Modelling the Determinants of Job Creation: Microeconometric Models Accounting for Latent Entrepreneurial Ability

Abdelfatah Ichou

Zoetermeer, July 2010

SCALES
SCientific AnaLysis of Entrepreneurship and SMEs

Dit onderzoek is mede gefinancierd door het programmaonderzoek MKB en Ondernemerschap (www.ondernemerschap.nl)

EIM Research Reports

reference number	H201018
publication	July 2010
emailaddress corresponding author	a.ichou@stratusbv.nl
address	EIM Bredewater 26 P.O. BOX 7001 2701 AA Zoetermeer The Netherlands Phone: +31 79 343 02 00 Fax: +31 79 343 02 03 Internet: www.eim.nl

Modelling the Determinants of Job Creation: Microeconometric Models Accounting for Latent Entrepreneurial Ability

Abdelfatah Ichou

EIM Business and Policy Research Zoetermeer, The Netherlands

July 4, 2010

Abstract

During the last decades, most developed countries have shown a remarkable increase in entrepreneurship rates. Recent research suggests that this increase is, for a considerable part, caused by an increase in the share of solo self-employed. Nowadays, for example, more than half of all Dutch business owners are solo self-employed. This raises the question which factors determine whether an entrepreneur becomes an employer or remains solo self-employed. This paper is devoted to answering this question by means of an empirical analysis using data of Dutch start-ups founded between 1998 and 2000. Using various duration and count data models we are able to identify several factors that influence job creation by entrepreneurs.

Keywords: entrepreneurship, self-employment, job-creation, start-ups, duration analysis, count models

Acknowledgement:
Jan de Kok and Richard Paap provided helpful comments on an earlier draft. This paper has been written in the framework of the research program SCALES, carried out by EIM and financed by the Dutch Ministry of Economic Affairs.

Contents

Chapter 1

Introduction

1.1 Motivation

After the Second Industrial Revolution the importance of growing firms for the economic prosperity became obvious. In *The Theory of Economic Development* (Schumpeter, 1934) entrepreneurship is the leading element and entrepreneurs the ones who make the economy grow. The entrepreneurs described by Schumpeter were innovative individuals that created new products, opened new markets and introduced new production methods and technologies. However, less innovative entrepreneurs also can play an important role. By exploiting business opportunities, they make a contribution to the economy as they propagate new business techniques (Baumol, 1998).

Nowadays, entrepreneurship is a key policy issue and much attention is being paid to the economic importance of entrepreneurship within and across countries. One reason that we find entrepreneurship, and in particular, entrepreneurs important, lies in the fact that they have the ability to create jobs for others. Therefore, employment creation by entrepreneurial individuals is one way to characterize their economic importance. Once an individual successfuly has made the transition to self-employed, the path towards job creation can be tread. However, we observe that the majority of the entrepreneurial population does not enter into the challenge of job creation. For example, in 1997 63% of the Dutch business owners were solo self-employed individuals (OECD, 2000). The question that now arises is: what determines whether a self-employed individual becomes a job creator? Stated alternatively, which factors affect an entrepreneur's decision to hire employees? And, how many employees are employed and once the entrepreneur has decided to hire employees? This paper is devoted to answering these questions by means of an empirical analysis using data of Dutch start-ups founded between 1998 and 2000. In this paper we essentially pay attention to two decisions that have to made by an entrepreneur: (1) the decision to employ personnel, and hence, to switch over from solo self-employed to employer, and, conditional upon the first decision, (2) the employee decision, i.e. the decision to hire a certain number of employees. In this paper we will refer to the first decision as the employer decision, and to the second as the employee decision. By studying the determinants of job creation at the micro level we can provide policymakers with new insights that can contribute to the development of schemes aiming not only to increase the share of business owners, but that will also stimulate job creation and employment growth.

1.2 Theoretical background

In this section we give an overview of the literature on the determinants of job creation. This includes a brief description of a recent EIM study on this subject that is based on the same data set that we use. At the end of this section we explain how this paper contributes to the literature.

1.2.1 The determinants of job creation

Compared to the self-employment decision, relatively little is known about the entrepreneur's hiring decision (Burke et al., 2002). One of the first studies that focused on job creation by new firms is that of Birch (1978). This research shows that new and smaller growing firms account for 81.5% of the new jobs in the United States. Since this pioneer work of Birch, the attention that is paid to this subject significantly increased (Burke et al., 2002). However, the literature on the determinants of job creation by self-employed individuals is still quite limited. We will give an overview of the available literature on the determinants of job creation. Barkham (1994) studies the relationship between characteristics of the entrepreneur and the size of his firm. He finds that entrepreneurs that are highly motivated and possess the necessary human capital (e.g. managerial skills) and right market information are the best job creators. Westhead and Cowling (1995) also find a positive link between human capital (in terms of educational level of the founder) and employment growth. Furthermore, they find that entrepreneurs that have better access to financial resources at the start-up grow faster. Caroll et al. (1995) pay attention to the prevalence of becoming job creator. Their research suggests that this likelihood is affected by the personal tax income situation of the entrepreneurs: when the tax rate of a solo self-employed goes up, the probability that this individual will hire employees goes down. Furthermore, tax rates are found to subdue firm growth in case the entrepreneur has decided to hire labour. Van Praag and Cramer (2001) also analyze how several characteristics of the entrepreneur affect firm size. Their most important empirical finding is that risk attitude of the entrepreneur affects the number of employees hired in a positive way. To their knowledge *'this empirical result is new and confirms almost all recent and older theories developed'*. Finally, Henley (2005) specifies an order probit model for the number of employees hired. By doing so he combines both the decision to become employer as well as the decision regarding the actual number of employees to hire. To the best of our knowledge, this author is only one that investigates both decisions, however, within a single framework. The previously mentioned authors either investigate the decision to make the transition from solo self-employed to employer or the number of employees hired in case of an employer, but not both. The empirical results of Henley suggest that the best job creators are middle-aged males. Furthermore, his results also suggest the existence of a positive relationship between the amount of social capital the entrepreneur possesses and the likelihood of becoming a job creator.

1.2.2 Latent entrepreneurial ability

From the available literature on the determinants of job creation we can identify several factors that can help explain the employer and employee decision that has to be made by a self-employed individual. These factors include motivation, human capital (e.g. education and working experience), social capital (e.g. business contacts), risk attitude, gender and age of the entrepreneur. However, as pointed out by Bosma et al. (2004) some characteristics that influence the employer and employee decision remain unobserved. The unobserved part of these characteristics, which are only known by the entrepreneur, are measures of entrepreneurial talent or intelligence. To our knowledge, the only author that explicitly takes into account this unobserved characteristics is Henley (2005). Henley refers to this unobserved heterogeneity as the entrepreneur's latent entrepreneurial ability. From a statistical point of view it is important to account for this unobserved heterogeneity as this could lead to a misspecified model rendering biased parameter estimates. Bosma et al. (2004) refer to this bias in the parameters as the *'unobserved talent bias'*. Therefore, we will explicitly focus on the latent entrepreneurial ability. We will do this by formulating extensions of our models that incorporate unobserved heterogeneity. By doing so, we are able to detect misspecification.

1.2.3 Entrepreneurial age as determinant of job creation

Many studies into determinants of job creation include entrepreneurial age as a control variable. The expected changes in the age decomposition of the workforce justify a more thorough investigation of the role of age, that goes beyond treating age as a control variable. A recent study by EIM (De Kok et al., 2010) has examined the nature of the relationship between age and entrepreneurship, taking into account direct as well as indirect age effects. Similar to our study, De Kok et al. (2010) distinguish between the employer decision and the employee decision. A first conclusion of their study is that it is important to make the distinction between these two decisions: the employer decision depends on other factors than the employee decision. A second conclusion is that age has a negative relationship with the outcome of both decisions, but that these relationships are indirect: once potential mediating variables are included in the estimations, the direct age effects are no longer significant. They find that entrepreneurs who start at older age are less likely to work fulltime in their new venture, are less willing to take risks and have a lower perception of their entrepreneurial skills. Each of these factors has, in turn, a positive impact on the probability of employing personnel. For the number of employees a negative indirect effect of age is found through the effect of age on the perception of entrepreneurial skills.

This study has much in common with our own study: they examine the same separate decisions and use the same data set as we do. There are, however, also important differences. A first difference lies in the attention for age as determinant of job creation. In De Kok et al. (2010) this is the main determinant that is investigated, and direct as well as indirect relationships between age and entrepreneurship are examined. In the current study, one could say that most attention is paid to determinants that have not been observed: an important aspect of this this study is how to control for latent entrepreneurial ability. A second difference lies in the methodology that is applied. De Kok et al. (2010) use a probit model to examine the decision to switch from solo self-employed to employer. The dependent variable in this model is whether or not the entrepreneur employs any employees three years after his start-up. Consequently, this model is estimated on a sample of entrepreneurs for which observations for the first three years after start-up are available. Although these results are valid for the population of entrepreneurs that survived the first three years after start-up, it is not clear whether the results can be generalised to the whole population of starters. In our study, the dependent variable is the time spent as solo entrepreneur before the transition to employer is made. We estimate duration models to determine how this duration is affected by various (observed and unobserved) determinants (more on this in section 1.3). This approach poses less limitations on the available observations, and hence the duration model can be based on considerably more observations than the analyses presented by De Kok et al. (2010).

1.2.4 Contribution to the literature

Our contribution to the literature is twofold. Firstly, this paper contributes to the literature by considering both the employer and employee decision separately. Within the context of the employer decision we compare solo self-employed entrepreneurs with job-creating entrepreneurs, and for the employee decision we compare job-creating entrepreneurs with a lower number of employees with those that have a higher number of employees. As mentioned before, only Henley (2005) takes into account both decisions, but does this within a single framework. To our knowledge, modelling both decisions separately has not been done before[1].

Secondly, we will formulate models that account for the latent entrepreneurial ability. By comparing and extensively discussing models with and without the latent entrepreneurial ability, we expand the list of scarce readings on this topic[2].

1.3 General methodological set-up

As already mentioned before, we will focus on the employer and employee decision. In this section we give a brief overview of how we have conducted this study on the determinants of job creation. After we give a short description of the data set, we describe the models for the employer and employee decision respectively.

The data set

The data set we used for the this research contains data of a panel of Dutch entrepreneurs that founded their firm between 1998 and 2000. From that moment they were monitored annually by means of a questionnaire. This resulted in an unbalanced panel. The data set contains several

[1] Except for the EIM study by De Kok et al. (2010).

[2] The EIM study by De Kok et al. (2010) did not account for latent entrepreneurial ability.

characteristics of the entrepreneurs, including measures for human and social capital, the age of the entrepreneur at start-up and gender. Next to this, the data set also contains information about the firm, such as industry, innovativeness and firm size (measured as the number of employees). From the literature we know that this information can be useful for modelling the employer and employee decision. To control for a possible business cycle effect (at macro level), we enriched this data set by adding a measure of the business cycle to it. Hence, we added the annual GDP growth rate to the data base.

The employer decision

The employer decision can be regarded as a binary choice and will be modelled using a discrete time transition model. Thus, to study the employer decision we will perform a duration analysis. By doing so, we fully exploit the panel structure of the data set. Furthermore, duration models give us the ability to test whether the age of the firm influences the employer decision. In the duration analysis we will include all the entrepreneurs in the data set, job creators and non job creators. To account for the latent entrepreneurial ability we will specify extensions of the transition model that account for unobserved heterogeneity.

The employee decision

For the employee decision we will only consider job creators and model the number of employees they hire in the year they made the transition from solo self-employed to job creator. Hence, the analysis for the employee decision will be cross-sectional. To model the number of employees hired, we will use count models. Again, we will formulate extensions of the model that account for unobserved heterogeneity. The models for the employee decision will include the same set of explanatory variables as the models we specify in the duration analysis.

1.4 Outline

The remainder of this paper is organised as follows. In the next chapter we will introduce the data set. We will give a detailed overview of the available measures and present some descriptives. In Chapter 3 we discuss the employer decision. In this chapter we will introduce the discrete time transition modelling framework and use this framework to model the employer decision. Chapter 4 considers the employee decision. In this chapter we will formulate the count data models for the number of employees hired by job creators and discuss their outcomes. In the final chapter, Chapter 5, we present a summary and the results of this research.

Chapter 2

Data and Sample Description

2.1 Data set

To model the decision of entrepreneurs regarding job creation we use data from EIM Business & Policy Research. The data set contains information of recently started entrepreneurs and is obtained from the so-called Start-Up Panel. This panel consists of three cohorts of Dutch entrepreneurs who started a business in 1998, 1999 or 2000. In each of these years about 500 new entrepreneurs entered the panel. From that moment there were monitored annually by means of a written questionnaire. The data in the panel covers various topics including personal characteristics of the entrepreneur (gender, age, education and (entrepreneurial) experience), firm characteristics (firm size and sector), and objectives and strategy (growth goals and R&D activities). For the analysis that will be done in the next chapters, a measure of the business cycle is needed. Therefore, we included the Dutch annual GDP growth obtained from the IMF World Economic Outlook Database, 2009 in our data base.

The annual results have been merged into a single data set containing the annual observations of 1,402 entrepreneurs. For each entrepreneur the maximum number of years for which we have data available is 9. That is, each entrepreneur is interviewed up to 9 times. Since not all entrepreneurs cooperated in each year, we end up with an unbalanced panel data set containing in total 6,239 cases. To denote an entrepreneur we use the cross-sectional index $i = 1, \dots, 1,402$. To denote a year we use the time index $t = 1, \dots, 9$. Since t is 1 in the year in which the entrepreneur entered the panel, it can be interpreted as the firm's age.

A description of all the variables in the data set is presented in Table 2.1. Note that only the number of employees in the business and the GDP growth vary over time. All other variables are only measured in the first year. For the analysis this will be done in the next chapters that is, however, sufficient.

Table 2.1: Description of variables used in the analysis

Variable	**Description**	**Time-varying**
Number of employees	Number of workers employed in the business (excluding family members)	yes
Age	Age was measured as an ordinal scale with 10 age categories of 5 years each. For each category we take the median age, except for the first and last category. The resulting scale has a minimum of 18 and maximum of 65 years and is treated as a continuous variable	no
Controls		
Male	Dummy indicating that then entrepreneur is male	no
Industry	We include 8 industry dummies: Manufacturing; Construction; Wholesale; Retail; Hotels and restaurants; Sale and repair of motor vehicles; Transport; Business and financial services; Other services	no
Start-up motives	*Dummies indicating the most important start-up motive for the entrepreneur*	
Intrinsic	The main start-up motive is wish to be own boss	no
Push	The main start-up motive is (threat of) unemployment or dissatisfaction with wage job	no
Opportunist	The main start-up motive is discovery of market opportunity or the opportunity to earn higher income as self-employed as compared to paid employment	no
Work-life	The main start-up motive is better possibilities to combine work and personal life or necessity due to personal circumstances	no
Other	The main start-up motive is to be able to have availability of own financial means or just happened	no
Entrepreneurial objectives	*Dummy variables indicating that the entrepreneur finds the objective important*	
Improve own expertise	Important entrepreneurial objective is to improve expertise at start-up	no
Improve quality of products	Important entrepreneurial objective is to to improve the product quality at start-up	no
Maximize profits	Important entrepreneurial objective is to to maximize profits at start-up	no
Maximize revenues	Important entrepreneurial objective is to to maximize revenues at start-up	no
Competencies		
Education level	Dummies for education categories: (1) Low: primary school and pre-vocational secondary education, (2) Middle: general secondary education, (3) High: tertiary education and/or graduate level	no
Industry experience	Dummy with value 1 if respondent worked in the same industry in wage-employment	no
Entrepreneurial experience	Dummy with value 1 if respondent started at least one firm before this one	no
Entrepreneurial self-efficacy	To what extent does the entrepreneur believe that (s)he possesses entrepreneurial competencies? (1) very weak, (2) weak, (3) strong nor weak, (4) strong, (5) very strong. In the analysis we treat this variable as a continuous variable.	no
Risk attitude	To what extent does the entrepreneur believe that (s)he possesses the courage to take risk? (1) very weak, (2) weak, (3) strong nor weak, (4) strong, (5) very strong In the analysis we treat this variable as a continuous variable.	no
Social capital	Dummy with value 1 if the entrepreneur has frequent contact with other entrepreneurs outside of the regular business contacts	no
Firm-specific factors		
Full-time	Dummy variable indicating that the entrepreneur works at least 40 hours per week	no
Innovativeness	Dummy indicating that a large share of products and services are based on techniques that were not applied three years ago	no
Business cycle measure		
GDP growth	Annual percentage change of the Dutch GDP. Obtained from the IMF World Economic Outlook Database, 2009	yes

2.2 Descriptives

In this section we will present some descriptive statistics. The most important statistics will be stated and discussed over here. Remaining statistics can be found in section A.1.

