use a new, detailed dataset on the earliest stage venture creation processes, the Perfect Timing Database. This unique data source, published with this paper, allows researchers to link precise details on acquisition of financial resources and team formation to speed, profitability, and innovativeness of NVC in the ICT sector across four countries. We hope this dataset may serve as a template for further data collection and analyses in this area.

The remainder of our paper is structured as follows. Section 2 presents our theory and model. Section 3 presents our empirical methodology and data. Section 4 presents our empirical results, and Section 5 concludes.

II. Theory and Model

II.1 Basic Set-up

In line with our empirical definition of the firm formation process, we assume a venture creation process starts with the registration of a firm. At that point in time the founder(s) commit an amount of founder equity, e_0, to the venture and starts supplying its exogenous supply of labor input, f, inelastically to the venture. From that point onwards the founder(s) can decide to hire external labor, $h(t)$ at an exogenously given wage rate, w. The venture has a technology to create revenue that we characterize by:

$$R(t) = I(t)^2 r(e(t), f + h(t))$$

Where we assume that $r(.)$ is positive and concave in both arguments and $I(t)^2$ is a draw from a normal distribution with mean μ and variance $\sigma(t)^2$, where $\sigma(t)^2 = \mu P\left(\int_0^t f(\tau)d\tau = tf\right)$, where we assume $P(.) > 0$, $P'(.) < 0$, $P''(.) < 0$, such that the variance falls with time and founder labor input to 0 as either goes to infinity. This process captures the Evans and Jovanovic (1989) idea that working in their venture reveals the innovativeness of the venture to the founders. Total costs for the venture are given by $C(t) = wh(t)$ and the budget constraint for the venture is:

$$e(t) = e_0 + \int_0^t R(\tau)\, d\tau - \int_0^t C(\tau)\, d\tau \geq 0$$

We assume that the founders want to maximize the value of their venture at the point of break-even, $R(t) = C(t)$. This value is given by:

II.2 Equilibrium

II.3 Hypothesis Development and Trade-Offs

III. Empirical Method and Data

III.1 Empirical Method

In the model above we have derived our hypotheses under the assumption that NVC is an efficient process and venture creation is a predictable and plannable process. Of course, in reality, every new venture is unique and develops under deep uncertainty, also on behalf of the founders. The success of the venture creation process depends to a large extent on the complex and unpredictable interaction between founders' talents and resources, the technology that is being explored and the environment in which the venture is launched. We cannot usefully model this as a process where the decision maker can rationally and efficiently employ resources by setting market prices equal to (expected) marginal value products. Instead, the uncertainty must be reduced and although external labor can be employed at market wages, it is the proprietary resources, committed to the venture by the founders in a leap of faith, that drive the process. When testing the hypotheses derived above, we need to take into account that in the data, all heterogeneity that is caused by the complex interaction between such resources, the technology and the environment of the venture, cause heterogeneity across NVC processes, potentially increasing their time to market or reducing their revenue for given levels of innovativeness. Such "inefficiency" is to be expected. We do assume, however, that it does not affect the trade-offs at the efficient frontier and will model this as a level shock to the NVC.

To introduce our empirical method for doing so, it will be helpful to introduce some formal notation. If we conceptualize a NVC process, I, as setting up a new organization, then, at that admittedly high level of abstraction, it is a transformation process in which an entrepreneurial team acquires and (re)arranges a set of N resources that we can represent by a Nx1 input vector $\mathbf{x^N_i}$.

These inputs are used to achieve a set of M objectives that can be represented by an Mx1 vector of $\mathbf{y^M}_\mathbf{j}$. We then follow production theory in assuming that the mapping of inputs into outputs is stable over all i and can be described by a mathematical function, $\mathbf{y^M} = f(\mathbf{x^N})$. Empirical research then typically proceeds with the implicit assumption that all observations in a dataset (in our case venture creation processes) are drawn from the same data generation process. That is, all venture creation processes follow a common transformation process, and the variation across observations can be used to identify the parameters of that process by assuming observations are randomly distributed around the true model. Assuming that inputs are (log) linearly combined (a Cobb-Douglas specification) into a single objective or performance measure, one would, for example, estimate:[1]

$$\ln y_i = \alpha + \sum_{n=1}^{N} \beta_n \ln x_i^n + \varepsilon_i \tag{1}$$

where i indexes the NVC processes over which we generalize. The empirical literature on NVC processes then tries to identify relevant inputs (such as founder team characteristics, environmental variables and investor inputs) by estimating the output elasticities β for inputs, which are theorized to affect venture creation, in a dataset of nascent ventures. When we also reduce the number of inputs to 1 for exposition purposes, Figure 1 below shows the familiar graph from Statistics 101.

{Figure 1}

[1] It is possible to estimate more general specifications like the Constant Elasticity of Substitution or Translog specification, that allows for the elasticity of substitution between inputs to be different from 1 or even dependent on the level of inputs used. In this paper we keep that part of the modelling simple and develop our argument around a standard Cobb-Douglas specification. What limits us in pursuing more complex models is primarily the size of the dataset.

Regression analysis fits a curve through the data points available and interpret all points away from that curve as randomly distributed errors in measurement, or specification. More sophisticated studies could identify the marginal contribution of inputs to venture creation in regressions where a dummy, or variable, distinguishes between different types of new ventures and estimate the difference in output elasticities between these groups –- again, interpreting the residual as a random, normally distributed error.

Both approaches are problematic, however, if we cannot assume that the underlying data generating process is similar across all units of observation. Importantly, the literature (Gartner, 1985; Davidsson and Gruenhagen, 2020) has frequently made the point that the assumption of homogeneity is particularly problematic for venture creation processes. Indeed, some would go so far as to suggest that every venture creation process is unique and idiosyncratic; this would imply generalization across these processes is impossible and one cannot learn from comparing across NVC processes. This contrasts with traditional regression analysis, which assumes the same data generating process across all observations but also that deviations from the true model occur only through random measurement error. We seek a middle path between these views.

A way to recognize the role of heterogeneity in NVC processes, while at the same time learning from generalization across instances of venture creation, is to model individual venture creation processes as yielding different outcomes for the same vector of inputs. In production theory, scholars have developed Stochastic Frontier Analysis (Farrell, 1957; Kumbhakar and Lovell, 2003) just for such cases. Unlike the baseline Figure 1, Figure 2 shows how, using the same observations, a stochastic frontier model separates between the firms at (area A) and below (area B) an "efficient frontier". The slope of the frontier represents the output elasticity of the input, whereas the model allows for observations to lie below the line for a host of (unobserved) reasons.

{Figure 2}

By making some assumptions on the distribution of the additional, one-sided error term, the model to be estimated is now:

$$\ln y_i = \alpha + \sum_{n=1}^{N} \beta_n \ln x_i^n + \varepsilon_i - v_i \tag{2}$$

where the additional, error term v_i is strictly positive, truncated normal, exponentially or half normally distributed, and measures the vertical distance from observation i to the maximum attainable output at the frontier. One advantage of this approach is that the output elasticities are thus estimated at the frontier. That is, in our simple one-output example, we estimate the marginal contribution of input factors to NVC output amongst the firms that attain the highest levels of output in the sample.[2] The ventures at the frontier are also most likely to be constrained by the measured inputs in trying to achieve the measured outcomes.

As can be verified visually by comparing Figures 1 and 2, the estimated parameters of the transformation process can differ significantly between these methods. This is a consequence of the bias created by estimating the parameters using data points for which, for unobserved reasons, the input constraints were not, or less binding on the outcomes considered. If these (unobserved)

[2] More precisely, all observations are used to estimate the slope of the frontier, but the estimation procedure takes into account that not all observations are at the frontier. The assumption here is that all observations face the same output elasticity (slope), but need not have the same intercept in their production function. This gives more weight to the observations close to the frontier in estimating the common output elasticities, as their remaining distance to the regression line will reduce the likelihood function most.

external factors, for example different initial conditions, explain why the ventures in area B could not operate at the frontier, then the missing variable bias can be addressed by accounting, in our estimation method, for the unexplained distance from the frontier. If we do not allow for this, missing variable bias could potentially affect the estimated trade-offs and substitution elasticities in the venture creation process. The traditional estimation methods therefore yield biased and imprecise results.