The number of observations per year

In table 2.2 we stated an overview of the number of observations per year. We see that directly after the first year more than 400 entrepreneurs exit the panel. Still, we have a substantial number of observations left each year. About one fifth (18.90%) of the entrepreneurs stayed in the panel until the end. The reason for firms dropping out of the sample remains unknown. One of the explainations could be survival. Firms that do not survive drop out of the sample. Hence, still being in the sample after t years is conditional upon surviving up to year t. Since each year t

Table 2.2: Observations per year t

t	Observations	Percent	Percent of entrepreneurs
1	1402	22.47	100.00
2	1065	17.07	75.96
3	850	13.62	60.63
4	707	11.33	50.43
5	618	9.91	44.08
6	528	8.46	37.66
7	441	7.07	31.46
8	363	5.82	25.89
9	265	4.25	18.90
Total	6,239	100	

in the panel corresponds with multiple calendar years, we also listed the number of observations per calendar. This overview is stated in table A.1.

The number of employers

Of the 1,402 entrepreneurs 321 start employing personnel somewhere between $t = 1$ and $t = 9$, i.e. somewhere within the period of observation. In table 2.3 we stated the number of transitions per year. We see that a large share of the entrepreneurs that decided to employ personnel directly did this within the first year after start-up. Furthermore, we see that the vast majority (almost 70%) do this within the first three years after start-up. Thus, we find that about 23% of the entrepreneurs can be marked as job creators since they hire employees within the period of observations. This means that a large share of the entrepreneurs (77%) acts independently as solo-entrepreneur.

Table 2.3: Number of transitions from solo-entrepreneur to employer per year

t	Entrepreneurs	Percent
1	104	32.40
2	63	19.63
3	52	16.20
4	28	8.72
5	23	7.17
6	21	6.54
7	9	2.80
8	12	3.74
9	9	2.80
Total	321	100

The number of employees

If we only consider the group of 321 employers, then from table 2.4 we find that they employ 2.29 employees in the year they started employing. The maximum number of employees reported equals 30. Table 2.5 shows the distributions of the number of employees hired by the group

Table 2.4: Descriptive statistics of the number of employees hired by employers

Mean	Std.	Minimum	Maximum
2.29	3.16	1	30

321 employers. The majority of the employers hires a single employee in the year they start employing. The vast majority hires no more than 4 employees. A small group of 8.75% hires directly more than 5 employees.

Table 2.5: Distribution of the number of employees

Number of employees	Observations	Percent
1 employee	179	55.94
2 to 4 employees	113	35.31
more than 4 employees	28	8.75
Total	321	100

Summary statistics for the remaining variables

For an overview of summary statistics for the remaining variables used in this study, we refer to section A.1.

Chapter 3

The Transition from Solo Entrepreneur to Job Creator

3.1 Introduction

In this chapter we will focus on the transition from solo entrepreneur to employer. To model this transition we will specify several duration models. Duration models are used for modelling the duration of the time spent in one state before a transit to another state is made. We measure the duration within the state of solo self-employed by a nonnegative discrete random variable $\{T_i\}_{i=1}^{M}$, where M denotes the number of entrepreneurs. We can interpret this variable as the number of years since the start-up of the enterprise the entrepreneur operates without any employees. Thus, it is the age of the firm measured in years at the moment the entrepreneur hires his first employee(s). As we already have seen in the data description, a large share of the entrepreneurs do not have any employees during period there were tracked. For this group we do not observe a transition, but only know that $T_i > C_i^*$, where C_i^* is the number of consecutive years the entrepreneur was interviewed. In this case we speak of censored spells or censored durations. Still, there is much variation in the durations T_i across entrepreneurs. This chapter is devoted to finding the causes for these differences, and hence, to finding the determinants that underlie the decision to hire employees. As mentioned before, we will do this by means of a discrete time duration analysis (Lancaster, 1990). In the econometrics literature there are also duration models that treat the duration T_i as a continuous random variable (Lancaster, 1990). The main reason for this is that in continuous-time modelling one can rely on more elegant mathematics. On the other hand, in continuous-time modelling it is more difficult to incorporate time-varying characteristics of individuals into the models, while this is relatively easy for discrete-time models. Another advantage of discrete-time over continuous-time models is that the quantities that we derive from these models (such as the hazard and survivor function) have a clearer interpretation. The set-up of this chapter is as follows. We start with specifying the models. Next, we discuss how the parameters of the models are estimated. Then we present the results by discussing and comparing their implications. In the final section we will summarize our findings.

3.2 Model specification

3.2.1 The hazard function

Consider the random variable T_i. Suppose that the probability of making the transition from solo entrepreneur to employer would be the same for all individuals and all time periods and equal to λ. The probability that an entrepreneur will start as an employer is then given by:

$$\Pr[T_i = 1] = \lambda,\ i = 1, \ldots, M. \tag{3.1}$$

In the econometrics literature a duration is also called a spell. Hence, expression (3.1) is the probability that the spell will end within the first year after start-up. When a spell has ended the transit to the other state has been made. In our case that is the transition from solo self-employed to job creator. The probability that the spell ends after two years is equal to $\lambda(1-\lambda)$, and in general we can write

$$\Pr[T_i = t] = \lambda(1-\lambda)^{(t-1)},\ t = 1, 2, 3, \ldots$$

This is the probability mass function (pmf) of the geometric distribution with success probability λ. Hence, T_i is random variable that follows a geometric distribution. The cumulative density function (cdf) for this random variable therefore equals:

$$\begin{aligned} F(t) = \Pr[T_i \leq t] &= \Pr[T_i = 1] + \Pr[T_i = 2] + \ldots + \Pr[T_i = t] \\ &= \sum_{i=1}^{t} \lambda(1-\lambda)^{(i-1)} = \lambda \frac{1-(1-\lambda)^t}{\lambda} = 1-(1-\lambda)^t. \end{aligned} \tag{3.2}$$

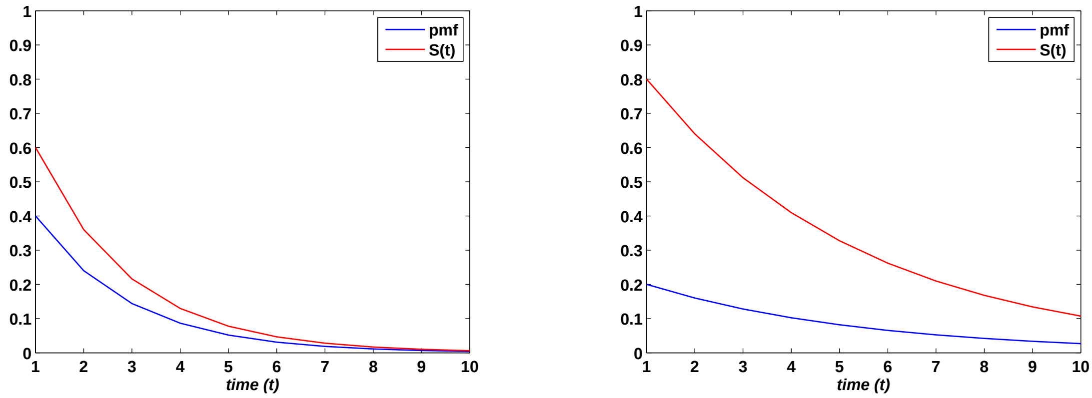
Note that we made use of the fact that $F(t)$ is the sum of a geometric series with $(1-\lambda)$ as the common ratio. We can now easily see that $F(0) = 0$ and $\lim_{t\to\infty} F(t) = 1$. $F(t)$ can be interpreted as the fraction of entrepreneurs that has become employer within the first t years after start-up. The survivor function $S(t)$, defined as $S(t) = 1 - F(t)$, is the remaining fraction that still acts as solo entrepreneur at time t.

The hazard function $\theta(t)$ is defined as the probability of becoming an employer at time t given that the entrepreneur was a solo entrepreneur until t. That is,

$$\theta(t) = \Pr[T_i = t | T_i \geq t] = \frac{\Pr[T_i = t]}{\Pr[T_i \geq t]} = \frac{\Pr[T_i = t]}{1 - F(t-1)} = \frac{\Pr[T_i = t]}{S(t-1)} = \lambda. \tag{3.3}$$

The hazard function (or hazard rate) is another way of characterizing the distribution of the durations T_i. Once the hazard function is known, the pmf and cdf of the duration can be derived. Since $\theta(t) = \lambda$, the hazard function can be interpreted as the fraction of entrepreneurs that become an employer in year t given that they were not in any of the years before. Since this fraction is assumed to be constant over time and across individuals, the hazard function is constant too.

A constant hazard rate may, however, not be a plausible assumption. It may for example vary over time and also be influenced by characteristics of the entrepreneur. To overcome this problem it is very common to specify the hazard function as some function of a set explanatory variables and time. Before we discuss the inclusion of explanatory variables in the model in more detail, let us first see how the functions introduced in this section look like. In figure 3.1 the the pmf (blue) and the survivor function (red) are shown for $\lambda = 0.4$ and $\lambda = 0.2$. Both the pmf and survivor function are exponentially declining towards zero. The higher the hazard rate the faster the functions decline. This is of course not surprising as a higher hazard rate implies that a greater share of the population of solo entrepreneurs becomes employer within t-th year after start-up.



(a) $\lambda = 0.4$ (b) $\lambda = 0.2$

Figure 3.1: The probability mass function and survivor function for different values of the hazard rate

3.2.2 Adding explanatory variables

A constant hazard rate may not be a plausible assumption. To allow the hazard probabilities to vary across individuals we parameterize the hazard function $\theta(t)$ by a function G and a vector of explanatory variables $\mathbf{x}_i$ (Cameron and Trivedi, 2005; Franses and Paap, 2001; Lancaster, 1990):

$$\theta_i = G(\alpha + \mathbf{x}_i'\beta), \tag{3.4}$$

where α is an intercept parameter and β is a parameter vector. θ_i now denotes the conditional probability of making the transition. Note that, for shortness, we use θ_i instead of $\theta_i(t)$. The index i is added to characterize the interindividual differences of the hazard function. Suitable choices for G are the standard normal distribution function (probit) and the logistic distribution function (logit), since θ_i is a probability. Discrete-time transition models are in that sense closely related to binary choice models. In each time period the entrepreneur faces the choice of hiring employees or not. Therefore, it is very common to use a logit or probit specification for the

probability that a spell will end given that it has not ended yet.
To complete the model specification, we also introduce time dependence. That is, we will not only allow the probability to vary across individuals, but also across time periods. We do this by also including a vector of time-varying variables $\mathbf{w}_{it}$ and a function of time t itself in (3.4), such that we obtain:

$$\theta_{it} = G(\alpha + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2),$$

where parameter vectors γ and $\delta = (\delta_1, \delta_2)$ are additional parameters to be estimated. The time-varying variables will include a measure of the business cycle that we added to the database. That is, $\mathbf{w}_{it}$ will contain the annual GDP growth and a one-year lag of the annual GDP growth. Using the logistic distribution function for G, the full model now reads:

$$\begin{aligned} \Pr[T_i = t] &= \theta_{it} \prod_{q=1}^{t-1} (1 - \theta_{iq}), \\ \theta_{it} &= G(\alpha + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2), \\ G(z) &= \frac{\exp(z)}{1 + \exp(z)}. \end{aligned} \tag{3.5}$$

We interpret the parameters by assessing how changes in the regressors $\mathbf{x}_i$, $\mathbf{w}_{it}$ and t affect the hazard probabilities. Since $G(\cdot)$ is a nonlinear function the effect of a change in any of the herefore mentioned quantities on the hazard probabilities is not immediately clear. If we consider the partial effect of $\mathbf{x}_i$ on the hazard rate we obtain:

$$\frac{\partial \theta_{it}}{\partial \mathbf{x}_i} = \beta \frac{\partial G}{\partial \mathbf{x}_i} = \beta \theta_{it}(1 - \theta_{it}).$$

Note that we make use of the fact that $dG(z)/dz = G(z)(1-G(z))$. Hence, the partial effects vary over the evaluation points of $\mathbf{x}_i$ due to nonlinearity. The question that arises is at which point of $\mathbf{x}_i$ should the partial effects be evaluated? Cameron and Trivedi (2005) propose to make use of the average sample transition probabilities (i.e. the average empirical hazard rate) That is, they propose to choose the evaluation point in such a way that $\theta_{it} = \bar{\theta}_{it}$. The measure for the partial effects is then given by $\hat{\beta}\bar{\theta}_{it}(1 - \bar{\theta}_{it})$. This measure, also known as the mean partial effect, is easy to compute, but has an non-negligible drawback. It is only easy to interpret the partial effect in case the regressor of interest is measured on a continuous scale. In case the independent variable is a binary indicator, this measure is not valid anymore as θ_{it} is not differentiable with respect to that variable. Therefore, we consider the odds ratio. This is the ratio between probability of becoming an employer and the probability of staying a solo entrepreneur in year t:

$$O(\mathbf{x}_i, \mathbf{w}_{it}, t) = \frac{\theta_{it}}{1 - \theta_{it}} = \exp(\alpha + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2).$$

Suppose that we would have a single explanatory variable x_i in our model with coefficient β. In that case the odds ratio would be equal to $O(w_i) = \exp(\alpha + \beta x_i)$. If x_i would be a binary indicator we can compare the odds ratios for both values of x_i:

$$O(x_i) = \begin{cases} \exp(\alpha) & \text{if } x_i = 0 \\ \exp(\alpha + \beta) & \text{if } x_i = 1 \end{cases}$$

The odds ratio corresponding with $x_i = 1$ equals the odds ratio for $x_i = 0$ multiplied by a factor $\exp(\beta)$. Hence, the effect of x_i on the odds ratio can be measured by $\exp(\beta)$. For $\beta > 0$ we have that individuals for which $x_i = 1$ the relative probability of becoming an employer at a given point in time is $\exp(\beta)$ times larger than for individuals with $x_i = 0$. In the same way we can use the odds ratio to interpret the parameters of continuous variables. In case x_i is continuous $\exp(\beta)$ can be interpreted as the multiplication factor for odds ratio that is associated with a one unit increase of x_i. When discussing the parameter estimates we will also pay attention to the proportionate increase of the odds ratios that is associated with a unit increase of several explanatory variables.

3.2.3 Incorporating unobserved heterogeneity

Observed heterogeneity refers to differences across individuals that we measure by the observed regressors $\mathbf{x}_i$, $\mathbf{w}_{it}$ and t. All other differences are known as unobserved heterogeneity. Both the observed and unobserved heterogeneity affect the hazard probabilities that are implied by model (3.5). If we would neglect unobserved heterogeneity this can affect the parameter estimates, and thus, the hazard probabilities. Cameron and Trivedi (2005) argue that accurate statements about duration dependent variables require the inclusion of unobserved heterogeneity in the model. To give a better understanding of this phenomenon, we will consider unobserved heterogeneity in a linear regression model. Suppose that a data generating process (dgp) is given by

$$y_i = \beta x_i + \gamma z_i + \varepsilon_i.$$

This model describes the dependent variable y_i as a linear function x_i and z_i. The error term ε_i is uncorrelated with both x_i and z_i. Suppose that we did not observed z_i and y_i is regressed on x_i alone. The model we obtain is then

$$y_i = bx_i + (\gamma z_i + \varepsilon_i) = bx_i + \eta_i,$$

where γz_i is absorbed into error term η_i. If we would apply OLS in this model, the estimator of β will be consistent if x_i and η_i are uncorrelated. This can only be the case if x_i and z_i are uncorrelated. In case there is no correlation, the unobserved heterogeneity is not an issue as the conditional expectation $\mathrm{E}[y_i|x_i]$ remains unchanged. On the other hand, if x_i and z_i are correlated we end up with the so-called omitted variable bias (see Heij et al. (2004) and Cameron and Trivedi (2005), among others). The conditional expectation of y_i given x_i will then be different as b is biased. In nonlinear models (such as duration models) unobserved heterogeneity causes more problems, even if there is no correlation between the observed and unobserved variables. As a consequence the estimated hazard rates will be biased (Heckman and Singer, 1984a). A motivating example that has widely been used to illustrate this effect is that of the high-risk and low-risk group described in Trussell and Richards (1985). Suppose that the sample under investigation can be divided into two groups: a high-risk group that has a high constant hazard rate, and a low-risk group that has a low constant hazard rate. If we are naware of this grouping, then the estimated overall hazard, which is a weighted average of the hazard rates of the two subgroups, does not have to be constant over time: suppose that both groups are equally represented in the sample and consist out of 100 individuals each, and the hazard for the high group equals 0.6 and the low group 0.2. In table 3.1 the transitions for each group are shown for the first three years.