The Stochastic Frontier Analysis (SFA) method has the advantage that it does not require the consideration of all possible factors in a complete model. Without controls, the distance to the frontier captures a significant share of the unobserved heterogeneity and isolates the bias that would otherwise affect our parameter estimates. Moreover, the one-sided error term gives us an indication of the importance of unobserved heterogeneity. A final, and in the case of venture creation processes, critical advantage of this approach is that it allows for multidimensionality, not only in the input vector, but also in the outcome vector. We know from the literature reviewed by Davidsson and Gruenhagen (2020), that founding teams can have a variety of different objectives. To then estimate the importance of inputs in the process for achieving one or the other objective, misses the point that the same resource inputs may have been employed to achieve objectives other than the one being modelled. Again, that will bias estimated elasticities (down) and while this is complicated to capture in standard regression models, multiple output frontiers are a straightforward extension of the single output SFA model in equation (2).

Thus, building on the single output production frontier, Appendix A shows that, under (mild) additional assumptions, notably that the frontier is homogeneous of degree one in outcomes, one can also estimate a multiple output model:

$$\ln y_i^1 = \alpha + \sum_{n=1}^{N} \beta_n \ln x_i^n + \sum_{m=2}^{M} \gamma_m \ln \frac{y_{it}^m}{y_{it}^1} + \varepsilon_i - v_i. \quad (3)$$

As in the single objective case, the variance of v over the total variance can be interpreted as a measure of importance of unobserved heterogeneity in factors that prevents venture from achieving its objectives; that is of its inefficiency (Kumbhakar et al., 2015). In equation 3, the distance to the frontier is $-v_i \leq 0$, which is not assumed to follow a standard normal distribution, $N(0, \sigma^2)$. Rather its distribution can be modelled alternatively as half (or folded) normal, truncated normal, or exponential. ε_i is the usual mean-zero normally distributed noise component, that is independently and identically distributed (Kumbhakar and Lovell, 2003). Because the distribution of v_i is asymmetric, so is the distribution of the composite error term $\epsilon_i = \varepsilon_i - v_i$.

Equation 3 is the frontier model we use to analyze which resources (inputs) contribute towards new ventures' achievement of their objectives.

III.2 Data

To estimate our model, we utilize data on individual NVC processes, and to illustrate the generalizability of our method to multiple outcome and input dimensions, we use observations on three parallel outcomes, as well as a number of relevant input variables. We draw our proprietary data from a unique firm-level dataset containing information on the start-up processes of 331 usable observations on nascent ventures in ICT, collected with an explicit focus on how their activities were sequenced. Founders were interviewed about their activities since the start of the venture creation, in the period which we labelled above as the earliest stage NVC process. The interviews were carried out in two waves between 2011 and 2018, based on computer-assisted telephone interviews in US, Germany, UK, and Italy. The population considered includes ICT-based ventures of all legal forms apart from sole proprietorship, registered between 2004 and 2014. From this population, founders were randomly selected and invited to participate in an interview about the venture creation process. A structured interview guide was developed for a survey-based interview on Computer Assisted Telephone Interviews (CATI). The interview guide made it possible to trace how venture creation processes evolved on a monthly basis. The questionnaire records the venture details and circumstances of venture creation, such as the venture's *location*, *year of registration*, *legal form*, *business idea (product or service)*, *novelty*, and *degree of innovativenes*s. It also identifies its start (in our case: registration) and end date (in our case: achieving a break-even point) of the earliest stage NVC. In line with the process-oriented entrepreneurship literature (Reynolds, 2018; Davidsson and Gruenhagen, 2020), we use the registration date as the start date of venture creation. The end of the NVC process was defined as the point in time when the new venture generates *profits* for more than 3 months.[3] If this had not occurred by the date of the interview, the process of venture creation was categorized as ongoing, to a maximum of 84 months. The shortest venture creation process in the sample is three months. Furthermore, the questionnaire traces – on a monthly basis – how many *founders*, *employees*, and *service providers* worked for the venture on a part-time or full-time basis respectively, and when. It also reports the different financial sources that the venture acquired that we categorize into *founder capital* and *loans and subsidies/grants.*

[3] It is possible to estimate profit frontier models that explicitly model profit maximizing behavior (Kumbhakar et al., 2015), but this would require information about prices of inputs and outputs that we do not have.

We next describe how we constructed our dependent variables (objectives), our resource inputs, as well as the explanatory variables in new venture creation processes. More detail can be found in the online material with this article where we include the Stata™ do files we have compiled to construct our variables and generate the results.

The dependent variables in equation (3) are the three objectives of the new venture. For the *Speed to Break Even Point (SPD),* we calculate the time (in months) that elapsed between the venture registration and the first month of three consecutive months of non-negative profits. As is standard in the frontier analysis, we then cut off the extreme values (top and bottom 1%) so they do not distort the frontier location, and computed speed to break-even point as:

$$SPD=1/MTP \tag{4}$$

where MTP is months to non-negative profit. SPD would then take a value of 1 for the firm that has the smallest number of months to non-negative profits (about 1) and will approach 0.01 for firms that score the highest number of months. We present descriptive statistics in Table 1, and the resulting variable in a histogram in Appendix B (Figure B1). The variable has a right skewed distribution, suggesting that many firms are close to the slowest firms in completing their firm formation process.

Our proxy for the second objective, *revenue*, is the level of revenue at the end of the venture creation process as defined above, labelled *REV*. To ensure that our variable is well behaved, we again drop the outliers above 99% and below 1% respectively. Figure B2 provides the histogram of this variable, both in the original form and in the logarithmic transformation we will use in estimations.

Our proxy for the third objective, *innovativeness of the venture (INN)*, is constructed as follows. In the survey, founders were asked to assess whether their product or service was new to the customers (CUS: 1=yes, 2=no), indicate the novelty of their product or service (NOV: 1=radical, 2=incremental, 3=replicative), and to list if the firm was (1=yes, 0=no) developing a new product (D^{PT}), process (D^{PS}), service (D^{SE}), technology (D^{TY}) or application (D^{AN}) and selling the product or service abroad (D^{EX}). Based on their answers, we defined our proxy for innovativeness as:

$$INN=1+99*[(2-CUS)+(3-NOV)+D^{PT}+D^{PS}+D^{SE}+D^{TY}+D^{AN}+D^{EX}]/9 \qquad (5)$$

The variable takes a value of 100 for firms that have the maximum score on all components in the index and takes a value 1 in the opposite case. Figure B3 presents a histogram. Our INN variable has only 9 possible outcomes by construction. This, however, is enough to treat this variable as more or less continuous for the purpose of our estimations. The distribution is relatively flat, indicating that many new firms in our sample are not very innovative. Given that we have sampled newly founded firms randomly, we would expect this from the empirical evidence.

To estimate equation (3), we also need to specify the inputs used in the production process. We follow the standard micro-economics perspective, largely shared by the entrepreneurship literature (Casson and Yeung, 2008), to classify the inputs needed by new ventures into two major categories: labor and capital. The standard model of production posits that increases in both categories of input, labor and capital, allows the firm to attain higher levels of its output (Estrin et al., 2013). To measure these inputs, we use the labor and the financial capital (our proxy for capital) utilized between the moment of registration and the end of the commercialization phase.

[TABLE 1]

4. Results

We first estimate the model in equation 3 using two 'naïve' estimators: corrected ordinary least squares (COLS) and corrected median absolute deviation (CMAD) estimators (Kumbhakar et al., 2015). The corresponding deviations from the frontier $\varepsilon_i \in [0,1]$ are presented in Appendix B. For COLS, the frontier is defined as if there is no measurement error or noise around it, and by construction, these estimates are sensitive to outliers (Kumbhakar et al., 2015). We can see this when comparing the deviations from the frontier in COLS with those obtained from the CMAD model, which is less sensitive to outliers. In the latter model, the distribution is more skewed. Nevertheless, in both models the distribution of deviations from the frontier resembles a truncated normal distribution: most firms are close to the frontier, but the density decreases in close proximity to the frontier. This is illustrated in Figure B4, which also shows tests of the distributional assumptions; we show that our data satisfy the conditions needed to justify the use of a frontier framework. The interpretation we might offer with this result is that some ventures, arguably those that care most about the outcomes we measure, operate at the frontier, but a substantial mass of observations lies below the frontier. If all ventures were to use labor and capital efficiently to achieve speed to break-even, revenue and innovativeness, then we would expect to see the mass of firms clustered close to the frontier. The fact that they do not, gives us an indication either of inefficiency or that other objectives and/or other inputs may have played a role in explaining why not all firms operate at the estimated frontier. At the same time, for those that do, we may estimate the outcome elasticities for the inputs as well as trade-offs between the observed objectives using our stochastic frontier model.

Table 2 presents the estimates of the frontier model where we first utilized the two inputs, labor and financial capital, but estimated the coefficients utilizing alternative distributional assumptions. This represents our benchmark model. The first two columns represent results of the COLS and MAD models. The next three present the estimates with an assumed half-normal, truncated normal, and exponential distribution for the one-sided residual distribution, respectively.

From the estimates of equation 3, we find the expected negative signs on trade-offs between the output dimensions to be highly significant. To be specific, for the exponential distribution, we find the estimated coefficient on the trade-off between speed to break-even and innovativeness equal to -0.500. Similarly, we find the trade-off between speed to break-even and the level of initial revenue with the estimated coefficient equal for the exponential distribution to -0.0425.[4] In a post-estimation test on the difference of coefficients, we find the difference is significant at 0.1% for the COLS and half normal distribution (Models 1 and 3 in Table 2) and at 1% for the MAD (Model 2), with weaker results at 10% for models based on a truncated and exponential distribution of distance from the frontier (Models 4 and 5).

[TABLE 2]

We find strong general support for the idea that there is significant unobserved heterogeneity in start-up performance; Distance from the common frontier accounts for about 75% of the variation

[4] We cannot directly test the trade-off between initial revenue and innovativeness. This trade-off has to be excluded to avoid perfect correlation. However, by implication given the sign of the other two trade-offs it must also be negative, as we also verified by choosing the other outcome variables as our benchmark/dependent variable and running alternative models. No other results were affected.

in outcomes across new ventures. This 75% is basically the measure of our remaining ignorance; it is the unexplained residual.

We next use our data to analyze in more detail the impact of factor inputs by taking more careful account of the heterogeneity in both capital and labor inputs. Thus, we distinguish between two forms of labor input: that of founders, and of hired labour (employees and service providers) as inputs to the venture creation process.[5] Furthermore, we analyze whether some forms of capital are more valuable than others, with a two-way categorization into founder equity, and other forms of finance (debt and grants).[6]

Starting with labor, while increases in all labor inputs should allow the new venture to attain higher levels of all three objectives, the literature highlights the need to distinguish between the entrepreneurial team and hired labor in the representation of the labor input (e.g., Bolzani et al., 2019). Founders (i.e. owners and managers) may have greater directly relevant capabilities for the new venture and may also be more strongly motivated and therefore work more productively than paid employees or hired service providers (Santos and Carden, 2018). Alternatively, following the logic of resource-based view (RBV) (Barney, 1991; Alvarez and Barney, 2005), it is the founders' labor that is firm-specific and hard to source from outside, whereas hired labor constitutes a market resource that is scarce, but does not constrain the choices of a venture at the frontier.

Similarly, we also expect different types of capital to have different effects on the objectives achieved. We start from the idea that financial inputs will have a greater impact on output in the

[5] We also experimented with models where we further distinguished between labor input and externally hired services. The differences in coefficients between the two were insignificant, while labor input of founders remained significant. We therefore report the more parsimonious models.
[6] Similar to labor, further distinctions in finance proved insignificant, while equity remained highly significant, which again led us to report the more parsimonious model.

form of equity than of debt. Equity finance, especially founders' equity investment, implies that there is incentive compatibility between the providers of finance and the management of the venture. Both downside risks and upside gains are shared equally. In contrast, when the new venture takes on external debt, there is asymmetry in the gains and losses because debt is a fixed financial-cost contract. This implies that providers of debt face a potential moral hazard problem, because the borrowers may gamble. As a result, debt providers usually insist on collateral from the debtors to protect themselves by securing their loans and debt finance comes with a higher risk of foreclosure by banks[7]. Thus, depending on the way it is secured, debt may lead to too little or too much risk taking and is less likely to lead to an optimum level of risk taking than with equity finance. Government grants, although formally equity, face similar problems, if the granting bodies are held accountable for how the money is spent (Parker, 2018).

Appendix C explains how we construct the categories of labor and of capital inputs. Our dataset provides information on the first five founders, employees and service providers and we are also able to compute founder equity, debt, and grants that were invested in the NVC process. In the estimations that follow, we utilize the same estimation methods as in the benchmark model. The results are reported in Table 3.

[TABLE 3]

We find support in Table 3 for the view that there is significant variation in the impact of different categories of factor inputs on the new venture's desired outcomes. We find that labor input by founders, and capital input through (founder) equity, have a significantly greater effect

[7] Entrepreneurs typically collateralize on their real estate assets. As Parker (2018) notes, this may induce them to take too little risk.

than the other categories of each input. Thus, we find that labor input by founders has a highly significant positive impact on the output frontier (p<0.001 in four models, and p<0.05 in MAD model), in contrast to that of service providers' and hired employees' labor (coefficients are always insignificant). To evaluate the difference between the two forms of labor formally, we need post-estimation tests of the differences in coefficients. These are presented in two rows of Table 3 located below the coefficient's rows. Correspondingly, these chi-square tests values of the difference in coefficients are highly significant (at p<0.001) in all models but MAD (p<0.05). This confirms the view that the labor effort of founders in the early stage is critical to the performance of the new venture.

There is a parallel result for capital: the p value for the equity coefficients is always below 0.001 in models 1 to 5 reported in Table 3. In sharp contrast, the coefficients on other forms of capital are always insignificant. Consistent with this, the chi-square tests on the difference in coefficients indicate that founders' equity finance matters more for entrepreneurial outputs in the early phase than external equity, debt, or grants.

5. Discussion

The results confirm the importance of 'sweat equity' in line with recent findings by Bhandari and McGrattan (2021). Likewise, the findings are also consistent with the RBV (Peteraf, 1993; Rumelt, 1984; Wernerfelt, 1984), in that it is the firm specific proprietary inputs that matter most for the ventures that operate at the frontier. In contrast, factors that can be hired or attracted in more or less open markets do not constrain the venture creation process as they can be adjusted to fit the ventures' needs and hired or acquired up to the point that marginal costs equal marginal benefits.

More generally, we propose a novel modeling approach to help us gain a deeper understanding of the earliest stage NVC process. We follow the literature and model the process of venture creation as a transformation process over time in which resources constrain the degree to which entrepreneurs can achieve competing objectives. Our approach allows us to link strategic choices in the allocation and acquisition of resources to the achievement of entrepreneurial objectives, and to systematically research the remaining sources of heterogeneity in the performance during the venture creation processes.

We identified three entrepreneurial objectives, each critical in the earliest stage of the new venture creation process: speed to break-even point, the initial level value of revenue when the break-even point is hit, and the innovativeness of the venture's offering. Our first set of results confirm the existence of trade-offs at the frontier between these objectives. At the frontier, we identify negative trade-offs between speed to break-even and the size of the initial revenue; between speed to break-even and innovativeness; and between the size of the initial revenue and innovativeness. We also find that the coefficient on innovation-speed to break-even is greater than that on revenue-speed to break even, suggesting that the former trade-off is more binding than the latter. This implies that, in terms of the entrepreneur's strategic allocation of time and other resources between these objectives (Levesque and Stephan, 2020), the opportunity cost of choosing more innovative strategies in terms of lower speed to break-even is greater than the opportunity cost of choosing strategies that generate higher levels of early revenue. This suggests that, while entrepreneurs may emphasize different objectives in different stages of NVC processes, innovation is the most expensive in terms of resources.[8] In themselves, these findings are not new, but our novel modelling approach allows us to quantify these trade-offs with more confidence that the estimates are not biased by unobserved heterogeneity.

Our second set of results relates to what extent new ventures succeed in attaining the studied combination of objectives. In our dataset we find that heterogeneity in this respect is very large. For the half-normal and exponential models for which we estimated the ratio of variance in distance to the frontier to variance in residual noise (λ), the former plays a tangible role as reported in Table 2.

Our third set of results confirms the basic intuition both in the ‘sweat equity’ approach (Bhandari and McGrattan, 2021) and in the Resource Based View of the firm (Peteraf, 1993) that state it is the firm specific inputs that are relevant constraints on building the sustainable competitive advantages that may sustain a new venture. We find that imitable resources available on the external markets, such as hired labor and debt finance, do not contribute significantly to achieving the objectives we have considered in our model. Instead, it is the labor inputs of the founding team and its equity investments that drive ventures at the frontier of new venture creation.

These results re-affirm that indeed the new venture creation process should be considered an important bottleneck in the process of innovation. If proprietary resources are what drives NVC, then new ideas cannot find their way to the market unless entrepreneurs are willing and able to organize and mobilize these resources around that idea. The trade-offs entrepreneurs face in bringing new products and services to the market and thereby to create Schumpeter’s “gale of

[8] The labor elasticity at the frontier ranges from 0.30 to 0.34, while the capital elasticity is 0.03 in all specifications of Table 2. Our results thus indicate that outputs respond more strongly to a proportional change in labor than to a proportional change in capital.

creative destruction" constrain then in turning knowledge into innovation and ultimately economic growth. And although entrepreneurship, like knowledge, human capital, and capital accumulation, can never be more than a proximate cause of growth, identifying proprietary resources in new venture creation as an important bottleneck also brings into focus a new set of institutions as fundamental causes. The institutions that motivate entrepreneurs to apply proprietary resources they need to build new ventures and challenge the status quo in markets, are the institutions that help turn knowledge creation into actual growth.