Table 3.1: An illustration of the effect of unobserved heterogeneity

t	Exits high-risk	Exits low-risk	Aggregated hazard rate
1	$0.6 \times 100 = 60$	$0.2 \times 100 = 20$	$80/200 = 0.4$
2	$0.6 \times (100 - 60) = 24$	$0.2 \times (100 - 20) = 16$	$40/120 = 0.33$
3	$0.6 \times (100 - 60 - 24) = 10$	$0.2 \times (100 - 20 - 16) = 13$	$23/80 = 0.29$

In the final column the aggregate hazard rate is shown. We see a declining aggregated hazard rate, while the hazard rate for each group is constant. As a consequence we would erroneously conclude that the hazard rate is declining over time. This bias may then lead to a bias in the estimated parameters. The question that now arises is how to deal with unobserved heterogeneity in our model. In the literature there are two types of methods that incorporate unobserved heterogeneity: the random and fixed effects approach. To be able to explain the difference between the two methods, let $\mathbf{z}_i^*$ be the vector of unobserved covariates. In case these we would have observed this vector, the hazard function would have been written as:

$$\theta_{it} = G(\alpha + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2 + \mathbf{z}_i'^*\zeta),$$

where ζ is an additional parameter vector. Since $\mathbf{z}_i^*$ is unobserved it is impossible to obtain estimates of ζ. If we define $\alpha_i = (\alpha + \mathbf{z}_i'^*\zeta)$, then we characterize the interindividual differences that we do not observe by a random variable α_i. Writing the hazard rate as a function of α_i yields:

$$\theta_{it}(\alpha_i) = G(\alpha_i + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2).$$

In the random effects approach α_i is parameterized. Usually α_i is assumed to have particular distributional form. In that case the likelihood is written as a function of the covariate parameters and the parameters that are associated with the assumed distribution of α_i. In the fixed effects approach no assumptions about the distribution of α_i are made. The α_i's are then assumed to be parameters that have to be estimated for each individual, although they are not of primary interest. The advantage of the fixed effects approach is that no distributional form for the unobserved heterogeneity has to be specified and that the unobserved factors can be correlated with the observed covariates that are included in the model. On the other hand the number of parameters that has to be estimated increases dramatically and can therefore not be estimated consistently with maximum likelihood (Yamaguchi, 1986). Therefore we will apply the random effects approach in this research. Within the random effects approach there are two possibilities: (1) we can assume a continuous distribution (like the normal distribution) for the random effects α_i and (2) we can assume a so-called mass point distribution, where α_i can take a on a limited number of values. We will consider both options and formulate two extensions on model (3.5).

Continuous random effects (CRE)

We will assume a normal distribution for the random coefficients which is centered around α, i.e. $\alpha_i \sim N(\alpha, \sigma^2_{\alpha_i})$. The variance σ^2_α measures the amount of unobserved heterogeneity. The higher the variance, the greater unobserved differences there are. The probability of T_i conditional on α_i is now given by

$$\Pr[T_i = t|\alpha_i] = \theta_{it}(\alpha_i) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\alpha_i)) .$$

To obtain the unconditional probability we have to integrate with respect to α_i, yielding the pmf

$$\begin{aligned} \Pr[T_i = t] &= \int_{\alpha_i} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_i - \alpha}{\sigma_\alpha}\right) \Pr[T_i = t|\alpha_i] d\alpha_i \\ &= \int_{\alpha_i} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_i - \alpha}{\sigma_\alpha}\right) \left\{\theta_{it}(\alpha_i) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\alpha_i))\right\} d\alpha_i, \end{aligned}$$

where $\phi(z) = (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)$, the density function of the standard normal distribution. This integral does not exists in closed form and therefore needs to approximated using quadrature methods. The complete model specification with continuous random coefficients is:

$$\begin{aligned} \Pr[T_i = t] &= \int_{\alpha_i} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_i - \alpha}{\sigma_\alpha}\right) \left\{\theta_{it}(\alpha_i) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\alpha_i))\right\} d\alpha_i, \\ \theta_{it}(\alpha_i) &= G(\alpha_i + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2), \\ \alpha_i &\sim N(\alpha, \sigma_\alpha), \\ G(z) &= \frac{\exp(z)}{1 + \exp(z)}. \end{aligned} \tag{3.6}$$

Compared to the basic model there is one additional parameter to be estimated, namely σ_α. In $\sigma_\alpha \to 0$, the model with CRE simplifies to the basic model. In that case the intercept α_i does not vary across individuals and the unobserved heterogeneity is not of an issue. Hence, to test the model with CRE against the basic model, we have to test wheteher σ_α signifcantly differs from zero. This can be done using a likelihood ratio (LR) test. Note that, however, this is a one-sided test since the variance component σ_α^2 is always greater that zero. For variance components the LR test can still be done, but the LR test statistic will not follow a standard χ^2 distribution (Self and Liang, 1987). Later on in this chapter we will provide more details on this test.

Discrete random effects (DRE)

When we assume discrete random effects α_i is assumed to follow a mass point distribution. In that case α_i can take on limited number, say J, values $\{\xi_1, \xi_2, \ldots, \xi_J\}$. Let $\{\pi_j\}_{j=1}^J$, $\sum_{j=1}^J \pi_j = 1$, be the probability that α_i equals ξ_j. The pmf of T_i is then equal to

$$\begin{aligned}\Pr[T_i = t] &= \sum_{j=1}^{J} \pi_j \Pr[T_i = t | \alpha_i = \xi_j] \\ &= \sum_{j=1}^{J} \pi_j \theta_{it}(\xi_j) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\xi_j)) .\end{aligned}$$

This model is also known as a latent class or finite mixture model, which assumes that the population is composed out of J different subpopulations, with mixing proportions (or prior probabilities) $\{\pi_j\}_{j=1}^J$. An individual then is assigned a random coefficient depending on to which group he belongs. Hence, the model with DRE assumes that the unobserved heterogeneity can be characterized by a categorical variable. This approach has important advantage compared to the CRE approach: as argued in Heckman and Singer (1984b) it is difficult to verify whether a normal distribution is suitable in case of CRE as the α_i's are not observable. Therefore the latent class approach is more flexible in the sense that it can generate various kind of distributional shapes for the random effects. The full model specification for the model with discrete random coefficients is:

$$\begin{aligned}\Pr[T_i = t] &= \sum_{j=1}^{J} \pi_j \theta_{it}(\xi_j) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\xi_j)) , \\ \theta_{it}(\alpha_i) &= G(\alpha_i + \mathbf{x}_i'\beta + \mathbf{w}_{it}'\gamma + \delta_1 t + \delta_2 t^2), \\ \alpha_i &\in \{\xi_1, \xi_2, \ldots, \xi_J\} \\ \pi_j &= \Pr[\alpha_i = \xi_j], \\ G(z) &= \frac{\exp(z)}{1 + \exp(z)}. \end{aligned} \tag{3.7}$$

In sum, we have derived three discrete-time transition models for explaining the differences in durations T_i across entrepreneurs:

1. model (3.5) without unobserved heterogeneity,
2. model (3.6), incorporating unobserved heterogeneity as continuous random variable following a normal distribution (CRE),

3. model (3.6), incorporating unobserved heterogeneity as a discrete random variable (DRE).

We will now turn over to the discussion about estimating the model parameters and comparison of these models.

3.3 Parameter estimation and model comparison

In this section we will consider parameter estimation of each of the models that we specified before. Furthermore, we will also discuss methods that give us the ability to compare the models. The models will be estimated using maximum likelihood. Therefore, we will first start with deriving the likelihood function for each of the three models and discuss issues regarding maximization and inference.

3.3.1 Estimation of model (3.5)

Let t_i^* be the duration for entrepreneur i and let c_i^* be the number of consecutive years entrepreneur i was interviewed. We only observe the a duration t_i^* for entrepreneur i if the spell ends within the interval $[1, c_i^*]$. Define $t_i = \min(t_i^*, c_i^*)$ and let the binary variable $\{\{y_{it}\}_{i=1}^{M}\}_{t=1}^{t_i}$ indicate the state in which entrepreneur i is at time t. That is,

$$y_{it} = \begin{cases} 1 & \text{if entrepreneur } i \text{ has at least 1 employee in year } t \\ 0 & \text{if entrepreneur } i \text{ has no employees in year } t \end{cases}$$

For entrepreneurs that make a transition we observe a sequence $\{y_{i1}, y_{i2}, \ldots, y_{it_i}\}$, with $y_{it_i} = 1$ and $y_{i1} = y_{i2} = y_{i,t_i-1} = 0$. In that case t_i is the observed duration, and thus, a realization of the random variable T_i. In case of a censored spell $\{y_{it}\}_{t=1}^{t_i}$ is a sequence of zeros and t_i is simply the number of consecutive years the entrepreneur was interviewed. For entrepreneurs for which the spell ended at time t_i the likelihood equals the probability:

$$\Pr[T_i = t_i] = \theta_{it_i} \prod_{t=1}^{t_i-1} (1 - \theta_{it}) \,.$$

For entrepreneurs that do not start employing within the time interval $[1, t_i]$ we only know that $T_i > t_i$. Therefore the likelihood for a censored spell equals the probability of that event. Making use of result (3.2) we obtain

$$\Pr[T_i > t_i] = 1 - \Pr[T_i \leq t_i] = 1 - F(t_i) = \prod_{t=1}^{t_i} (1 - \theta_{it}) \,.$$

By using the censoring indicator y_{it} we can write the total likelihood L as:

$$L(\Psi) = \prod_{i=1}^{M} \prod_{t=1}^{t_i} \theta_{it}^{y_{it}} (1 - \theta_{it})^{(1-y_{it})} \,, \tag{3.8}$$

where Ψ the vector of parameters to be estimated. The contribution to the likelihood by individual i is a product of t_i terms. If t_i is equal to 1 for all individuals, the likelihood function simplifies to that of the binary choice (logit) model. In that sense the likelihood function in (3.8) can be regarded as a product of likelihood functions of a sequence of binary choice models (Jenkins, 1995). In fact, by stacking all observations and estimating a logit model with the censoring indicator as dependent variable, we can obtain unbiased estimates of Ψ (Jenkins, 1995). By running a stacked logit routine, we maximize the log-likelihood instead of the likelihood function itself for numerical purposes. The log of the likelihood function in (3.8) is given by:

$$\ln L(\Psi) = \sum_{i=1}^{M} \sum_{t=1}^{t_i} \left(y_{it} \ln \theta_{it} + (1 - y_{it}) \ln(1 - \theta_{it}) \right). \tag{3.9}$$

Hence, estimating model parameters for model (3.5) is a matter of constructing a censoring indicator y_{it} and organizing the data in such a way that we have t_i observations for each individual. Since the log-likelihood in (3.9) has no analytical solutions to the first order conditions, we numerically maximize this function using the Newton-Raphson method.

3.3.2 Estimation of model (3.6)

In the same way we derived the log-likelihood function for model (3.5) we can derive this for model (3.6). The likelihood conditional on the random coefficients can be written as:

$$L(\Psi | \alpha_1, \ldots, \alpha_M) = \prod_{i=1}^{M} \prod_{t=1}^{t_i} \theta_{it}(\alpha_i)^{y_{it}} \left(1 - \theta_{it}(\alpha_i)\right)^{(1-y_{it})}.$$

To obtain the unconditional likelihood we have to integrate individual i's likelihood with respect to α_i, resulting in

$$L(\Psi) = \prod_{i=1}^{M} \int_{\alpha_i} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_i - \alpha}{\sigma_\alpha}\right) \left\{ \prod_{t=1}^{t_i} \theta_{it}(\alpha_i)^{y_{it}} \left(1 - \theta_{it}(\alpha_i)\right)^{(1-y_{it})} \right\} d\alpha_i. \tag{3.10}$$

As mentioned before, the integral in (3.10) has no closed form solution and is usually computed numerically using the Gauss-Hermite quadrature. We will now briefly explain this method. Let l_i be likelihood of individual i, that is

$$l_i = \int_{\alpha_i} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_i - \alpha}{\sigma_\alpha}\right) \left\{ \prod_{t=1}^{t_i} \theta_{it}(\alpha_i)^{y_{it}} \left(1 - \theta_{it}(\alpha_i)\right)^{(1-y_{it})} \right\} d\alpha_i = \int_{\alpha_i} g(\alpha_i) d\alpha_i. \tag{3.11}$$

The Gauss-Hermite quadrature approximates integrals of the form $\int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx$ as follows:

$$\int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx \approx \sum_{j=1}^{K} w_j f(x_j),$$

where w_j is a weight and x_j an evaluation point and K the number of evaluation points. The weight is given by

$$w_j = \frac{2^{K-1} n! \sqrt{\pi}}{K^2 H_{j-1}(x_j)^2},$$

where $H_j(x)$ are known as the Hermite polynomials that are generated using the recursion

$$H_{j+1}(x) = \sqrt{2/(j+1)} x H_j(x) - \sqrt{j/(j+1)} H_{j-1}(x),$$

with $H_{-1} = 0$ (by definition) and $H_0 = \pi^{-1/4}$. The K evaluation points are obtained by solving $H_K(x) = 0$, i.e. they are the roots of the K-th order polynomial $H_K(x)$. To obtain the Gauss-Hermite approximation for the integral in (3.11) we write

$$\begin{aligned} l_i &= \int_{\alpha_i} g(\alpha_i) d\alpha_i = \int_{\alpha_i} \exp(-\alpha_i^2) \exp(\alpha_i^2) g(\alpha_i) d\alpha_i \\ &\approx \sum_{j=1}^{K} w_j \exp(\alpha_{ij}^2) g(\alpha_{ij}) \\ &= \sum_{j=1}^{K} w_j \exp(\alpha_{ij}^2) \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_{ij} - \alpha}{\sigma_\alpha}\right) \left\{ \prod_{t=1}^{t_i} \theta_{it}(\alpha_{ij})^{y_{it}} \left(1 - \theta_{it}(\alpha_{ij})\right)^{(1-y_{it})} \right\}, \end{aligned}$$

where $\{\alpha_{ij}\}_{j=1}^{K}$ is the sequence of evaluation points. Both the weights and evaluation points are already given in tables for a given K. The number of evaluation points K usually does not exceed 30. Using this approximation it is straightforward to derive the total approximated log-likelihood, which is given by:

$$\ln L(\Psi) \approx \sum_{i=1}^{M} \ln \left(\sum_{j=1}^{K} w_j \exp(\alpha_{ij}^2) \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha_{ij} - \alpha}{\sigma_\alpha}\right) \left\{ \prod_{t=1}^{t_i} \theta_{it}(\alpha_{ij})^{y_{it}} \left(1 - \theta_{it}(\alpha_{ij})\right)^{(1-y_{it})} \right\} \right). \tag{3.12}$$

To obtain estimates of Ψ we maximize this approximated log-likelihood using the Newton-Raphson method.

3.3.3 Estimation of model (3.7)

The final model for which we describe concerns regarding estimation is the finite mixture model in (3.7). The likelihood function for this model is given by

$$L(\Psi) = \prod_{i=1}^{M} \sum_{j=1}^{J} \pi_j \prod_{t=1}^{t_i} \theta_{it}(\xi_j)^{y_{it}} \left(1 - \theta_{it}(\xi_j)\right)^{(1-y_{it})}$$

and consequently the log-likelihood

$$\ln L(\Psi) = \sum_{i=1}^{M} \ln \left(\sum_{j=1}^{J} \pi_j \prod_{t=1}^{t_i} \theta_{it}(\xi_j)^{y_{it}} \left(1 - \theta_{it}(\xi_j)\right)^{(1-y_{it})} \right). \tag{3.13}$$

Again, we have to rely on numerical optimization methods to obtain estimates of Ψ. Since the likelihood function in (3.13) is very complicated, one usually maximizes this likelihood using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm assumes there is a latent variable $s_i \in \{1, \dots, J\}$ for each individual that indicates to which (latent) class he belongs. If s_i would have been observed we can incorporate this information into the likelihood function, yielding the so-called complete-data likelihood (denoted by L_c):

$$L_c(\Psi|\mathbf{s}) = \prod_{i=1}^{M}\prod_{j=1}^{J}\left(\pi_j \prod_{t=1}^{t_i} \theta_{it}(\xi_j)^{y_{it}}\,(1-\theta_{it}(\xi_j))^{(1-y_{it})}\right)^{I[s_i=j]},$$

where $\mathbf{s} = \{s_i\}_{i=1}^{M}$ and $I[\cdot]$ denotes the indicator function. Each individual is assumed to belong to a single class, that is $\sum_{j=1}^{J} I[s_i = j] = 1$. The EM algorithm is an iterative two-step procedure that applies a missing data augmentation scheme to obtain estimates of s_i. The two steps, the E-step and M-step, are described below.