6. Conclusions

This paper proposes that allowing for unobserved heterogeneity in venture creation processes by estimating Stochastic Frontier models would be a major step forward in studying the inherently heterogeneous venture creation process. Understanding what multidimensional vector of inputs drives new ventures in achieving a multidimensional vector of objectives may go a long way in understanding entrepreneurship, innovation and growth. It is our contention that, once we understand how new ventures achieve their objectives, we can help them improve their performance by choosing a more appropriate mix of objectives, setting initial configurations, as well as by improving the environmental factors that prevent ventures from being best in class relative to their objectives. Our work therefore has important implications for future researchers, for practitioners and for policy makers. Our method is easily extended to contexts where different objectives, resources, characteristics and environmental variables are deemed relevant: for example, in social entrepreneurship one might consider multiple non-monetary objectives, whereas in corporate venturing one might zoom in on managerial talent and access to parent firm distribution networks and knowledge base as strategic inputs.

Our primary aim in this work has been to analyze the new venture creation process in terms of trade-offs among alternative objectives, dealing with resource heterogeneity and analyzing heterogeneity in entrepreneurial performance. A secondary aim has been to operationalize that approach empirically and to quantify the principal trade-offs and input-output relationships. Furthermore, we have stressed the distinction between proprietary and market resources that are subject to strategic decision making by entrepreneurs (i.e., labor and capital inputs that we further differentiated between the imitable and proprietary). We have chosen not to go in the initial study into the factors that drive the residual heterogeneity[9] but rather sought at this stage to quantify its importance. Future research might focus on providing a fuller and more nuanced account of the factors (for example at the level of the firm, region and industry) that might explain the trade-offs at and distance to the frontier among start-up processes.

Our results also have important policy implications. While additional external resources always allow entrepreneurs to achieve more of their objectives, such a policy will be less effective than the state supporting and incentivizing the provision of both founders' equity and labor. Also, policy makers will be interested in factors that enhance the long-term prospects of new ventures by making them more innovative, especially since this seems to trade-off especially sharply against short-term factors like the need for company income (revenue) and speed to break-even point. Our

[9] Primarily because of sample size. In small samples, the conditional heteroskedastic estimators lack precision for the parameters of variance.

framework helps to understand how these objectives are inter-related and how interventions may relax constraints and affect the way entrepreneurs may better handle these trade-offs.

References

Acs, Z. J., & Sanders, M. (2012). Patents, knowledge spillovers, and entrepreneurship. *Small Business Economics*, *39*(4), 801-817.

Acs, Z. J., & Sanders, M. W. (2013). Knowledge spillover entrepreneurship in an endogenous growth model. *Small Business Economics*, *41*(4), 775-795.

Acs, Z. J., Stam, E., Audretsch, D.B., and O'Connor, A. (2017) 'The lineages of the entrepreneurial ecosystem approach', *Small Business Economics*, 49(1), pp. 1–10.

Aghion, P., Akcigit, U., & Howitt, P. (2014). What do we learn from Schumpeterian growth theory?. In *Handbook of economic growth* (Vol. 2, pp. 515-563). Elsevier.

Dinopoulos, E., & Thompson, P. (1998). Schumpeterian growth without scale effects. *Journal of Economic Growth*, *3*(4), 313-335.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*, 323-351.

Alvarez, S. A., & Barney, J. B. (2005). How do entrepreneurs organize firms under conditions of uncertainty?. *Journal of management*, 31(5), 776-793.

Autio, E., Nambisan, S., Thomas, L.D. and Wright, M. (2018) 'Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems', *Strategic Entrepreneurship Journal*, 12(1), pp. 72–95.

Barney, J. (1991) 'Firm Resources and Sustained Competitive Advantage', *Journal of Management*, 17(1), pp. 99–120.

Baumol, W. J. (1990). Productive, unproductive and destructive. *Journal of Political Economy*, 98(5), 893-921.

Bhandari, A., & McGrattan, E. R. (2021). Sweat equity in US private business. *The Quarterly Journal of Economics*, *136*(2), 727-781.

Bolzani, D., Fini, R., Napolitano, S. and Toschi, L. (2019) 'Entrepreneurial Teams: An Input-Process-Outcome Framework', *Foundations and Trends in Entrepreneurship*, 15(2), pp. 56–258.

Carlsson, B., Braunerhjelm, P., McKelvey, M., Olofsson, C., Persson, L. and Ylinenpaa, H. (2013) 'The evolving domain of entrepreneurship research', *Small Business Economics*, 41(4), pp. 913–930.

Carter, N. M., Gartner, W. B. and Reynolds, P. D. (1996) 'Exploring start-up event sequences', *Journal of Business Venturing*, 11(3), pp. 151–166.

Casson, M., Yeung, B., Basu, A. and Wadeson, N.(eds) (2008) *The Oxford Handbook of Entrepreneurship*. Oxford: Oxford University Press.

Davidsson, P. (2016) *Researching Entrepreneurship: Conceptualization and Design*. 2nd edn. Cham: Springer International Publishing (International Studies in Entrepreneurship).

Davidsson, P. and Gruenhagen, J. H. (2020) 'Fulfilling the Process Promise: A Review and Agenda for New Venture Creation Process Research', *Entrepreneurship Theory and Practice*.

Davidsson, P. and Honig, B. (2003) 'The role of social and human capital among nascent entrepreneurs', *Journal of Business Venturing*, 18(3), pp. 301–331.

Delmar, F. and Shane, S. (2004) 'Legitimating first: organizing activities and the survival of new ventures', *Journal of Business Venturing*, 19(3), pp. 385–410.

Dilli, S., Elert, N. and Herrmann, A. (2018) ‘Varieties of entrepreneurship: exploring the institutional foundations of different entrepreneurship types through “Varieties-of-Capitalism” arguments’, *Small Business Economics*, 51(6), pp. 1–28.

Dushnitsky, G. and Matusik, S. F. (2019) ‘A fresh look at patterns and assumptions in the field of entrepreneurship: What can we learn?’, *Strategic Entrepreneurship Journal*, 13(4), pp. 437–447.

Elert, N., Henrekson, M. and Sanders, M. (2019) *The Entrepreneurial Society: A Reform Strategy for the European Union*. Berlin: Springer Publishers (International Studies in Entrepreneurship).

Estrin, S., David, L. and Dietrich, M. (2013) *Microeconomics*. 6th edition. London: Pearson.

Evans, D. S., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of political economy*, 97(4), 808-827.

Farrell, M. J. (1957) ‘The Measurement of Productive Efficiency’, *Journal of the Royal Statistical Society. Series A (General)*, 120(3), pp. 253–290.

Foo, M., Uy, M. and Baron, R. (2009) ‘How do feelings influence effort? An empirical study of entrepreneurs’ affect and venture effort.’, *The Journal of applied psychology*, 94, pp. 1086–94.

Garnsey, E. (1998) ‘A Theory of the Early Growth of the Firm’, *Industrial and Corporate Change*, 7(3), pp. 523–556.

Gartner, W. B. (1985) ‘A Conceptual Framework for Describing the Phenomenon of New Venture Creation’, *The Academy of Management Review*, 10(4), pp. 696–706.

Gartner, W. B. (1988) ‘“Who Is an Entrepreneur?” Is the Wrong Question’, *American Journal of Small Business*, 12(4), pp. 11–32.

Held, L., Herrmann, A. M. and van Mossel, A. (2018) ‘Team formation processes in new ventures’, *Small Business Economics*, 51(2), pp. 441–464.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. *Review of Economics and Statistics*, *95*(2), 347-361.

Hartog, J., Van Praag, M., & Van Der Sluis, J. (2010). If you are so smart, why aren't you an entrepreneur? Returns to cognitive and social ability: Entrepreneurs versus employees. *Journal of Economics & Management Strategy*, 19(4), 947-989.

Herrmann, A. M. (2019). A plea for varieties of entrepreneurship. *Small Business Economics*, *52*(2), 331-343.

Herrmann, A. M., Storz, C. and Held, L. (2020) ‘Whom do nascent ventures search for? Resource scarcity and linkage formation activities during new product development processes’, *Small Business Economics*.

Jones, C. I. (1995). R & D-based models of economic growth. Journal of political Economy, 103(4), 759-784.

Jones, C. I. (2005). Growth and ideas. In Handbook of economic growth (Vol. 1, pp. 1063-1111). Elsevier.

Koellinger, P. D., & Roy Thurik, A. (2012). Entrepreneurship and the business cycle. *Review of Economics and Statistics*, 94(4), 1143-1156.

Kumbhakar, S. C. and Lovell, C. A. K. (2003) *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S. C., Wang, H.-J. and Horncastle, A. P. (2015) *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge: Cambridge University Press.

Lazear, E. P. (2004). Balanced skills and entrepreneurship. *American Economic Review*, 94(2), 208-211.

Lévesque, M. and Stephan, U. (2020) 'It's Time We Talk About Time in Entrepreneurship', *Entrepreneurship Theory and Practice*, 44(2), pp. 163–184.

Levine, R., & Rubinstein, Y. (2017). Smart and illicit: who becomes an entrepreneur and do they earn more?. *The Quarterly Journal of Economics*, 132(2), 963-1018.

Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics,* 508-523.

McMullen, J. S., & Dimov, D. (2013). Time and the entrepreneurial journey: The problems and promise ofstudying entrepreneurship as a process. *Journal of Management Studies*, 50(8), 1481–1512.

Martinez, M. and Aldrich, H. (2011) 'Networking strategies for entrepreneurs: Balancing cohesion and diversity', *International Journal of Entrepreneurial Behaviour & Research*, 17(1), pp. 7–38.

McDonald, R. and Gao, C. (2019) 'Pivoting Isn't Enough? Managing Strategic Reorientation in New Ventures', *Organization Science*, 30(6), pp. 1289–1318.

Meza, D. D., & Southey, C. (1996). The borrower's curse: optimism, finance and entrepreneurship. *The Economic Journal*, 106(435), 375-386.

O'Connor, A., Stam, E., Sussan, F. and Audretsch, D.B. (eds) (2018) *Entrepreneurial Ecosystems: Place-Based Transformations and Transitions*. Berlin: Springer International Publishing (International Studies in Entrepreneurship).

Parker, S. C. (2018) *The Economics of Entrepreneurship*. 2nd edn. Cambridge: Cambridge University Press.

Peteraf, M. A. (1993) 'The Cornerstones of Competitive Advantage: A Resource-Based View', *Strategic Management Journal*, 14(3), pp. 179–191.

Reynolds, P. D. (2018) *Business Creation: Ten Factors for Entrepreneurial Success*. Cheltenham: Edward Elgar Publishing.

Reynolds, P. and Miller, B. (1992) 'New firm gestation: Conception, birth, and implications for research', *Journal of Business Venturing*, 7(5), pp. 405–417.

Romer, P. M. (1990). Endogenous technological change. Journal of political Economy, 98(5, Part 2), S71-S102.

Romer, P. M. (1994). The origins of endogenous growth. Journal of Economic perspectives, 8(1), 3-22.

Rumelt, R. P. (1984) 'Towards a strategic theory of the firm', in Lamb, R. (ed.) *Competitive Strategic Management*. Englewood Cliffs, NJ: Prentice-Hall, pp. 556–570.

Sanders, M., Marx, A. and Stenkula, M. (eds) (2020) *The Entrepreneurial Society: A Reform Strategy for Italy, Germany and the UK*. Berlin: Springer Publishers (International Studies in Entrepreneurship).

Santos, S. C. and Cardon, M. S. (2019) ‘What’s love got to do with it? Team entrepreneurial passion and performance in new venture teams.’, *Entrepreneurship Theory and Practice*, 43(3), pp. 475–504.

Sarasvathy, S. D. (2001) ‘Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency’, *The Academy of Management Review*, 26(2), pp. 243–263.

Schumpeter, J. A. (1934) *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.

Segerstrom, P. S., Anant, T. C., & Dinopoulos, E. (1990). A Schumpeterian model of the product life cycle. *The American Economic Review*, 1077-1091.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217–226.

Shepherd, D. A., Wennberg, K., Suddaby, R., & Wiklund, J. (2019). What are we explaining? A review and agenda on initiating, engaging, performing, and contextualizing entrepreneurship. *Journal of Management*, 45(1), 159–196.

Welter, F. (2011) ‘Contextualizing Entrepreneurship—Conceptual Challenges and Ways Forward’, *Entrepreneurship Theory and Practice*, 35(1), pp. 165–184.

Wernerfelt, B. (1984) ‘A Resource-Based View of the Firm’, *Strategic Management Journal*, 5(2), pp. 171–180.

Wiklund, J. Davidsson, P., Audretsch, D.B. and Karlsson, C. (2011) ‘The future of entrepreneurship research.’, *Entrepreneurship Theory and Practice*, 35(1), pp. 1–9.

Table 1. Descriptive Statistics, ICT sector

Variable	Obs	Mean	Std. Dev.	Min	Max
Frontier					
SPD: Speed to break-even point	499	0.203	0.183	0.011	0.982
INN: Innovativeness	563	31.972	17.255	9.091	81.818
RVN: Revenue	363	15382	32259	0	400000
SPD: Speed to break-even point (log)	499	-2.126	1.139	-4.532	-0.018
INN: Innovativeness (log)	563	3.293	0.622	2.207	4.405
RVN: Revenue (log)	362	8.603	1.577	2.303	12.899
INN - SPD (log difference)	499	5.396	1.399	2.869	8.572
RVN - SPD (log difference)	334	10.719	1.817	5.527	15.969
Inputs					
Financial capital (log, scaled)	361	2.975	3.400	0.000	23.516
Equity capital (log, scaled)	366	1.089	1.013	0.000	4.508
Loans, grants (log, scaled)	361	2.869	3.272	0.000	14.926
Labour (log, scaled)	361	0.347	1.641	0.000	23.491
Founders labour (log, scaled)	363	0.885	0.884	-0.002	4.456
Employees and services (log, scaled)	361	0.640	0.709	0.000	3.722
Explanatory variables for technical inefficiency					
Spin-off	563	0.146	0.353	0.000	1.000
1 founder	403	0.330	0.471	0.000	1.000
2 founders	403	0.392	0.489	0.000	1.000
3 founders	403	0.171	0.377	0.000	1.000
4 and more founders	403	0.107	0.309	0.000	1.000
More than one sector	563	0.089	0.285	0.000	1.000
Germany	563	0.309	0.463	0.000	1.000
Italy	563	0.162	0.368	0.000	1.000
Netherlands	563	0.028	0.166	0.000	1.000
UK	563	0.231	0.422	0.000	1.000
US	563	0.270	0.444	0.000	1.000

Table 2. Estimates of the Productivity Frontier (Speed to Market). Benchmark models (ICT sector)

Dependent: Speed to Break-Even Point (log)					
	(1) OLS	(2) MAD	(3) Half-normal	(4) Truncated	(5) Exponential
Innovativeness - Speed (log difference)	-0.500***	-0.485***	-0.502***	-0.514***	-0.500***
	(0.0231)	(0.0283)	(0.0238)	(0.0255)	(0.0231)
Revenue - Speed (log difference)	-0.0407**	-0.0342*	-0.0450**	-0.0505**	-0.0425**
	(0.0142)	(0.0173)	(0.0146)	(0.0169)	(0.0143)
Capital (log scaled)	0.0340***	0.0309*	0.0306**	0.0276**	0.0325**
	(0.0102)	(0.0124)	(0.0102)	(0.00959)	(0.00996)
Labour (log scaled)	0.304***	0.342***	0.303***	0.297***	0.306***
	(0.0381)	(0.0466)	(0.0385)	(0.0372)	(0.0378)
Constant	2.409***	2.385***	2.769***	3.228***	2.599***
	(0.0589)	(0.0719)	(0.0892)	(0.0796)	(0.0799)
logarithm of total variance			-2.562***		-2.185***
			(0.289)		(0.179)
μ				0.769***	
				(0.0754)	
Λ (variance in inefficiency / variance in noise)			1.600		0.569
R squared	0.884				
Observations	331	331	331	331	331

Standard errors in parentheses

+ p<0.10; * p<0.05; ** p<0.01; *** p<0.001

Table 3. Estimates of the Productivity Frontier (Speed to Market). Models with Heterogenous Labour and Capital