E-step: compute the estimated log complete-data likelihood function with respect to $s|\Psi^{(v)}$ (i.e. the conditional distribution of s_i given the estimates of Ψ in that iteration). That is, we compute

$$Q(\Psi, \Psi^{(v)}) = \mathrm{E}_{s_i|\Psi^{(v)}}[\ln L_c(\Psi|s)],$$

where the log complete-data likelihood is given by

$$\begin{aligned}\ln L_c(\Psi) \;=\; & \sum_{i=1}^{M}\sum_{j=1}^{J}\sum_{t=1}^{t_i} I[s_i=j]\,(y_{it}\ln\theta_{it}(\xi_j) + (1-y_{it})\ln(1-\theta_{it}(\xi_j))) \\ & + \sum_{i=1}^{M}\sum_{j=1}^{J} I[s_i=j]\ln\pi_j. \end{aligned} \tag{3.14}$$

The unobserved cluster indicators $\{I[s_i = j]\}_{j=1}^{J}$ in this function are augmented by the estimated component memberships, i.e. the posterior probabilities. The estimated posterior probabilities $\hat{p}_{ij}$ are obtained by using Bayes' rule:

$$\hat{p}_{ij} = \frac{\pi_j \prod_{t=1}^{t_i} \theta_{it}(\xi_j)^{y_{it}}\,(1-\theta_{it}(\xi_j))^{(1-y_{it})}}{\sum_{j=1}^{J}\pi_j \prod_{t=1}^{t_i} \theta_{it}(\xi_j)^{y_{it}}\,(1-\theta_{it}(\xi_j))^{(1-y_{it})}}.$$

M-step: maximize the expected log complete-data likelihood with respect to Ψ to obtain the update $\Psi^{(v+1)}$:

$$\Psi^{(v+1)} = \arg\max_{\Psi} Q(\Psi, \Psi^{(v)}).$$

The two steps are repeated until there is no improvement in the log-likelihood in (3.13). Maximization of the log complete-data likelihood can be done using the Newton-Raphson method. Note that the two terms in (3.14) can be maximized separately. The second term can be maximized analytically, yielding the maximization problem

$$\max_{\pi_j} \sum_{i=1}^{M} \sum_{j=1}^{J} \hat{p}_{ij} \ln \pi_j$$
$$\text{s.t.} \sum_{j=1}^{J} \pi_j = 1.$$

Setting up the Lagrangian function we obtain

$$\begin{aligned} \Lambda(\pi_1, \ldots, \pi_J, \lambda) &= \sum_{i=1}^{M} \sum_{j=1}^{J} \hat{p}_{ij} \ln \pi_j + \lambda \left(\sum_{j=1}^{J} \pi_j - 1 \right) \\ &= \sum_{j=1}^{J} \ln \pi_j \sum_{i=1}^{M} \hat{p}_{ij} + \lambda \left(\sum_{j=1}^{J} \pi_j - 1 \right) \\ &= M \sum_{j=1}^{J} \ln \pi_j \bar{p}_j + \lambda \left(\sum_{j=1}^{J} \pi_j - 1 \right) \end{aligned}$$

where λ is the Lagrangian multiplier and $\bar{p}_j = M^{-1} \sum_{i=1}^{M} \hat{p}_{ij}$, i.e. the average posterior probability in class j. Taking the first order derivatives with respect to π_j and setting them equal to zero renders the first order conditions

$$\frac{\partial \Lambda}{\partial \pi_j} = M \frac{\bar{p}_j}{\pi_j} + \lambda = 0, \text{ for } j = 1 \ldots, J.$$

Solving this with respect to π_j gives

$$\pi_j = -\frac{M \bar{p}_j}{\lambda}. \tag{3.15}$$

The value for λ can be obtained by solving the equality constraint

$$1 = \sum_{j=1}^{J} \pi_j = -\frac{M}{\lambda}(\bar{p}_1 + \ldots + \bar{p}_J) = -\frac{M}{\lambda},$$

and hence, λ is set equal to $-M$. Plugging this back into equation (3.15) gives $\hat{\pi}_j = M^{-1} \sum_{i=1}^{M} \hat{p}_{ij}$. That is, in the M-step the updates of π_j are set equal to the average posterior probabilities within class j.

Creating clusters

After the EM algorithm has converged the posterior probabilities can be used to create clusters of individuals. We assign an individual i to cluster $j = 1, \ldots, J$ if $\hat{p}_{ij} = \max(\hat{p}_{i1}, \hat{p}_{i2}, \ldots, \hat{p}_{iJ})$. That is, an individual is assigned to cluster j if the posterior probability of belonging to that cluster is greater that the posterior probability of belong to any of the other clusters.

Selecting the number of latent classes J

The number of latent classes J is not known a priori. We will therefore run the EM algorithm several times for different values of J. The value of J that yields the best model will be selected. But the that question arises is: how do we find the best model? A simple approach would be to make use of the Likelihood Ratio (LR) test to test 2 models, one with J classes and another one with $(J + 1)$ classes, against each other. However, this test is strictly not valid as the additional parameter estimated in the alternative hypothesis (the model with $(J + 1)$ classes) is not identified under the null. This problem is also known as the Davies (1977) problem. Since there is no formal test that can be used to select the number of latent classes one usually relies on information criteria such as the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Both criteria assess the quality as a trade-off between model fit (in terms of likelihood) and number of estimated parameters. The BIC is defined as

$$\text{BIC} = k \ln N - 2 \ln L(\hat{\Psi}),$$

where $N = \sum_{i=1}^{M} t_i$, the total number of observations, $\hat{\Psi}$ is an MLE estimate of Ψ and k the dimension of Ψ. The AIC is defined as

$$\text{AIC} = 2k - 2 \ln L(\hat{\Psi}).$$

The only difference between both criteria is the way in which the number of estimated parameters is penalized. AIC uses a penalty term of 2 and BIC a term $\ln N$. The decision for the number of latent classes will be based on both information criteria.

3.3.4 Model comparison tests

There is no formal test for testing whether the models that we specified are the true models. However, it is possible to compare the models (in pairs of two) to see which of both is closest to the true model. In this way we can test the presence of unobserved heterogeneity, and thus, of the latent entrepreneurial ability. When comparing two models, we make a distinction between nested and nonnested models. If two models are said to be nested, then one model is a special case of the other. It can, for example, be obtained by a parameter restriction in the other model. This is the case for the basic model and CRE model. The basic model is nested in the CRE model and can be obtained from the CRE model by setting the variance of the random intercept equal to zero. If this is not possible, then we speak of nonnested models. This is the case for the remaining models.

Comparing nested models: the LR test

To test nested models against each other we make use of the LR test. The LR test is used for comparing the the basic model with the CRE model. As we mentioned before we can test whether the variance of the random intercept signficantly differs from zero using an LR test. Since this is a one-sided test, the distribution of the LR test statistic will not follow $\chi^2(1)$ distribution. Self and Liang (1987) show that for one-sided tests the asymptotic distribution of the LR statistic is approximately a 50:50 mixture of $\chi^2(0)$ and $\chi^2(1)$ distribution.

Comparing nonnested models: the Vuong test

To test the DRE against the basic model and against the CRE model we will make use of Vuong's closeness test (Vuong, 1989). The Vuong test is used to test which of two nonnested models is closest to the true model. The Vuong closeness test is an LR test based on the Kullback-Leibler Information Criterion (KLIC) (Kullback and Leibler, 1951). The null hypothesis is that the two models are equally close to the true model. The KLIC is used to measure of distance between two statistical models and defined as:

$$\text{KLIC} = \text{E}[\ln h(y_i|x_i)] - \text{E}[\ln f(y_i|x_i, \hat{\beta})],$$

where $h(y_i|x_i)$ is the true conditional density of y given x and $f(y_i|x_i, \hat{\beta})$ the density that is implied by the model. When we assume that $f(y_i|x_i, \hat{\beta})$ is not the true conditional density of y given x then the parameter estimates $\hat{\beta}$ are said to be the pseudo-true values of β. The model $f(y_i|x_i, \hat{\beta})$ that minimizes the KLIC is said to be the best model. Since KLIC ≥ 0 by definition, the best model is the model that maximizes $\text{E}[\ln f(y_i|x_i, \hat{\beta})]$. Stated alternatively, the best model is the model that yields the highest likelihood. Hence, a model should be chosen over another if the difference in likelihood is significantly positive. Suppose that there would be a competing model for f that yields the conditional density $g(y_i|x_i, \hat{\gamma})$. Then the null hypothesis of the Vuong closeness test is given by:

$$H_0\text{: } \text{E}_0\left[\ln \frac{f(y_i|x_i, \hat{\beta})}{g(y_i|x_i, \hat{\gamma})}\right] = 0,$$

where $\text{E}_0[\cdot]$ denotes the expectation under the null hypothesis. This hypothesis implies that the likelihood of model f equals that of model g. Let $\ln L_f$ and $\ln L_g$ be the log-likelihood of model f and g respectively. It can be shown (see Vuong (1989)) that the likelihood ratio $LR = \ln L_f - \ln L_g$ converges almost surely to the expectation under the null hypothesis, i.e.

$$\frac{1}{M} LR \overset{as}{\to} \text{E}_0\left[\ln \frac{f(y_i|x_i, \hat{\beta})}{g(y_i|x_i, \hat{\gamma})}\right].$$

Therefore under the null hypothesis it holds that

$$V = \frac{LR_M}{\sqrt{M}\hat{\omega}} \overset{d}{\to} N(0,1),$$

where $\hat{\omega}^2$ is defined as the sample variance of the individual likelihood ratios:

$$\hat{\omega}^2 \quad = \quad M^{-1} \sum_{i=1}^{M} \left[\ln \frac{f(y_i|x_i, \hat{\beta})}{g(y_i|x_i, \hat{\gamma})}\right]^2 - \left[M^{-1} \sum_{i=1}^{M} \ln \frac{f(y_i|x_i, \hat{\beta})}{g(y_i|x_i, \hat{\gamma})}\right]^2.$$

The null hypothesis will be rejected if V exceeds the critical values of the standard normal distribution. Since the number of parameters are allowed to vary across the models, a small adaption to the likelihood ratio LR_M is made to correct for the degrees of freedom. This correction is based on the AIC information criterion (BIC may also be used Vuong (1989)). Vuong proposes the following adapted version of the likelihood ratio:

$$\tilde{LR}_M = \ln L_f - \ln L_g - (k_f - k_g)$$

where k_f and k_g is the number of parameters that is associated with model f and g respectively.

3.4 Results

In this section we will apply the models to the data set that we described in the data description. Using information of the $M = 1,420$ entrepreneurs, the three duration models that we specified within this chapter will be estimated. That is, we will apply the following models:

1. model (3.5), i.e. the basic model not accounting for unobserved heterogeneity;
2. model (3.6), i.e. the continuous random effects model (CRE) that incorporates unobserved heterogeneity as continuous random variable following a normal distribution;
3. model (3.7), i.e. the discrete random effects (DRE) models incorporating unobserved heterogeneity as a discrete random variable. We sometimes refer to this models as 'the latent class model'.

The set-up of this section if as follows. First we will present the parameter estimates of the models in a comparative way. Next we will compare the models by means of the model comparison tests. We will end our discussion with policy implications and recommendations.

3.4.1 Parameter estimates

We shall now present the results implied by the models. We will start our discussion with model selection of the latent class model in (3.7). We ran the EM algorithm with 5 random starts for $J = 2$ up to $J = 5$. Then, for each J, we selected the outcome with the highest likelihood and computed the information criteria. The results of this procedure are shown in table 3.2. The

Table 3.2: Model selection for the latent class model

J	$\ln L$	BIC	AIC
2	-1101.68	2509.21**	2273.36
3	-1093.72	2510.76	2261.43**
4	-1093.72	2528.24	2265.43
5	-1093.71	2545.71	2269.43

BIC prefers a model with 2 classes, the AIC prefers a model with 3 classes. Going from 2 to 3 classes increases the likelihood, while going from 3 classes to a higher number of classes does not change anything to the likelihood. Therefore we will report the parameter estimates of the model 3 latent classes. A summary of this model can be found in table 3.3.

Table 3.3: Summary of the latent class model with 3 classes

Class	π_j	Observations	Entrepreneurs
1	0.38	2248	350
2	0.21	478	203
3	0.41	3513	849

We find that all classes contain a substantial number of entrepreneurs. Later on in this section we will interpret the clusters and describe some speficic features of them. We will now focus on the parameter estimates. Using the DRE model with 3 classes, we have listed the estimates of all three models in table 3.4. From left to right we have the estimates of the basic model (model (3.5)), the estimates of the CRE model (model (3.6)) and the DRE model with 3 latent classes (model (3.7)). Something that immediatly strikes is that both models that incorporate unobserved heterogeneity are very similar to each other in terms of significant relationships of certain explanatory variables. Furthermore, we find that the variance of the random intercept of the CRE model significantly differs from zero. This indicates that unobserved heterogeneity is of an issue. We will discuss and compare the results implied by the models step by step.

Age and gender

All three models indicate there is an age effect on the hazard probabilities. We find a significantly positive sign for age and a significantly negative sign for the squared age. This suggests that the relationship between age and the hazard probability is inverse U-shaped. A (crude) estimation of the age that maximizes the hazard probability can be found by dividing minus the parameter estimate of age by 2 times that of the squared age[1]. The estimates are shown in table 3.5. The hazard maximizing ages lie relatively close to each other for the different models. We find that entrepreneurs are most likely to become employers somewhere at the end of their thirties.
Regarding gender, we find no strong evidence for the existence of a gender effect, although in the basic model and DRE model the coefficient for male entrepreneurs is significantly positive at ten percent level.

Start-up motives

There is strong evidence that entrepreneurs who started a company to improve their work-life balance are less likely to become employers compared to the entrepreneurs with an intrinsic motivation, something that is not surprisingly: hiring employees would mean more time has to be spent in on tasks such as managing and coordinating the employees. Hence, less time will be left to be at home.
Furthermore, we find that the CRE model implies that entrepreneurs that were pushed into

[1] The standard error of this term is computed using the Delta method (Oehlert, 1992)

Table 3.4: Estimated model parameters for all three duration models. The standard errors are in parenthesis.

	Basic model		CRE		DRE	
Age	0.162***	(0.055)	0.375***	(0.106)	0.436***	(0.111)
Age squared	-0.002***	(0.001)	-0.005***	(0.001)	-0.006***	(0.001)
Male	0.251*	(0.149)	0.551*	(0.297)	0.447	(0.282)
Start-up motives						
Intrinsic (base)						
Push	-0.293	(0.180)	-0.676*	(0.358)	-0.522	(0.361)
Opportunist	0.063	(0.195)	0.353	(0.405)	0.924*	(0.490)
Work-life	-0.457**	(0.219)	-0.964**	(0.423)	-1.048***	(0.382)
Other	-0.094	(0.247)	-0.093	(0.511)	-0.163	(0.456)
Objectives						
Improve own expertise	-0.179	(0.149)	-0.240	(0.304)	0.172	(0.298)
Improve product quality	-0.136	(0.130)	-0.292	(0.265)	-0.378	(0.258)
Maximize profits	-0.183	(0.131)	-0.380	(0.270)	-0.286	(0.278)
Maximize revenue	0.569***	(0.133)	1.222***	(0.288)	1.174***	(0.295)
Competencies						
Educational level						
-Low (base)						
-Middle	0.069	(0.165)	0.061	(0.327)	-0.010	(0.307)
-High	-0.053	(0.180)	-0.201	(0.360)	-0.088	(0.377)
Industry experience	0.505***	(0.142)	1.130***	(0.300)	1.621***	(0.339)
Entrepreneurial experience	0.484**	(0.196)	1.193***	(0.428)	1.459***	(0.452)
Entrepreneurial self-efficacy	0.270***	(0.092)	0.648***	(0.194)	1.051***	(0.240)
Risk attitude	0.243***	(0.093)	0.420**	(0.190)	0.547***	(0.195)
Social capital	0.192	(0.154)	0.517	(0.321)	0.669**	(0.317)
Firm-specific factors						
Fulltime	0.882***	(0.139)	1.957***	(0.314)	2.143***	(0.297)
Innovativeness	0.103	(0.142)	0.333	(0.296)	0.704**	(0.321)
Time-varying covariates						
t	-0.020	(0.149)	0.816***	(0.250)	0.937***	(0.254)
t^2	0.002	(0.015)	-0.048**	(0.022)	-0.055**	(0.023)
GDP growth	-0.063	(0.065)	-0.137*	(0.083)	-0.143*	(0.085)
GDP growth (one-year lag)	0.195***	(0.068)	0.304***	(0.084)	0.324***	(0.088)
Intercepts						
Constant	-8.278***	(1.285)	-18.757***	(2.946)		
σ_α			2.900[a]	(0.346)		
ξ_1					-28.583***	(4.036)
ξ_2					-19.297***	(3.259)
ξ_3					-23.447***	(3.627)
Estimation info						
log-likelihood	-1110.22		-1098.80		-1093.72	
M (entrepreneurs)	1402		1402		1402	
N (effective sample)	6239		6239		6239	

* significant at 10% level
** significant at 5% level
*** significant at 1% level
[a] The variance σ_α^2 of the random intercept significantly differs from zero according to the LR test. The LR test-statistic equals 22.84 and is signifcant at 1% level.
Note: Industry dummies are included, but not reported in this table. For a complete table we refer to table A.10.