Dependent: Speed to Break-Even Point (log)	(1) OLS	(2) MAD	(3) Half-normal	(4) Truncated	(5) Exponential
Innovativeness - Speed (log difference)	-0.518***	-0.489***	-0.523***	-0.536***	-0.518***
	(0.0236)	(0.0324)	(0.0244)	(0.0251)	(0.0234)
Revenue - Speed (log difference)	-0.0379*	-0.0515*	-0.0433**	-0.0423**	-0.0401**
	(0.0150)	(0.0206)	(0.0156)	(0.0142)	(0.0150)
Equity (log scaled)	0.0499***	0.0556***	0.0467***	0.0450***	0.0486***
	(0.0106)	(0.0145)	(0.0102)	(0.00976)	(0.0102)
Loans and grants (log scaled)	0.00479	-0.00439	0.00498	0.00855	0.00410
	(0.0130)	(0.0179)	(0.0131)	(0.0113)	(0.0127)
Founders' labour (log scaled)	0.308***	0.280***	0.294***	0.295***	0.305***
	(0.0440)	(0.0605)	(0.0431)	(0.0405)	(0.0428)
Non-Founders' labour (log scaled)	-0.0350	0.0402	-0.0274	-0.0481	-0.0312
	(0.0457)	(0.0629)	(0.0441)	(0.0413)	(0.0448)
Constant	2.469***	2.398***	2.843***	3.228***	2.650***
	(0.0598)	(0.0822)	(0.100)	(0.0899)	(0.0848)
Founders v. Non-founders: *chi2*	20.67***	5.34*	19.70***	26.38***	20.91***
Equity v. Loans & grants: *chi2*	6.71**	6.27*	6.06*	7.02***	6.98**
logarithm of total variance			-2.545***		-2.115***
			(0.334)		(0.177)
μ				0.667***	
				(0.098)	
Λ (variance in inefficiency / variance in noise)			1.626		0.521
Observations	329	329	329	329	329

Standard errors in parentheses; + p<0.10; * p<0.05; ** p<0.01; *** p<0.001

Figure 1. Unobserved heterogeneity in NVC treated as noise

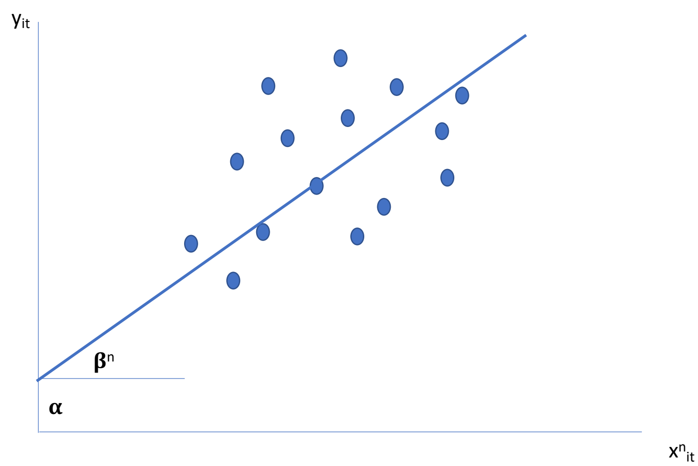
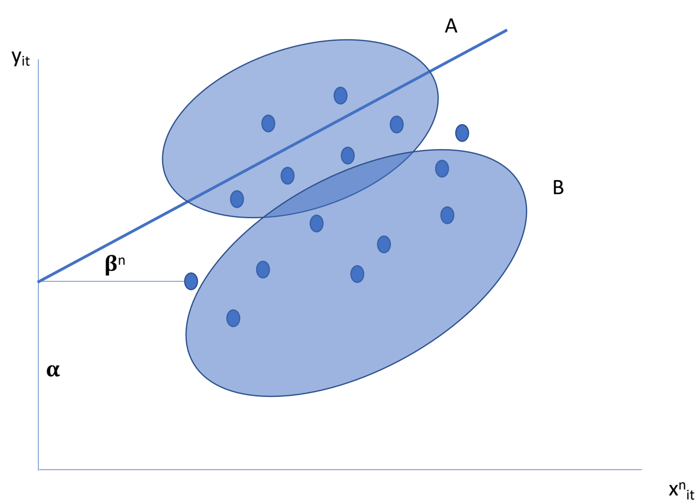


Figure 2. Explicit modeling of unobserved heterogeneity in NVC



Appendix A

Suppressing the *it* subscripts to save on notation, we follow Bos, Lamers, Li, Sanders and Schippers (2020) and define the distance to the frontier as:

$$D(\mathbf{y^M}, \mathbf{x^N}) = \operatorname*{argmin}_{\theta} \left(\frac{\mathbf{y^M}}{\theta} \in f^M(\mathbf{x^N}) \right) \tag{A1}$$

where $\mathbf{y^M} = (y_1, \ldots, y_M)$ is the vector of *M* maximum attainable outputs using a vector of *N* inputs or resources, $\mathbf{x^N}=(x_1, \ldots, x_N)$. To separate the inputs from the outputs, we assume that the distance function is homogeneous of degree one in $\mathbf{y^M}$ so that we can divide by $\mathbf{y^M}$ on both sides. Dividing by the first output y_1, we can then write:

$$\frac{D(\mathbf{y^M}, \mathbf{x^N})}{y_1} = h\left(\mathbf{x^N}, \frac{y_2}{y_1}, \frac{y_3}{y_1}, \ldots, \frac{y_M}{y_1}\right) \tag{A2}$$

where $h(\cdot)$ is a parametric regression function. Taking the logarithms on both sides of the equation and assigning a normally distributed disturbance term to the right-hand side, we have:

$$\ln D(\mathbf{y^M}, \mathbf{x^N}) - \ln y_1 = \ln h\left(\mathbf{x^N}, \frac{y_2}{y_1}, \frac{y_3}{y_1}, \ldots, \frac{y_M}{y_1}\right) - \varepsilon_i \tag{A3}$$

Finally, denoting $\ln D(\mathbf{y^M}, \mathbf{x^N}) \equiv -v_i \leq 0$ and assuming the standard Cobb-Douglas (loglinear) form for $\ln h\left(\mathbf{x^N}, \frac{y_2}{y_1}, \frac{y_3}{y_1}, \ldots, \frac{y_M}{y_1}\right)$, our empirical model becomes:

$$\ln y_i^1 = \alpha + \sum_{n=1}^{N} \beta_n \ln x_i^n + \sum_{m=2}^{M} \gamma_m \ln \frac{y_{it}^m}{y_{it}^1} + \varepsilon_i - v_i \tag{3}$$

References to Appendix A

Bos, J., Lamers, M., Li, R., Sanders, M. and Schippers, V. (2020). 'Towards Good Governance: How Do Institutions Affect the Efficiency of Disaster Relief?' Maastricht University Mimeo.

Appendix B: Distributions

Figure B1. Distribution of the Speed to Break-Even Point Variable (ICT sector)

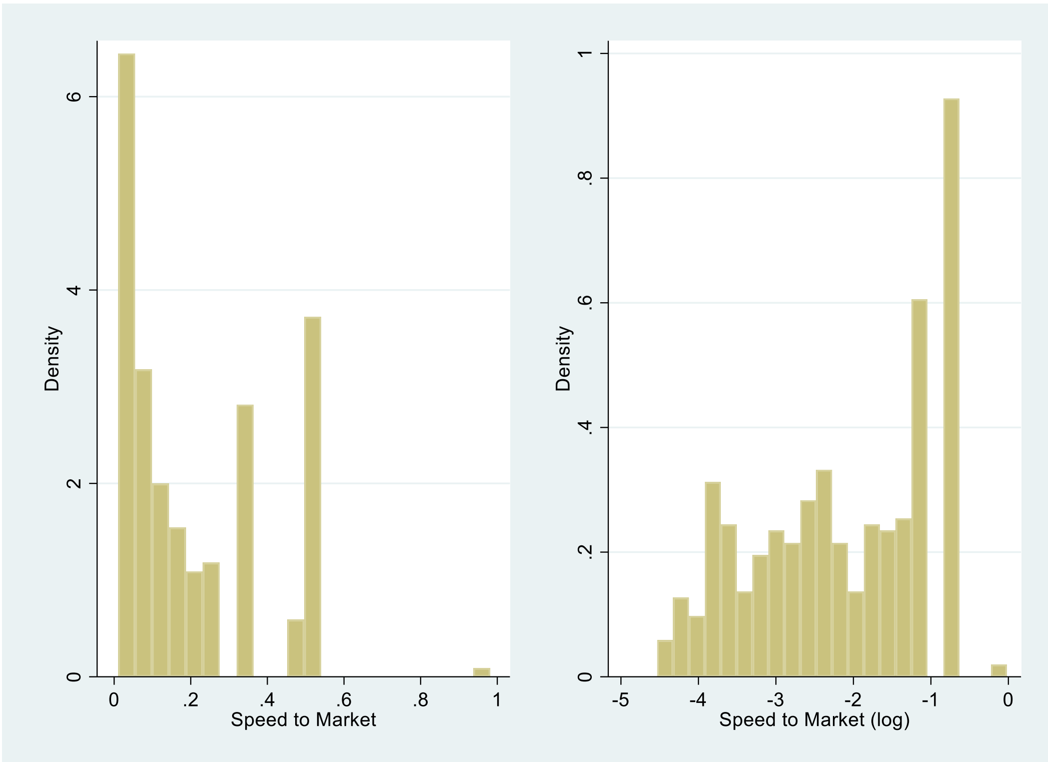


Figure B2. Distribution of the Revenue Variable (ICT sector)

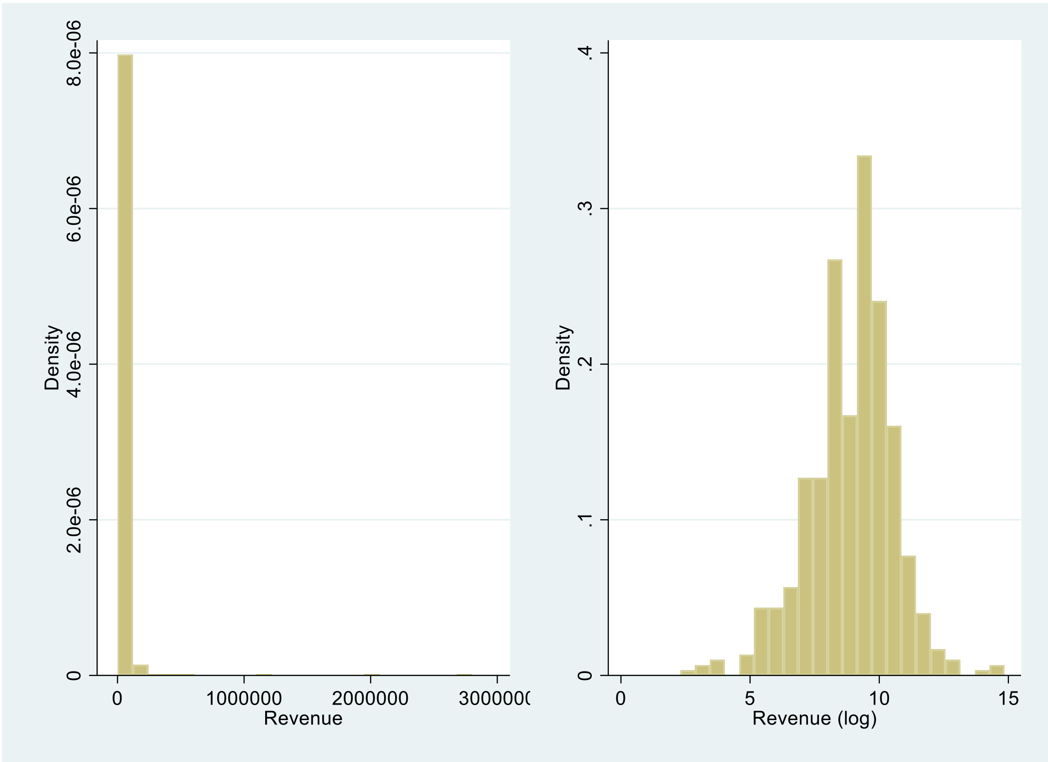


Figure B3. Distribution of the Innovation Variable (ICT sector)

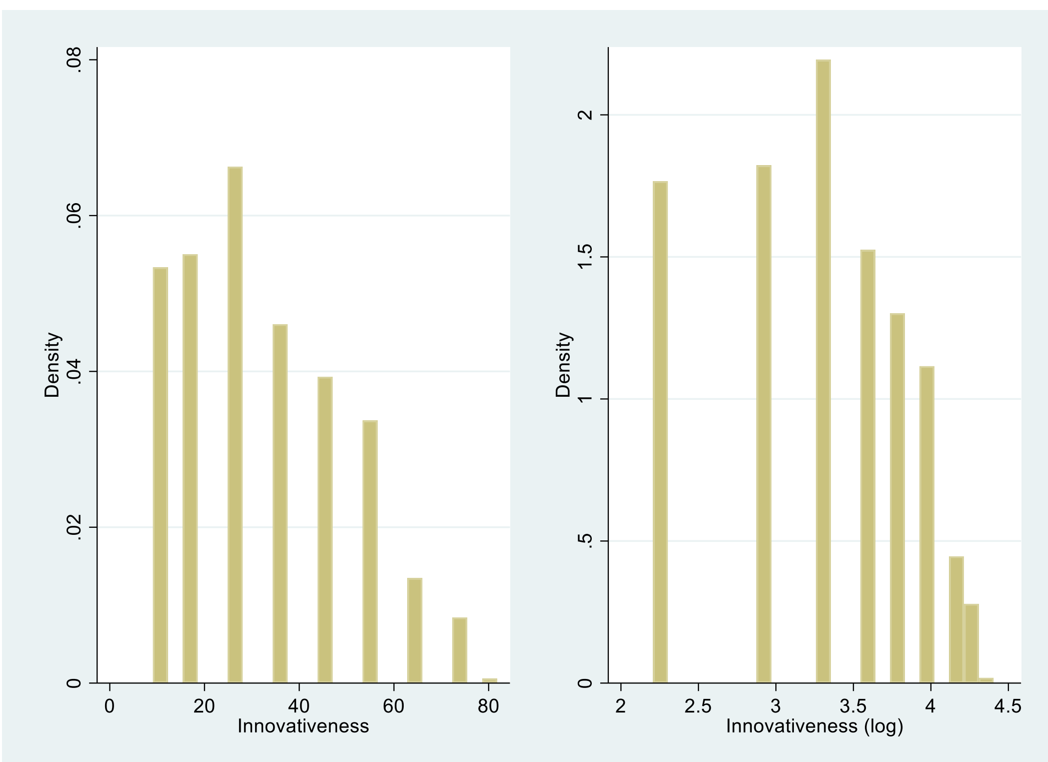


Figure B4. Distribution of technical efficiency: COLS and CMAD (ICT sector)