Table 3.5: Estimated hazard maximizing ages for all three models. The standard are in parenthesis and are computed using the Delta method

Model	Age
Basic model	33 (2.32)
CRE model	36 (1.80)
DRE model	39 (1.47)

entrepreneurship are also less likely to hire employees. Although the relationship is not very strong, this also is not surprising.

Next two these two motives, in the DRE model we also find a significantly positive coefficient for the opportunists. This would suggest that entrepreneurs that started a company out of a market opportunity or opportunity to earn more are more likely to hire employees than the ones that started out of an intrinsic motivation. This could be due to the fact that opportunists might be more ambitious.

In sum, we find a strong link between the hazard probability and the work-life balance motivation.

Objectives

Regarding objectives, we find that especially when entrepreneurs want to maximize their revenue they also create jobs. To increase revenue more activities have to be undertaken, and hence, the entrepreneur will need employees.

Competencies

We find no relationship between educational level and the hazard probability. We do find that experience is an important factor. Both industry and entrepreneurial experience are found to have a significantly positive effect on the hazard probability. An entrepreneur with more entrepreneurial and industry experience will be more familiar issues regarding entrepreneurship and the market in which (s)he operates. This awareness can help the entrepreneur to identify market opportunities, which in turn could lead to an expansion.

Next to these two effects the entrepreneurial self-efficacy and risk attitude are found to have a significantly positive effect on the hazard probability. The self-efficacy measures the entrepreneur's confidence in his or her entrepreneurial skills. Not surprisingly, we find that more confident entrepreneurs are more likely to hire employees. The same can be said for the risk attitude, as there are risks associated with employing personnel. Entrepreneurs that do not restrain from these risks are more likely to become employers.

According to the DRE model, the entrepreneur's social capital also affects the hazard probability in a positive way. More contacts with entrepreneurs outside the business might help the entrepreneur to gain knowledge about issues that each entrepreneur has to deal with. This knowledge and experiences might take away some of the entrepreneur's uncertainties, which in turn may lead to a higher hazard rate.

Firm-specific factors

If we consider firm-specific factors we find that the time the entrepreneurs spends in his business at start-up is an important predictor for the choice of becoming an employer. We find a strong positive relation, indicating that entrepreneurs that spend more time in their business have a greater likelihood of becoming employers.
The DRE model also implies that innovativeness, measured by a dummy indicating that the firms uses techniques that were not applied three years before, leads to a higher hazard rate.

Time-varying covariates

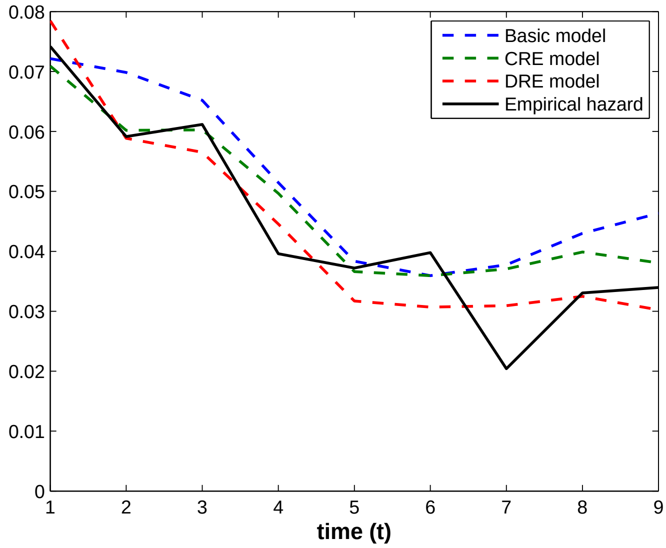
Finally, we discuss the time-varying covariates. In the basic model we find that there is no effect of time (firm age) itself on the hazard rate. In the models that incorporate unobserved heterogeneity we do find a time effect: the hazard probabilities increase over time, implying that as firm age increases the entrepreneur is more inclined to hire employees. As the the square of the firm's age is significantly negative in both models, the relationship with time becomes less steep as the firm grows older.
Furthermore, there is also evidence for the existence of a business-cycle effect on the hazard probabilities. In all models we find that the one-year lag of the GDP growth affects the hazard probabilities in a positive way. This effect is stronger in the CRE and DRE model. This implies that there is a relation between the decision of hiring employees and the economic climate in that period. The instanteneous GDP growth has a significantly negative effect on the hazard probability in the models that incorporate unobserved heterogeneity. This effect, however, is not as strong as that for the lagged GDP growth. The positive relationship with the lagged GDP growth indicates that economic growth might increase the entrepreneur's demand for labour, and hence, might imply a causal relationship between economic growth and job creation.

3.4.2 Model comparison

The aggregated hazard function

In the previous section we have seen that the models yield results that are quite the same. Especially the models that incorporate unobserved heterogeneity lie close to each other. Now that we have compared and the parameter estimates we will consider another feature of the models, namely the predicted hazard rates. Using the parameter estimates and the definition of the hazard rate in (3.3) we computed the hazard probabilites for all entrepreneurs. In the CRE model an integral has to be evaluated when computing probability $\Pr[T_i = t]$. We do this again using the Gauss-Hermite quadrature. For each entrepreneur we computed the hazard rates for $t = 1, \ldots, 9$. The aggregated hazard is obtained by taking averages of the predicted hazard for each year t. The aggregated hazard rates are plotted in figure 3.2. The black solid line indicates the empirical aggregated hazard function. We observe that all models do a good job in reproducing the downward shape of the aggregated hazard rate. The basic model slightly overestimates the hazard rate. This could be a direct consequence of the unonbserved heterogeneity. The models

Figure 3.2: Aggregated hazard functions



that incorporate unobserved heterogeneity have a signifcantly positive sign for the time. At first sight, this might be inconsistent with the picture of the aggregated hazard rates, since the rates seem to increase over time at micro level. But as we already have seen in the example of Trussell and Richards (1985) there might be a grouping in the data that we are unaware of. The downward shape may thus be a direct consequence of aggregating over these groups. Using the latent class model we were able to identify three groups of entrepreneurs in the data. Will now take a closer look at the clusters that we found with this model.

Interpreting the latent classes

Using the estimated posterior probabilites $\hat{p}_{ij}$ we have divided the entrepreneurs into 3 groups. An entrepreneur is assigned to a cluster if the posterior probability of belong to that group is higher than for any of the other groups. In table 3.3 we already gave a summary of these latent classes. In table 3.6 we provide some more details. The intercept for class 1 is lowest among all,

Table 3.6: An interpretation of the latent classes

Class	Label	ξ_j	Entrepreneurs	Employers	$\mathrm{E}[T_i \mid T_i \leq 9]$
1	Conscious sole	-28.583	350	4 (1.14%)	5.25
2	Aspired employer	-19.297	203	203 (100%)	2.35
3	Necessity employer	-23.447	849	114 (13.4%)	4.21

implying that the (baseline) hazard rate for this class is by definition lower that that of the other two groups. Out of the 350 entrepreneurs within this group only 4 entrepreneurs become employers in the period of observation. Hence, it is very unlikely that an entrepreneur that belongs to this group will start employer personnel. The 4 entrepreneurs that became employers did this

5 years after start-up averagely. Hence, hiring employees for this group is not an option: working as sole proprietor was a conscious choice.

In the second group, consisting out of 203 entrepreneurs, we find that all of them started employing within the first 9 years after start-up. This also happens relatively fast, on average 2 years after start-up. This could indicate that they already had these plans before they started their firm. Hence, hiring employees was no large step for this group of entrepreneurs, and might have been something that they have aspired.

The final group, which has an intercept that lies between that of group 1 and 2, consists out of a relative small group of employers. The entrepreneurs in this group that started employing personnel did this on average 4 years after start-up. This indicates that for this group of entrepreneurs hiring employees is a large(r) step, but not necessarily something that is unreachable. The entrepreneurs in this group that did not start employing during the observation period can be regarded as potential employers. Hence, hiring employees for this group is something that may take a while and needs to be necessary.

We will now consider the aggregated hazard rates per class. For each latent class we computed the empirical aggregated hazard rate. Next to the empirical hazard rate, we also computed the hazard rate that is implied by the DRE model. For each entrepreneur we did this using only the random intercept that was assigned to the group in which he or she belongs, i.e. we used the pmf

$$\Pr[T_i = t] = \theta_{it}(\xi_j) \prod_{q=1}^{t-1} (1 - \theta_{iq}(\xi_j)) \text{ when entrepreneur } i \text{ belongs to class } j.$$

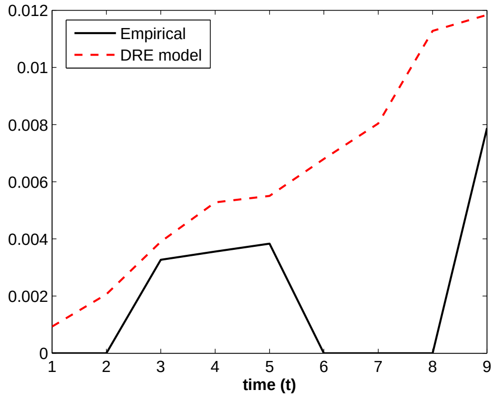
The hazard rates are then obtained by dividing this probability by $S(t-1)$, for $t = 1, .., 9$ and $S(0) = 1$. That is, we again made use of result (3.3). The rates we obtained are shown in figure 3.3. We now see that the empirical hazard rates are (on average) increasing over time[2]. The DRE model fails to reproduce the increasing shape of the hazard for the second class. Using the latent classes that are provided by the DRE model, we are now able to see that the hazard rate indeed increases over time.

Results of model comparison tests

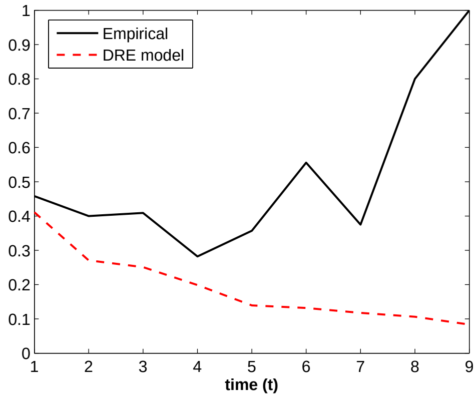
The results of the model comparison tests are stated in table 3.7. The LR test, used for testing the basic model against the CRE model, favours the CRE model. Using the Vuong test to test the basic model and DRE model against the DRE model resulted in a tie. That is, according to the Vuong test the basic model and DRE model are equally close to the true model. The same can be said for the CRE and DRE model. A reason for the fact that the Vuong test fails to reject the null hypothesis in case of the basic model against the DRE, while the LR test does not, might lie in the fact that the DRE model has 5 additional parameters to be estimated. Compared to the basic model, the CRE model has only a single additional parameter to be estimated. Based

[2]Note that the scale on the y-axis is different for each class. For class 1 we indeed observe an increasing hazard rate, but this increase is relatively small.

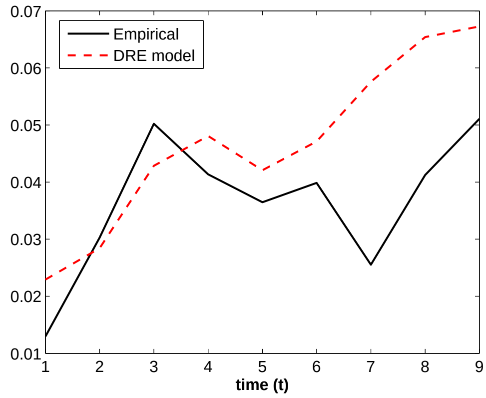
Figure 3.3: Aggregated hazard rates per latent class



(a) Class 1: conscious sole



(b) Class 2: aspired employer



(c) Class 3: necessity employer

on the LR test we may conclude the unobserved heterogeneity is signifcant in size.

Table 3.7: Results of model comparison tests

Compared models	Test	Test statistic	p-value	Preferred model
Basic model vs CRE	one-sided LR	22.84	< 0.0001	CRE
Basic model vs DRE	Vuong	-1.61	0.1072	none
CRE vs DRE	Vuong	-0.65	0.2313	none

3.4.3 Policy implications

To be able to develop policy that will increae the share of employers within the total entrepreneurial population it is not only important to know which factors influence the likelihood of becoming an employer, but also how large their impact is. To be able to make a judgement about the impact of the factors that do influence the hazard rate, we computed the proportionate increase of the odds ratio that is associated with a unit increae of these variables. The results are based on the DRE model and are stated in table 3.8. For age, we find that a one-year increase in

Table 3.8: Proporationate increase of the odds ratio that is associated with a unit increase of each of the variables based on the DRE model

Variable	Proporationate increase odds (%)
Age	55
Work-life (start-up motive)	-65
Maximize revenue	224
Industry experience	406
Entrepreneurial experience	331
Entrepreneurial self-efficacy	186
Risk attitude	73
Social capital	95
Fulltime	752
Innovativeness	102
t	155
GDP Growth (one-year lag)	38

age increases the relative probability of becoming an employer with 55%. As we have seen before the relationship with age is inverse u-shaped and peaking in the late thirties. Hence, after this age the hazard rate decreases.

Instruments that can be used to increase job creation by entrepreneurs are the entrepreneurial self-efficacy and risk attitude. This can be accomplished by introducing courses to increase the entrepreneur's self-efficacy and risk attitude. Also the age of the firm (measured by t) can be useful. As firm age increases the entrepreneur will be more likely to hire entrepreneurs. Hence, stimulation of job creation will be more effective when the firm already exists for a couple of years.

We also observe that entrepreneurs with industry experience are 4 times more likely to become employers that ones who do not. Entrepreneurs that had a firm before are 3.3 times more likely to become employers. Hence, experience can be regarded as an important factor. Therefore it should be made easier for individuals to start a new firm when they already entrepreneurial experience and/or experience within the industry they are trying to operate.

3.5 Conclusion

In this chapter we have considered the decision of the entrepreneur to hire employees. We did this using discrete-time duration models. We formulated a basic duration model and two extensions of this model that incorporate heterogeneity. In one extensions we treat unobserved heterogeneity as a continuous random variable and in the other as a discrete random variable. We found that the unobserved heterogeneity is of an issue. Hence, the latent entrepreneurial ability is significant in size.

We found that the age of the entrepreneur affects the decision to hire employees. This relationship is inversely U-shaped peaking at an age of about 36 years. We also find that entrepreneurs that entrepreneurs who started a firm to improve their work-life balance are less likely to hire employees. The remaining factors that we found to influence the employer decision do this all in a positive way. These factors include the objective of the entrepreneur to maximize revenue, experience within the industry in which he operates, his entrepreneurial experience, self-efficacy, risk attitude and the time that is spent in the company. Furthermore, we find that as firm age increases the prevalence of hiring employees also increases. Finally, we also found a business cycle effect: the choice of hiring employees is positive related to the economic situation in that period.

Chapter 4

Modelling Firm Size Using Count Data Models

4.1 Introduction

Once the entrepreneur has decided to hire employees, a conscious decision regarding the actual number of employees has to be made. Most of the entrepreneurs hire a single employee in the year they start employing personnel. Still, there is a great share of entrepreneurs that directly hires more than one employee. As we have seen in the data description this can go up to 30 employees. To explain these differences, we will specify several count data models using the same explanatory variables as we did in the previous chapter. The set-up will also be the same as in the previous chapter. That is, we will formulate a basic count data model, and also two extensions of this model that incorporate unobserved heterogeneity in a continuous and discrete way, respectively. An alternative to count models are linear regression models. An appealing feature of count models is that they take the non-negative and discrete character of the dependent variable into account, whereas the linear model does not. In that sense count models are regarded as an improvement over the linear model when the dependent variable is a count.

The analysis in this chapter will be based on the sample of entrepreneurs that decided to hire employees somewhere within the first 9 years after start-up. In the previous chapter(s) we have seen that of the 1,420 entrepreneurs within our sample, 321 made the transition from solo entrepreneur to employer. Thus, the sample will solely consist out of the group of 321 employers and the interest lies in the number of employees they hire in the year they switch over from solo-entrepreneur to employer. Since we only consider the year in which the entrepreneur decided to hire employees, the analysis in this chapter will be based on a cross-section of 321 entrepreneurs.

The set-up of this chapter is as follows. We start with specifying the models. Next, we discuss how the parameters of the models are estimated. Then we present the results by discussing and comparing their implications. In the final section we will summarize our findings.

4.2 Model specification

4.2.1 The dependent variable

Let w_i, $i = 1, \ldots, M$, be the number of employees hired by employer i and M be the total number of employers. Since the group of employers has at least 1 employee we have that $w_i > 0$, and $\Pr[w_i = 0] = 0$. However, (basic) count models assume that observing a zero outcome is not impossible. One way to overcome this problem is by constructing a zero-truncated count model that takes the specific feature of w_i being greater than zero into account. Another more easy, yet effective, method is by taking $y_i = (w_i - 1)$ as dependent variable, instead of w_i. In that case the dependent variable (y_i) can take on zero values but is still a count and has a slightly different interpretation. As w_i is the actual number of employees that is hired by the entrepreneur, y_i can be interpreted as the number of additional individuals that are employed by the entrepreneur given that he already has one employee. The count models in the remainder of this chapter will be specified using this definition of y_i.

4.2.2 The Poisson model

The most basic count model is the Poisson count model (Gilbert, 1982). This model is based on the Poisson probability distribution. The most basic version of this model assumes that the underlying random variable Y_i of y_i follows a Poisson distribution with parameter λ, i.e.

$$Y_i \sim \text{Poisson}(\lambda).$$

A random variable is said to follow a Poisson distribution if its probability mass function is given by

$$p(y_i) = \Pr[Y_i = y_i] = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\ , y_i = 0, 1, 2, \ldots.$$

The mean and variance of Y_i are then equal to $\mathrm{E}[Y_i] = \mathrm{V}[Y_i] = \lambda$. This model does not include explanatory variables. Explanatory variables are added by parameterizing the mean of the random variable, i.e. the mean of Y_i is written as a function of the explanatory variables and some set of parameters. Let $\mathbf{x}_i$ be the vector of regressors of individual i. In the standard framework one normally uses the exponential mean parameterization. That is, the conditional mean is written as:

$$\lambda_i = \mathrm{E}[Y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta). \tag{4.1}$$

Using this parameterization, an interpretation of the parameters can be given by considering the first-order derivatives:

$$\frac{\partial \mathrm{E}[Y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \beta \exp(\mathbf{x}_i'\beta) = \beta \mathrm{E}[Y_i|\mathbf{x}_i].$$

That is, a unit increase in the j-th regressor x_j leads to an increase of the expectation by $\beta_j \mathrm{E}[Y_i|\mathbf{x}_i]$. Hence, β_j measures the relative change in the conditional expectation due to a unit change in x_j. Suppose that $\beta_j = 0.1$. In that case a unit increase in x_j would increase the expectation with $0.1 \times \mathrm{E}[Y_i|\mathbf{x_i}]$. That is, the expectation would increase with 10 percent. The full model specification for the Poisson regression model reads as follows:

$$\begin{aligned} \Pr[Y_i = y_i|\lambda_i] &= \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \ldots \\ \lambda_i &= \mathrm{E}[Y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta). \end{aligned} \tag{4.2}$$

4.2.3 Extensions with unobserved heterogeneity

Again, we will consider unobserved heterogeneity. In the previous section we have seen that not accounting for unobserved heterogeneity adequately can lead to a bias in the parameter estimates. In Poisson regression models the conditional mean gets affected by unobserved heterogeneity. This can be seen as follows. Let $\mathbf{z}_i^*$ be a vector of unobserved covariates that should be included the model. In that case the true conditional expectation of y_i should be written as:

$$\begin{aligned} \tilde{\lambda}_i &= \exp(\mathbf{x}_i'\beta + \mathbf{z}_i^{*'}\zeta) \\ &= \exp(\mathbf{x}_i'\beta)\exp(\mathbf{z}_i^{*'}\zeta) \\ &= \exp(\mathbf{x}_i'\beta)u_i \\ &= \lambda_i u_i. \end{aligned}$$

Hence, the unobserved heterogeneity is now characterized by random variable $u_i = \exp(\mathbf{z}_i^{*'}\zeta)$. It is assumed that u_i is uncorrelated with any of the other regressors. Furthermore, $\mathrm{E}[u_i]$ can be normalized to 1 as long as the model has an overall constant (Winkelmann, 2003). Again, we will parameterize u_i and formulate two extensions on the Poisson model in (4.2). One extension assuming a continuous random distribution for u_i and another assuming a discrete random distribution.

Unobserved heterogeneity as continuous random variable: the Negative Binomial model

Since we are dealing with counts it must hold that $u_i > 0$, such that the conditional mean is non-negative. Hence, when assuming a continuous probability distribution we have to take into account that u_i is always positive. In the literature we find that the gamma distribution is most commonly used for this purpose (Greenwood and Yule, 1920). A random variable u is said to follow a gamma distribution if its pdf is given by

$$g(u|\alpha, \delta) = \frac{\alpha^\delta}{\Gamma(\alpha)} u^{\alpha-1} \exp(-\delta u), \; u, \alpha, \delta > 0,$$

where α and δ are free parameters such that $\mathrm{E}[u] = \alpha/\delta$ and $\mathrm{Var}[u] = \alpha/\delta^2$, and $\Gamma(\alpha)$ is the Gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty \exp(-t)t^{\alpha-1}dt.$$

Using this parameterization we are now able to derive the pmf of the counts y_i. When we normalize the mean of u_i to one (as we may), it must hold that $\alpha = \delta$, and thus, $\mathrm{Var}[u_i] = \alpha^{-1}$. That is, the number of free parameters now equals one instead of two. The pmf of y_i given the unobserved differences u_i is now given by

$$\begin{aligned}\Pr[y_i|\lambda_i, u_i] &= \frac{\exp(-\tilde{\lambda}_i)\tilde{\lambda}_i^{y_i}}{y_i!} \\ &= \frac{\exp(-\lambda_i u_i)(\lambda_i u_i)^{y_i}}{y_i!}.\end{aligned}$$

To obtain the pmf unconditional on the unobserved heterogeneity u_i we have to integrate with respect to u_i, yielding the density

$$\begin{aligned}\Pr[y_i|\lambda_i] &= \int_0^\infty f(y_i|\lambda_i, u_i) g(u_i, \alpha) du_i \\ &= \int_0^\infty \frac{\exp(-\lambda_i u_i)(\lambda_i u_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} u_i^{\alpha-1} \exp(-\alpha u_i) du_i \\ &= \frac{\alpha^\alpha \lambda_i^{y_i}}{y_i!\Gamma(\alpha)} \int_0^\infty \exp(-u_i(\lambda_i + \alpha)) u_i^{y_i+\alpha-1} du_i \\ &= \frac{\alpha^\alpha \lambda_i^{y_i}}{y_i!\Gamma(\alpha)} \frac{\Gamma(y_i+\alpha)}{(\lambda_i+\alpha)^{y_i+\alpha}} \\ &= \tilde{p}(y_i|\lambda_i).\end{aligned}$$

Now, by making use of the fact that $\Gamma(n) = (n-1)!$ for $n = 1, 2, 3, \ldots$, we can rewrite this density to

$$\begin{aligned}\tilde{p}(y_i|\lambda_i) &= \frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)\Gamma(y_i+1)} \frac{\alpha^\alpha \lambda_i^{y_i}}{(\lambda_i+\alpha)^{y_i+\alpha}} \\ &= \frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)\Gamma(y_i+1)} \left(\frac{\alpha}{\lambda_i+\alpha}\right)^\alpha \left(\frac{\lambda_i}{\lambda_i+\alpha}\right)^{y_i}.\end{aligned}$$

This density function is known as the negative binomial density. Hence, when assuming a gamma distribution for u_i we end up the Negative Binomial (NB) count model. The NB model is a more general version of the Poisson model. In the NB model the conditional mean of Y_i given $\mathbf{x}_i$ remains unchanged due to normalization of the mean of u_i:

$$\mathrm{E}[Y_i|\lambda_i] = \tilde{\lambda}_i = \lambda_i \mathrm{E}[u_i] = \lambda_i.$$

Therefore, parameter interpretation in the NB model stays the same as in the Poisson model. The conditional variance of Y_i given $\mathbf{x}_i$, however, does change to

$$\mathrm{V}[Y_i|\lambda_i] = \lambda_i \left(1 + \frac{1}{\alpha}\lambda_i\right).$$

Since $\alpha > 0$ the conditional variance is always greater than the conditional mean. In that sense the NB model is less restrictive than the Poisson model, as in the Poisson model the mean equals the variance. In case $\alpha \to \infty$ the NB model simplifies to the Poisson model. Usually α is not estimated directly, but instead we estimate $\omega = \alpha^{-1}$. Testing the NB model against the Poisson is then equivalent to testing whether $\omega = 0$. This can be achieved by means of a single-sided LR test.
Having $\omega = \alpha^{-1}$ the full NB model is given by:

$$\begin{aligned} \Pr[Y_i = y_i|\lambda_i, \omega] &= \frac{\Gamma(y_i + \omega^{-1})}{\Gamma(\omega^{-1})\Gamma(y_i + 1)} \left(\frac{\omega^{-1}}{\lambda_i + \omega^{-1}}\right)^{\omega^{-1}} \left(\frac{\lambda_i}{\lambda_i + \omega^{-1}}\right)^{y_i}, y_i = 0, 1, 2, \ldots \\ \lambda_i &= \mathrm{E}[Y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta). \end{aligned} \tag{4.3}$$

Unobserved heterogeneity as discrete random variable: the latent class (LC) Poisson model

In case we assume a discrete distribution for u_i, then u_i is assumed to take on a limited number of values (say J), i.e. $u_i \in \{\tau_1, \tau_2, \ldots, \tau_J\}$. Let $\pi_1, \pi_2, \ldots, \pi_J$, $\sum_{j=1}^{J} \pi_j = 1$, be the probabilities assigned to each of the possible outcomes of u_i, that is, $\pi_j = \Pr[u_i = \tau_j]$. The pdf of Y_i given λ_i and $u_i = \tau_j$ is now given by

$$\Pr[Y_i = y_i|\lambda_i, u_i = \tau_j] = \frac{\exp(-\lambda_i\tau_j)(\lambda_i\tau_j)^{y_i}}{y_i!}. \tag{4.4}$$

To guarantee that $u_i > 0$ is must hold that $\tau_j > 0$. This can be obtained by expressing τ_j as $\tau_j = \exp(\xi_j)$. Instead of estimating τ_j directly we estimate ξ_j, $j = 1, \ldots, J$. The parameters $\{\xi_j\}_{j=1}^{J}$ can be interpreted as random intercepts belonging to each of the latent classes. If we let $\tilde{\lambda}_{ij}$ be the conditional expectation of Y_i given $\mathbf{x}_i$ within class j, then this can be seen by writing:

$$\begin{aligned} \tilde{\lambda}_{ij} &= \mathrm{E}[Y_i|\lambda_i, u_i = \tau_j] = \lambda_i\tau_j \\ &= \exp(\mathbf{x}_i'\beta + \xi_j). \end{aligned}$$

This means that the model cannot have an overall intercept since it is not identified. Therefore, we are not allowed to normalize the mean of u_i to 1. In the LC Poisson model the expectation of Y_i given $\mathbf{x}_i$ therefore equals:

$$\mathrm{E}[Y_i|\lambda_i] = \lambda_i \mathrm{E}[u_i] = \lambda_i \sum_{j=1}^{J} \pi_j \exp(\xi_j).$$

That is, the conditional mean in the LC model is slightly different than that of the previous models. Parameter interpretation, however, still remains the same as the coefficients β still measure the relative change due to a unit increase in any of the covariates.
To complete the specification, the density of Y_i given $\mathbf{x}_i$ unconditional on u_i is needed. Using

(4.4) we obtain the unconditional density as follows:

$$
\begin{aligned}
\Pr[Y_i = y_i|\lambda_i] &= \sum_{j=1}^{J} \pi_j \Pr[Y_i = y_i|\lambda_i, u_i = \tau_j] \\
&= \sum_{j=1}^{J} \pi_j \frac{\exp(-\lambda_i\tau_j)(\lambda_i\tau_j)^{y_i}}{y_i!} \\
&= \sum_{j=1}^{J} \pi_j \frac{\exp(-\tilde{\lambda}_{ij})\tilde{\lambda}_{ij}^{y_i}}{y_i!}.
\end{aligned}
$$

That is, the unconditional density is a weighted average of J Poisson models, each having an own intercept parameter. Having derived this density, we now are able to specify the full LC Poisson model:

$$
\begin{aligned}
\Pr[Y_i = y_i|\{\tilde{\lambda}_{ij}\}_{j=1}^{J}] &= \sum_{j=1}^{J} \pi_j \frac{\exp(-\tilde{\lambda}_{ij})\tilde{\lambda}_{ij}^{y_i}}{y_i!}\ , y_i = 0, 1, 2, \ldots \\
\tilde{\lambda}_{ij} &= \mathrm{E}[Y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta + \xi_j).
\end{aligned} \tag{4.5}
$$

To summarize this chapter up to this point, we have derived a Poisson model for the number of employees that are hired at the moment the entrepreneur starts employing personnel. Next to this model we have also derived two extensions of this model that incorporate unobserved heterogeneity. That is, in total we have specified 3 models:

1. model (4.2): the basic Poisson model;
2. model (4.3):the Negative Binomial (NB) model incorporating unobserved heterogeneity as continuous random variable following a gamma distribution;
3. model (4.5): the Latent Class (LC) Poisson model incorporating unobserved heterogeneity as a discrete random variable.

The next section discusses parameter estimation of these models.

4.3 Parameter estimation

In section we will discuss parameter estimation for the count models that we specified. The basic Poisson model and NB model can be estimated straightforward using maximum likelihood. For the LC Poisson we will derive an EM algorithm.

4.3.1 Estimation of the Poisson model (4.2)

Let M be the number of employers and $\{y_i\}_{i=1}^M$ be the sequence of realizations of the underlying random variable Y_i. The likelihood function for the Poisson model is then given by:

$$L(\beta) = \prod_{i=1}^{M} \Pr[Y_i = y_i|\lambda_i].$$

The corresponding log-likelihood is therefore given by:

$$\begin{aligned}
\ln L(\beta) &= \sum_{i=1}^{M} \ln \Pr[Y_i = y_i|\lambda_i] \\
&= \sum_{i=1}^{M} \ln \left(\frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \right) \\
&= \sum_{i=1}^{M} \left(y_i \mathbf{x}_i'\beta - \exp(\mathbf{x}_i'\beta) - \ln y! \right).
\end{aligned}$$

This log-likelihood function is known to be globally concave. Therefore, optimization via the Newton-Raphson method will yield unique parameter estimates.