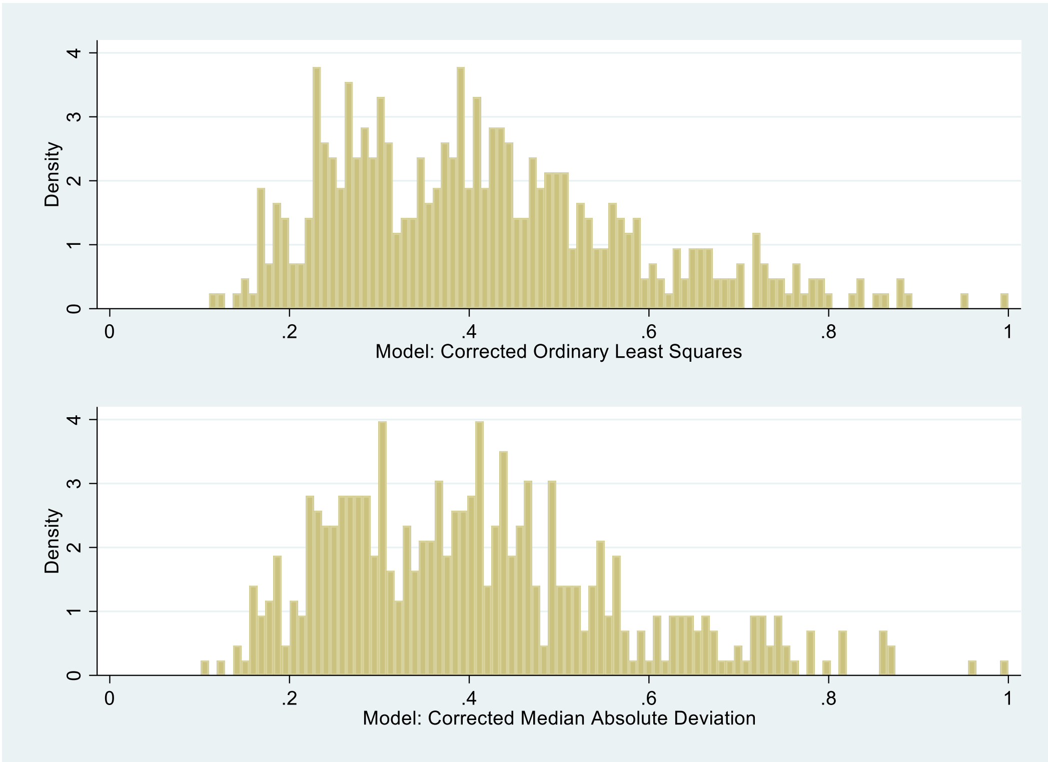
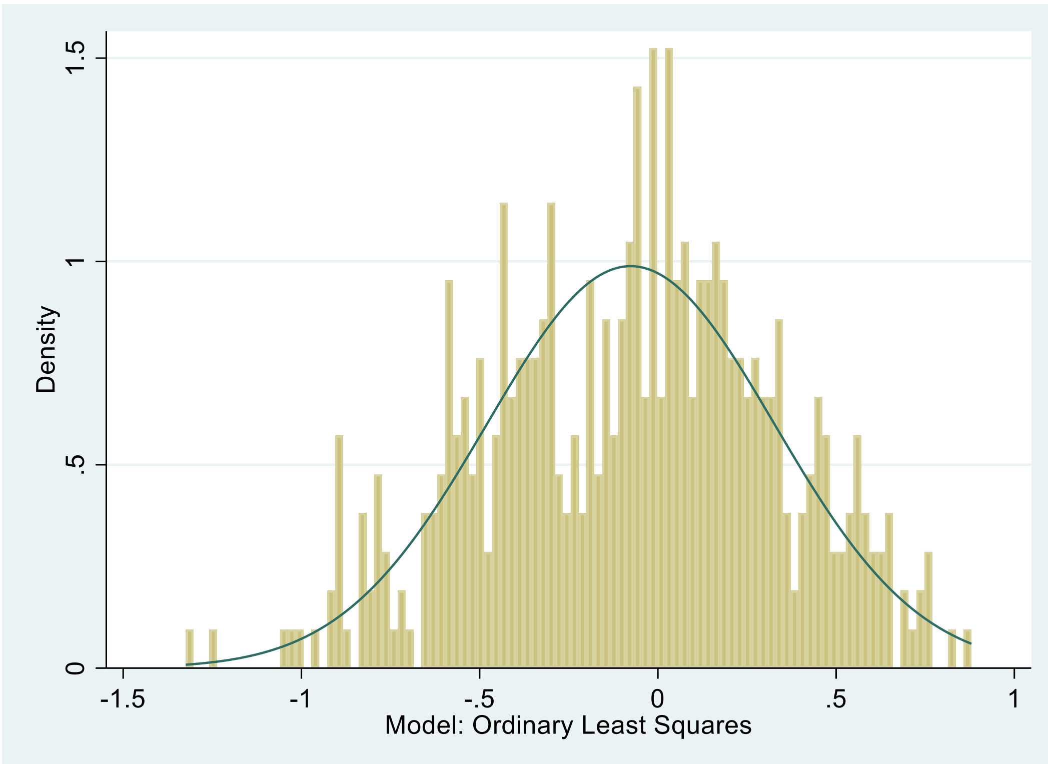


Figure B5. Error distribution from the OLS model (ICT sector)



The distributional assumptions that need to hold for the application of the frontier model can also be tested formally. As Figure B5 illustrates, we can overlay a truncated normal distribution on the actual distribution of errors from the OLS estimates. For a frontier function, we expect a left-skewed error term, and indeed this is what we get. We can confirm that by applying skewness and kurtosis tests for normality (D'Agostino, Belanger, and D' Agostino, 1990; Royston, 1992), for our data, the tests come significant below the p=0.05 level, implying a frontier specification is justified.

References to Appendix B

Coelli, T. (1995) 'Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis', *Journal of Productivity Analysis*, *6*(3), pp. 247-268.

D'Agostino, R. B., Belanger, A. and D'Agostino Jr, R. B. (1990). 'A Suggestion for Using Powerful and Informative Tests of Normality', *The American Statistician*, 44(4), pp. 316–321.

Royston, P. (1992) .'Tests for Departure from Normality', *Stata Technical Bulletin*, 1.

Appendix C: Specifying the Categories of Labor and Capital Inputs

Labor

First consider the founders. For each of the first five, we know the date they started and ended or changed their commitment in the venture as well as whether they were engaged full time or part time[10]. We compute total founder labor input by assuming part time involvement as 50%.

$$X_i^{lfft} = \sum_{t=0}^{T} X_{it}^{lfft}$$

where X_{it}^{lfft} is 1 if founder *i* was engaged full time in month *t* between start date 0 and end date *T*. Summing this over the entire firm formation period gives the person-months engaged in the firm by founder *i*. We have information only for the first five members of the founding team, but as there are very few firms with larger founding teams in our sample, we decided to ignore labour input from founders 6 and beyond. Summing over the first five founders we get for total full-time person months provided by the founder team:

$$X^{lfft} = \sum_{i=1}^{5} X_i^{lfft}$$

We can compute the number of part-time person months provided by the first five founders and compute total founding team labor input as:

$$X^{lfft} + 0.5X^{lfpt} = X^{lf}$$

This gives the number of full-time person months of labor input provided by the first five founders on the assumption that a part time engagement is 50%.

Similarly, we have information on the first five members of employed staff. Here, however, there are quite a few firms that employed more than five employees by the end of their firm formation process, and we do not want to introduce strong bias in our data by ignoring that. As we know the total number of employees in the firm, we decided to add this number minus 5 multiplied by the average labor input of the first five employees. We compute:

$$X_i^{left} = \sum_{t=0}^{T} X_{it}^{left}$$

and add over the first five employees to obtain:

$$X^{left} = \sum_{i=1}^{5} X_i^{left}$$

and weighting all part time employees by 0.5 we obtain:

$$X^{left} + 0.5X^{lept} = X^{le}$$

For service providers that were used, we do not have information on the intensity of their contract. We only know if and how long they have been engaged by the founding team. We therefore can only include the number of months during which a service provider was engaged. Moreover, as before, we only have information on the first five of these service providers. We compute the total number of months of external service engagement as:

[10] Unfortunately, we do not have information on the exact number of hours.

$$X_i^{ls} = \sum_{t=0}^{T} X_{it}^{ls}$$

and add over the first five service providers to obtain:

$$X^{ls} = \sum_{i=1}^{5} X_i^{ls}$$

where we again added the total number of service providers listed minus five times the average engagement for the first five providers to obtain our proxy for externally sourced labor inputs.

Total labor input in the firm formation process can then be computed as the sum of the three labor inputs above. Finally, all labor input variables are divided by the number of months between the start and end date of the venture creation process to create an average labor intensity value that proxies for the average number of person months of labor engaged in the firm during its formation period. If we did not do that, we would introduce a spurious negative correlation between the speed to break-even point and amount of labor, as when the start-up process takes longer, more month-person units are used.

Capital

For financial resources, we follow a similar procedure to compute total equity, formal and informal debt, and grants that were employed in the firm formation process. For equity, we obtained the start date and amount from the survey for the first five equity providers. The duration for these financial engagements is to the end date of the firm formation process, as equity does not leave the firm. For debt, we have the amount and the start and end date for the loan for the first five loan providers. If the end date is after the end date of the firm formation process, we only counted the months till the formation end date, *T*.

$$X_i^{kef} = \sum_{t=0}^{T} X_{it}^{kef}$$

with *X* now denoting the amount of equity invested in month *t* by investor *i*, summing over the first five investors we obtain:

$$X^{kef} = \sum_{i=1}^{5} X_i^{kef}$$

such that X^{kef} is the total amount of Euros [11] invested in the firm times the months these Euros were invested. We know the number of equity investors in the firm and added that number minus 5 times the average amount of equity invested when the number of equity providers was above 5.

Similarly, but now with the complication that the loans may expire between registration date and end date, we obtained for formal debt:

$$X_i^{kfl} = \sum_{t=0}^{\min(\tau_i, T)} X_{it}^{kfl}$$

[11] We used an exchange rate of 01-04-2019 at 1.12 dollar to the euro and 0.85 pound to the euro.

where X^{kfl} is the amount of formal debt provided in month t by lender i, and τ_i is the expiration date of the loan provided it falls before T. Summing over the first five lenders we obtain:

$$X^{kfl} = \sum_{i=1}^{5} X_i^{kfl}$$

A similar procedure was followed for informal debt. For grants, we assume, as with equity, that the financial resources stay in the firm from the date the grant is granted to the end of the firm formation process. Again, if more than five grants were collected, we added the average grant times the number of grants above 5. This, however, is extremely rare in our dataset. As before, these values were then scaled by the number of months to obtain the average amount of euros of equity, debt and grants engaged in the firm during the firm formation process. Summing all the sources of finance gives the total capital input for our benchmark equations.