4.3.2 Estimation of the Negative Binomial model (4.3)

For the NB model the log-likelihood is obtained as follows:

$$\begin{aligned}
\ln L(\beta, \omega) &= \sum_{i=1}^{M} \ln \Pr[Y_i = y_i|\lambda_i, \omega] \\
&= \sum_{i=1}^{M} \ln \left\{ \frac{\Gamma(y_i + \omega^{-1})}{\Gamma(\omega^{-1})\Gamma(y_i + 1)} \left(\frac{\omega^{-1}}{\lambda_i + \omega^{-1}} \right)^{\omega^{-1}} \left(\frac{\lambda_i}{\lambda_i + \omega^{-1}} \right)^{y_i} \right\} \\
&= \sum_{i=1}^{M} \Bigg\{ \ln \Gamma(y_i + \omega^{-1}) - \ln \Gamma(\omega^{-1}) - \ln \Gamma(y_i + 1) \\
&\quad + \omega^{-1} \left(\ln \omega^{-1} - \ln(\lambda_i + \omega^{-1}) \right) + y_i \left(\ln \lambda_i - \ln(\lambda_i + \omega^{-1}) \right) \Bigg\},
\end{aligned}$$

where $\lambda_i = \exp(\mathbf{x}_i'\beta)$. Note that the term $\ln \Gamma(y_i + 1)$ can be ignored when using the Newton-Raphson method as it does not depend on β nor ω.

4.3.3 Estimation of the Latent Class Poisson model (4.5)

For the LC Poisson model we will derive an EM algorithm since optimization of its log-likelihood via the Newton-Raphson method can give problems due to the complexity of this function. Let

$\Psi = \{\beta, \pi_1, \ldots, \pi_J, \xi_1, \ldots, \xi_J\}$. The likelihood function of model (4.5) is then given by:

$$\begin{aligned} L(\Psi) &= \prod_{i=1}^{M} \Pr[Y_i = y_i | \{\tilde{\lambda}_{ij}\}_{j=1}^{J}] \\ &= \prod_{i=1}^{M} \sum_{j=1}^{J} \pi_j \frac{\exp(-\tilde{\lambda}_{ij}) \tilde{\lambda}_{ij}^{y_i}}{y_i!}. \end{aligned}$$

Its corresponding log-likelihood function therefore reads as follows:

$$\begin{aligned} \ln L(\Psi) &= \sum_{i=1}^{M} \ln \Pr[Y_i = y_i | \{\tilde{\lambda}_{ij}\}_{j=1}^{J}] \\ &= \sum_{i=1}^{M} \ln \left(\sum_{j=1}^{J} \pi_j \frac{\exp(-\tilde{\lambda}_{ij}) \tilde{\lambda}_{ij}^{y_i}}{y_i!} \right). \end{aligned} \tag{4.6}$$

Maximization of the log-likelihood in (4.6) will be done using the EM algorithm. Again, the assumption is that there exists a latent variable $s_i \in \{1, \ldots, J\}$ for each individual that indicates to which (latent) class he belongs. Having observed this variable would give us the ability to incorporate this information into the likelihood function, yielding the complete-data likelihood function:

$$L_c(\Psi | \mathbf{s}) = \prod_{i=1}^{M} \prod_{j=1}^{J} \left(\pi_j \frac{\exp(-\tilde{\lambda}_{ij}) \tilde{\lambda}_{ij}^{y_i}}{y_i!} \right)^{I[s_i = j]},$$

where $\mathbf{s} = \{s_i\}_{i=1}^{M}$ and $I[\cdot]$ denotes the indicator function. The complete-data log-likelihood therefore equals

$$\begin{aligned} \ln L_c(\Psi | \mathbf{s}) &= \sum_{i=1}^{M} \sum_{j=1}^{J} \ln \left\{ \left(\pi_j \frac{\exp(-\tilde{\lambda}_{ij}) \tilde{\lambda}_{ij}^{y_i}}{y_i!} \right)^{I[s_i = j]} \right\} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{J} I[s_i = j] \ln \pi_j + \sum_{i=1}^{M} \sum_{j=1}^{J} I[s_i = j] \left(y_i(\mathbf{x}_i'\beta + \xi_j) - \exp(\mathbf{x}_i'\beta + \xi_j) - \ln y! \right). \end{aligned}$$

Having set-up the complete data likelihood, we are now able to describe the E and the M-step of the EM algorithm.

E-step: In the E-step we compute the expected log complete-data likelihood function with respect to $s|\Psi^{(v)}$. That is, we will augment the the cluster indicators $\{I[s_i = j]\}_{j=1}^{J}$ by making use of the conditional distribution of s_i given the estimates of Ψ in the v-th iteration. For each individual we compute the posterior probability of being in a specific cluster using Bayes rule.

Letting $\hat{p}_{ij}$ denote that probability, we have

$$\hat{p}_{ij} = \frac{\pi_j \left(\exp(-\tilde{\lambda}_{ij})(\tilde{\lambda}_{ij})^{y_i}\right)/y_i!}{\sum_{j=1}^{J}\pi_j \left(\exp(-\tilde{\lambda}_{ij})(\tilde{\lambda}_{ij})^{y_i}\right)/y_i!}.$$

Hence, the expected complete-data likelihood is given by

$$\begin{aligned} Q(\Psi, \Psi^{(v)}) &= \mathrm{E}_{s_i|\Psi^{(v)}}[\ln L_c(\Psi|s)] \\ &= \sum_{i=1}^{M}\sum_{j=1}^{J}\hat{p}_{ij}\ln\pi_j + \sum_{i=1}^{M}\sum_{j=1}^{J}\hat{p}_{ij}\left(y_i(\mathbf{x}_i'\beta+\xi_j) - \exp(\mathbf{x}_i'\beta+\xi_j) - \ln y!\right). \end{aligned}$$

M-step: In the M-step we maximize the expected log complete-data likelihood with respect to Ψ to obtain the update $\Psi^{(v+1)}$:

$$\Psi^{(v+1)} = \arg\max_{\Psi} Q(\Psi, \Psi^{(v)}).$$

Just as in the M-step of the EM algorithm for the DRE duration model, we can maximize the two terms separately. The latter term is maximized using the Newton-Raphson method. As we have shown before, updates of the prior probabilities are obtained by setting them equal to the average posterior probabilities $\hat{p}_{ij}$ of their latent class, i.e. $\pi_j^{(v)} = M^{-1}\sum_{i=1}^{M}\hat{p}_{ij}^{(v)}$, $j = 1, \ldots, J$. Selection of the number of classes J to use in the latent class model will again be based on the information criteria we introduced in the previous chapter.

4.4 Results

In this section we will present the results of the count data models that we specified in this chapter. Using the cross-section of 321 entrepreneurs that decided to make the transition from solo-entrepreneur to employer, we estimate the parameters of the count models. That is, we estimate the parameters of

1. model (4.2), i.e. the basic Poisson model that does not account for unobserved heterogeneity;
2. model (4.3), i.e. the Negative Binomial model that models unobserved heterogeneity as a gamma distributed random term with normalized mean;
3. model (4.5), i.e. the Latent Class Poisson model incorporating unobserved heterogeneity as a discrete random variable.

4.4.1 Model selection for the LC Poisson model

Before presenting the estimated model parameters, we will first pay attention to selection of the number of latent classes of the latent class (LC) Poisson model. As we did in the previous chapter for the DRE model, we ran the EM algorithm for $J = 2$ up to $J = 5$ classes, using 5 random starts. Then, for each J we selected the solution that yielded the highest likelihood and computed the corresponding values of the information criteria. The outcome of this procedure is shown in table 4.1.

Table 4.1: Model selection for the Latent Class Poisson model

J	$\ln L$	BIC	AIC
2	-429.27	1060.43**	928.54
3	-427.50	1068.43	929.01
4	-421.17	1067.31	920.35**
5	-421.17	1078.85	924.35

From table 4.1 we observe that the BIC criterion prefers a model with 2 classes while the AIC criterion prefers one with 4 classes. We have selected the model with 2 classes for two reasons. The first reason is that we prefer parsimony, just for the sake of simplicity. A second reason is that when creating clusters using the model with 4 latent classes we end up with an empty class. This makes it difficult to give an interpretation to that class. The solution with 2 classes is summarized in table 4.2. The first class contains 64 entrepreneurs and the second class 256. Hence, both classes

Table 4.2: Interpretation of the classes of the LC Poisson model

Class	Label	π_j	Entrepreneurs	Average firm size
1	High growth	0.23	64	5.95
2	Low growth	0.77	256	1.37

are substantially filled. The average number of employees hired by entrepreneurs that fall into class 1 is about 6 while this is about 1 for the ones that fall into class 2. That is, entrepreneurs in class 1 hire on average 6 times more employees than those in class 2. Therefore, class 1 can be interpreted as high growth firms, while class 2 can be interpreted as low growth firms.

4.4.2 Parameter estimates

Using the solution with 2 classes of the LC Poisson model we have listed the estimated parameters of all three models in table 4.4. We will move through this table in steps.

Age and gender

Just as in the duration models, there is support for the presence of an age effect. We again find that age itself is positive related to the size of the firm. Since the square of age is significantly negative, the relationship between firm size and age is inverse U-shaped. The top of this parabola can be found by again dividing minus the parameter estimate of age by 2 times that of age squared. For each model the age that maximizes firm size is stated in table 4.4.2. We find that entrepreneurs aged somewhere at the end of thirties are the best in creating jobs. This is consistent with the results that we found in the previous chapter regarding the decision to become

Table 4.3: Estimated model parameters for the count models. The standard errors are in parenthesis.

	Poisson		NB		LC Poisson	
Age	0.291***	(0.066)	0.234**	(0.107)	0.246***	(0.015)
(Age squared)/100	-0.398***	(0.087)	-0.334**	(0.141)	-0.333***	(0.005)
Male	-0.081	(0.134)	-0.058	(0.267)	-0.338*	(0.189)
Start-up motives						
Intrinsic (base)						
Push	0.910***	(0.148)	0.795***	(0.303)	1.125***	(0.208)
Opportunist	-0.246	(0.174)	-0.253	(0.331)	-0.025	(0.227)
Work-life	0.609***	(0.178)	0.369	(0.379)	-0.111	(0.237)
Other	-0.507**	(0.249)	-0.343	(0.436)	-0.072	(0.321)
Objectives						
Improve own expertise	-0.070	(0.127)	-0.068	(0.247)	0.039	(0.162)
Improve product quality	0.122	(0.116)	0.166	(0.229)	-0.329*	(0.170)
Maximize profits	0.155	(0.111)	0.251	(0.218)	-0.236	(0.158)
Maximize revenue	0.068	(0.124)	0.115	(0.234)	0.076	(0.182)
Competencies						
Educational level						
-Low (base)						
-Middle	0.821***	(0.172)	0.678**	(0.283)	0.675***	(0.222)
-High	0.646***	(0.190)	0.361	(0.313)	0.642**	(0.250)
Industry experience	-0.001	(0.135)	-0.202	(0.260)	-0.257	(0.198)
Entrepreneurial experience	0.581***	(0.141)	0.613**	(0.309)	0.509***	(0.177)
Entrepreneurial self-efficacy	0.327***	(0.087)	0.245	(0.176)	0.351***	(0.130)
Risk attitude	0.139	(0.092)	0.040	(0.183)	0.143	(0.126)
Social capital	-0.171	(0.137)	0.036	(0.264)	0.012	(0.189)
Firm-specific factors						
Fulltime	0.239*	(0.136)	0.281	(0.244)	0.446**	(0.182)
Innovativeness	0.513***	(0.126)	0.558**	(0.252)	0.703***	(0.176)
Time-varying covariates						
t	-0.578***	(0.137)	-0.829***	(0.266)	-0.439*	(0.226)
t^2	0.068***	(0.014)	0.088***	(0.028)	0.055**	(0.023)
GDP growth	-0.134**	(0.064)	-0.233*	(0.124)	-0.139	(0.098)
GDP growth (one-year lag)	0.278***	(0.077)	0.160	(0.133)	0.344***	(0.130)
Intercepts						
Constant	-8.784***	(1.506)	-5.915**	(2.473)		
ω			1.674[a]	(0.259)		
ξ_1					-6.946***	(1.446)
ξ_2					-9.602***	(1.453)
Estimation info						
log-likelihood	-583.26		-434.79		-429.27	
M (effective sample)[b]	320		320		320	

* significant at 10% level
** significant at 5% level
*** significant at 1% level
[a] The variance ω of the unobserved heterogeneity term significantly differs from zero according to the LR test. The LR test-statistic equals 296.94 and is significant at 1% level.
[b] The initial sample size for this analysis was 321. As 1 entrepreneur reported an erroneous number of employees, (s)he was left out of the analysis, bringing the effective sample down to 320.
Note: Industry dummies are included, but not reported in this table. For a complete table we refer to table A.2.

Table 4.4: Estimated turning points of the inverse U-shaped relation between age and firm size. The standard errors are in parenthesis and computed using the Delta method

Model	Age
Poisson model	37 (1.08)
NB model	35 (2.48)
LC Poisson model	37 (2.00)

an employer.
The results indicate that there is no evidence for the existence of a gender effect.

Start-up motives

If we consider the start-up motives we find that especially entrepreneurs who started their business because of a push factor hire relatively more employees once they decide to become employers. In the duration analysis done in the previous chapter this effect was not present. A possible explanation for this finding is that intrinsically motivated entrepreneurs might pursue less firm growth due to the fact that they became entrepreneurs to have the ability to work independently. Hiring more employees would hinder this, since in that case the entrepreneur would also be confronted with managing tasks. Entrepreneurs that were pushed into entrepreneurship might have to make that trade-off between working independently and managing employees to a lesser extent.

Objectives

We find that objectives do not affect the number of employees hired once the entrepreneur has become an employer. Hence, entrepreneurs base their decision regarding firm size on other aspects than objectives.

Competencies

In the previous chapter we did not find any relationship between the educational level and the decision to become an employer. In this chapter we do find evidence for the existence of a relationship between the firm size and the educational level of the founder. We find that in all three models the dummy variable indicating middle educated entrepreneurs is positively significant. Furthermore, we find that the dummy for highly educated entrepreneurs is also positively significant in both Poisson models. This result is not surprising, as managing a growing business requires the right knowledge and the ability to deal with complex situations. Highly educated entrepreneurs are more likely to posses the right knowledge to deal with difficulties.
Next to education, we again find that entrepreneurial experience and self-efficacy are factors that influence the firm size in positive way. These factors also play a role in the decision to become employer.

Firm specific factors

In the duration analysis done in the previous chapter we found that the time the entrepreneur spends in the firm strongly affected the decision to become an employer in a positive way. The Poisson models in this chapter indicate that the time the entrepreneur spends working for his company tends to increase the firm's size. This effect, however, is not as strong as the one we found in the duration analysis.

We do find strong evidence for the effect of innovativeness on firm size. It turns out that firms with a high share of products based on techniques that were not applied three years ago, have more employees. This suggests that innovativeness may lead to firm growth.

Time-varying covariates

In this analysis t denotes the number of years since start-up after which the entrepreneur made the transition from solo-entrepreneur to employer. We find a significantly negative sign, indicating that the longer it takes to make the transition, the less employees are hired. This means that entrepreneurs who start employing personnel in the first year after start-up are likely to hire more employees than ones who do that after, say, 3 years. We also find a significantly positive sign for t^2. This suggests that that relationship between t and actual firm size is non-linear and might be U-shaped.

If we consider the business cycle we find the following. Both Poisson models show a significant positive relationship between the lagged GDP growth and firm size. The instantaneous GDP growth has a negative effect on firm size according to the basic Poisson model and the Negative Binomial model. This effect, however, is not convincing as this effect is not found using the LC Poisson model and only holds at a significance of 10 percent. The lagged GDP growth is found to have a postive effect on firm size according the basic Poisson and LC Poisson model. This is consistent with the result that we found in the previous chapter. Therefore, the main finding here is that job creation by entrepreneurs is positively affected by the business cycle.

4.4.3 Model comparison tests

Having discussed the parameters of the models, we will now investigate which model lies closest to the the true model. To compare the basic Poisson model with the NB model, we use an one-sided LR test, since these models are nested. For the remaining comparisons we use the Vuong test described in the previous chapter. The results of these tests are shown in table 4.5. In turns out that the models that incorporate unobserved heterogeneity are favored. Testing the NB model against the latent class Poisson model results in a tie. Therefore, we may assume that the latent entrepreneurial ability is significant in size within the context of the employee decision.

4.5 Conclusion

This chapter is devoted to the decision regarding the actual number of employees to employ once the entrepreneur has decided to become an employer. Using the sample of 321 entrepreneurs that made the switch from solo-entrepreneur to employer in the observation period, we estimated the

Table 4.5: Results of model comparison tests

Compared models	Test	Test statistic	p-value	Preferred model
Basic Poisson vs NB	one-sided LR	296.94	< 0.0001	NB
Basic Poisson vs LC Poisson	Vuong	-3.90	0.0001	LC Poisson
NB vs LC Poisson	Vuong	-0.76	0.4454	none

parameters of several count models that describe the number of employees that were hired in the year that transition was made. The set-up was the same as in the previous chapter. That is, we formulated a basic count model and two extensions of this model that incorporate unobserved heterogeneity in continuous and discrete way respectively. The results give us reasons to believe that the latent entrepreneurial ability is significant in size.

Again we found that age is inversely U-shaped related to the firm size. The top of this shape is found that an age of approximately 36 years. If the entrepreneur started his firm out of necessity, then we find the he hires relatively more employees (compared to those who started out of an intrinsic motivation). We also find that the educational level, entrepreneurial experience and self-efficacy of the entrepreneur lead to a greater firm size. An other factor that increases firm size is innovativeness, measured by a dummy variable indicating that the entrepreneur uses a large share of products that are based on techniques that were not applied three years earlier. The moment in time at which the transition from solo-entrepreneur to employer is made, also plays are role. We find a negative relationship with firm age, indicating that the faster the switch is made, the more personnel will be employed. That is, entrepreneurs who decide to directly start employing hire on average more employees than ones who do that at a later moment in time.

Finally, the business cycle that we found in the previous chapter, is also found using the count data models. That is, we find that firm size is positively related to the GDP growth rate.

Chapter 5

Conclusion

In this paper we examine how several characteristics of an entrepreneur affects his decision to hire labour. This decision process was split up in two distinct decisions. The first decision is the employer decision, i.e. the (binary) decision whether or not to hire employees and thus to become job creator. The second decision, which is conditional upon the first decision, is the employee decision. This is the decision about the actual number of employees to hire. Using data of Dutch start-ups, we analyzed the first decision be means of a discrete-time duration analysis. The latter decision was analyzed using count data models. We paid special attention to the presence of unobserved heterogeneity. Within the context of both decision, unobserved heterogeneity can be interpreted as the latent entrepreneurial ability (i.e. talent or intelligence). A first conclusion of our study is that the latent entrepreneurial ability is significant in size: model comparison test conducted in this paper favour the models that incorporate unobserved heterogeneity. A second conclusion is that the best job creators are aged between 35 and 40, possess entrepreneurial experience and are confident about their entrepreneurial skills. Within the context of the each of the two decision we find the following.

The employer decision

We find that entrepreneurs who founded a firm to improve their work-life balance are less likely to become job creators. The remaining factors that we found to influence the employer decision do this all in a positive way. These factors include the objective of the entrepreneur to maximize revenue, experience within the industry in which he operates, his entrepreneurial experience, self-efficacy, risk attitude and the time that is spent in the company. We also find that the prevalence of becoming a job creator is positively related to the business cycle.

The employee decision

If we focus on the job creators and analyze how their characteristics influence firm size, then we find the following. Entrepreneurs that started a firms out of necessity, hire relatively more employees compared to those who started out of an intrinsic motivation. We also find that the educational level, entrepreneurial experience and self-efficacy of the entrepreneur lead to a greater firm size. An other factor that increases firm size is innovativeness, measured by a dummy variable indicating that the entrepreneur uses a large share of products that are based on techniques that were not applied three years earlier. The moment in time at which the transition from solo-entrepreneur to employer is made, also plays are role. We find a negative relationship with firm age, indicating that the faster the switch is made, the more personnel will be employed. That is, entrepreneurs who decide to directly start employing hire on average more employees than ones who do that at a later moment in time. Also, for the employee decision we find a positive relation with the business cycle.

Bibliography

Barkham RJ. 1994. Entrepreneurial characteristics and the size of the new firm: A model and an econometric test. *Small Business Economics* **6**: 117–125.

Baumol W. 1998. *Is Entrepreneurship always Productive?* New York: United Nations.

Birch D. 1978. The job creation process. Unpublished report, MIT Program on Neighbourhood and Regional change prepared for the Economic Development Administration. U.S. Department of Commerce. Washington, D.C.

Bosma NS, Van Praag CM, Thurik AR, De Wit G. 2004. The value of human and social capital investments for the business performance of startups. *Small Business Economics* **23**: 227–236.

Burke AE, FitzRoy FR, Nolan MA. 2002. Self-employment wealth and job creation: The roles of gender, non-pecuniary motivation and entrepreneurial ability. *Small Business Economics* **19**: 255–270.

Cameron AC, Trivedi PK. 2005. *Microeconometrics: Methods and Applications.* Camebridge University Press.

Caroll R, Holtz-Eakin D, Rider M, Rosen HS. 1995. Income taxes and entrepreneur's use of labor. Income Taxes and Entrepreneur's Use of Labor, Centre for Economic Policy Studies, Working Paper 32.

Davies RB. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**: 247–254.

De Kok JMP, Ichou A, Verheul I. 2010. New firm performance: does the age of founders affect employment creation? EIM Research Report H201015.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society* **39**: 1–38.

Franses PHF, Paap R. 2001. *Quantitative Models in Marketing Research.* Cambridge University Press.

Gilbert C. 1982. *Econometric models for discrete (integer valued) economic processes.* The Athens School of Economics and Business Science.

Greenwood M, Yule GU. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society* **83**: 255–279.

Heckman J, Singer B. 1984a. Econometric duration analysis. *Journal of Econometrics* **24**: 63–132.

Heckman J, Singer B. 1984b. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52**: 271–320.

Heij C, De Boer P, Franses PHF, Kloek T, Van Dijk HK. 2004. *Econometric Methods with Applications in Business and Economics.* Oxford University Press.

Henley A. 2005. Job creation by the self-employed: The roles of entrepreneurial and financial capital. *Small Business Economics* **25**: 175–196.

Jenkins S. 1995. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* **57**: 120–138.

Kullback S, Leibler RA. 1951. On information and sufficiency. *Annals of Mathematical Statistics* **22**: 79–86.

Lancaster T. 1990. *The Econometric Analysis of Transition Data.* Cambridge University Press.

OECD. 2000. *The partial renaissance of self-employment.* OECD Employment Outlook.

Oehlert GW. 1992. A note on the delta method. *The American Statistician* **46**: 27–29.

Schumpeter JA. 1934. *The Theory of Economic Development.* Oxford University Press: New York.

Self SG, Liang K. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**: 605–610.

Trussell J, Richards T. 1985. Correcting for unmeasured heterogeneity in hazard models using the heckman-singer procedure. *Sociological Methodology* **15**: 242–276.

Van Praag CM, Cramer JS. 2001. The roots of entrepreneurship and labour demand. *Economica* **68**: 45–62.

Vuong QH. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**: 307–333.

Westhead P, Cowling M. 1995. Employment change in independent owner-managed high-technology firms in great britain. *Small Business Economics* **7**: 111–140.

Winkelmann R. 2003. *Econometric Analysis of Count Data*. Springer Verlag.

Yamaguchi K. 1986. Alternative approaches to unobserved heterogeneity in the analysis of repeatable events. *Sociological Methodology* **16**: 213–249.

Appendix A

Additional tables

A.1 Summary statistics

In this appendix we provide summary statistics for the variables used in this study.

Table A.1: Observations per calendar year

Year	Observations	Percent
1998	475	7.61
1999	857	13.74
2000	1,081	17.33
2001	856	13.72
2002	718	11.51
2003	611	9.79
2004	529	8.48
2005	446	7.15
2006	357	5.72
2007	217	3.48
2008	92	1.47
Total	6,239	100

Table A.2: Industry

Industry	Firms	Percent
Manufacturing	45	3.21
Construction	210	14.98
Wholesale	78	5.56
Retail	143	10.20
Hotels and restaurants	20	1.43
Motor vehicles	21	1.50
Transport	31	2.21
Business and financial services	543	38.73
Other services	311	22.18
Total	1,402	100

Table A.3: Distribution of the entrepreneurial age by gender

	Gender		
Age	Male	Female	Total
Younger than 30 years	226	85	311 (22.18%)
30-44 years	507	294	801 (75.13%)
45 years or older	226	64	290 (20.68%)
Total	959 (68.40%)	443 (31.60%)	1,402

Table A.4: Start-up motives

Start-up motive	Entrepreneurs	Percent
Intrinsic	741	52.85
Push	234	16.69
Opportunist	137	9.77
Work-life	197	14.05
Other	93	6.63
Total	1,402	100

Table A.5: Entrepreneurial objectives

	Entrepreneurs	Percent
Improve own expertise	1,069	23.75
Improve product quality	779	44.44
Maximize profits	659	53.00
Maximize revenue	626	55.35

Table A.6: Distribution of the educational level of the entrepreneurs

Educational level	Entrepreneurs	Percent
Low	383	27.32
Middle	463	33.02
High	556	39.66
Total	1,402	100

Table A.7: Remaining dummy variables

Variable	Entrepreneurs	Percent
Industry experience	834	59.49
Entrepreneurial experience	109	7.77
Social capital	218	15.55
Fulltime	336	23.97
Innovativeness	651	46.43

Table A.8: Entrepreneurial self-efficacy

Entrepreneurial self-efficacy	Entrepreneurs	Percent
Very weak	7	0.50
Weak	70	4.99
Weak nor strong	546	38.94
Strong	621	44.29
Very strong	158	11.27
Total	1,402	100

Table A.9: Risk attitude

Risk attitude	Observations	Percent
Very weak	3	0.21
Weak	63	4.49
Weak nor strong	443	31.60
Strong	684	48.79
Very strong	209	14.91
Total	1,402	100

A.2 Extended parameter estimates

In this section we stated extended versions of the tables containing the parameter estimates. The tables stated over here also containing the parameter estimates for the industry dummies.

Table A.10: Estimated model parameters for all three duration models, including industry dummies. The standard errors are in parenthesis.

	Basic model		**CRE**		**DRE**	
Age	0.162***	(0.055)	0.375***	(0.106)	0.436***	(0.111)
Age squared	-0.002***	(0.001)	-0.005***	(0.001)	-0.006***	(0.001)
Male	0.251*	(0.149)	0.551*	(0.297)	0.447	(0.282)
Industry						
Manufacturing (base)						
Construction	-1.196***	(0.375)	-2.407***	(0.793)	-2.151***	(0.738)
Wholesale	-0.125	(0.404)	-0.310	(0.849)	-0.308	(0.724)
Retail	-0.596	(0.387)	-0.908	(0.808)	-0.736	(0.741)
Hotels and restaurants	0.863*	(0.483)	2.096*	(1.108)	3.379***	(1.006)
Motor vehicles	-0.115	(0.551)	-0.342	(1.190)	0.735	(1.012)
Transport	0.106	(0.455)	0.998	(0.980)	2.057**	(0.884)
Business and financial services	-0.231	(0.346)	-0.345	(0.727)	-0.178	(0.669)
Other services	-0.468	(0.361)	-0.815	(0.747)	-0.673	(0.704)
Start-up motives						
Intrinsic (base)						
Push	-0.293	(0.180)	-0.676*	(0.358)	-0.522	(0.361)
Opportunist	0.063	(0.195)	0.353	(0.405)	0.924*	(0.490)
Work-life	-0.457**	(0.219)	-0.964**	(0.423)	-1.048***	(0.382)
Other	-0.094	(0.247)	-0.093	(0.511)	-0.163	(0.456)
Objectives						
Improve own expertise	-0.179	(0.149)	-0.240	(0.304)	0.172	(0.298)
Improve product quality	-0.136	(0.130)	-0.292	(0.265)	-0.378	(0.258)
Maximize profits	-0.183	(0.131)	-0.380	(0.270)	-0.286	(0.278)
Maximize revenue	0.569***	(0.133)	1.222***	(0.288)	1.174***	(0.295)
Competencies						
Educational level						
-Low (base)						
-Middle	0.069	(0.165)	0.061	(0.327)	-0.010	(0.307)
-High	-0.053	(0.180)	-0.201	(0.360)	-0.088	(0.377)
Industry experience	0.505***	(0.142)	1.130***	(0.300)	1.621***	(0.339)
Entrepreneurial experience	0.484**	(0.196)	1.193***	(0.428)	1.459***	(0.452)
Entrepreneurial self-efficacy	0.270***	(0.092)	0.648***	(0.194)	1.051***	(0.240)
Risk attitude	0.243***	(0.093)	0.420**	(0.190)	0.547***	(0.195)
Social capital	0.192	(0.154)	0.517	(0.321)	0.669**	(0.317)
Firm-specific factors						
Fulltime	0.882***	(0.139)	1.957***	(0.314)	2.143***	(0.297)
Innovativeness	0.103	(0.142)	0.333	(0.296)	0.704**	(0.321)
Time-varying covariates						
t	-0.020	(0.149)	0.816***	(0.250)	0.937***	(0.254)
t^2	0.002	(0.015)	-0.048**	(0.022)	-0.055**	(0.023)
GDP growth	-0.063	(0.065)	-0.137*	(0.083)	-0.143*	(0.085)
GDP growth (one-year lag)	0.195***	(0.068)	0.304***	(0.084)	0.324***	(0.088)
Intercepts						
Constant	-8.278***	(1.285)	-18.757***	(2.946)		
σ_α			2.900[a]	(0.346)		
ξ_1					-28.583***	(4.036)
ξ_2					-19.297***	(3.259)
ξ_3					-23.447***	(3.627)
Estimation info						
log-likelihood	-1110.22		-1098.80		-1093.72	
M (entrepreneurs)	1402		1402		1402	
N (effective sample)	6239		6239		6239	

* significant at 10% level
** significant at 5% level
*** significant at 1% level
[a] The variance σ_α^2 of the random intercept significantly differs from zero according to the LR test. The LR test-statistic equals 22.84 and is signifcant at 1% level.

Table A.11: Estimated model parameters for the count models, including industry dummies. The standard errors are in parenthesis.

	Poisson	**NB**	**LC Poisson**
Age	0.291*** (0.066)	0.234** (0.107)	0.246*** (0.015)
(Age squared)/100	-0.398*** (0.087)	-0.334** (0.141)	-0.333*** (0.005)
Male	-0.081 (0.134)	-0.058 (0.267)	-0.338* (0.189)
Industry			
Manufacturing (base)			
Construction	1.088** (0.451)	1.327* (0.746)	0.920 (0.599)
Wholesale	1.236*** (0.468)	1.570** (0.785)	1.249* (0.640)
Retail	-0.078 (0.499)	0.540 (0.779)	0.753 (0.654)
Hotels and restaurants	1.903*** (0.464)	2.182*** (0.820)	1.871*** (0.589)
Motor vehicles	0.571 (0.620)	0.721 (1.000)	0.117 (0.773)
Transport	0.901* (0.519)	1.341 (0.862)	1.606** (0.674)
Business and financial services	0.797* (0.438)	1.246* (0.713)	1.189** (0.568)
Other services	1.310*** (0.434)	1.595** (0.699)	1.418*** (0.542)
Start-up motives			
Intrinsic (base)			
Push	0.910*** (0.148)	0.795*** (0.303)	1.125*** (0.208)
Opportunist	-0.246 (0.174)	-0.253 (0.331)	-0.025 (0.227)
Work-life	0.609*** (0.178)	0.369 (0.379)	-0.111 (0.237)
Other	-0.507** (0.249)	-0.343 (0.436)	-0.072 (0.321)
Objectives			
Improve own expertise	-0.070 (0.127)	-0.068 (0.247)	0.039 (0.162)
Improve product quality	0.122 (0.116)	0.166 (0.229)	-0.329* (0.170)
Maximize profits	0.155 (0.111)	0.251 (0.218)	-0.236 (0.158)
Maximize revenue	0.068 (0.124)	0.115 (0.234)	0.076 (0.182)
Competencies			
Educational level			
-Low (base)			
-Middle	0.821*** (0.172)	0.678** (0.283)	0.675*** (0.222)
-High	0.646*** (0.190)	0.361 (0.313)	0.642** (0.250)
Industry experience	-0.001 (0.135)	-0.202 (0.260)	-0.257 (0.198)
Entrepreneurial experience	0.581*** (0.141)	0.613** (0.309)	0.509*** (0.177)
Entrepreneurial self-efficacy	0.327*** (0.087)	0.245 (0.176)	0.351*** (0.130)
Risk attitude	0.139 (0.092)	0.040 (0.183)	0.143 (0.126)
Social capital	-0.171 (0.137)	0.036 (0.264)	0.012 (0.189)
Firm-specific factors			
Fulltime	0.239* (0.136)	0.281 (0.244)	0.446** (0.182)
Innovativeness	0.513*** (0.126)	0.558** (0.252)	0.703*** (0.176)
Time-varying covariates			
t	-0.578*** (0.137)	-0.829*** (0.266)	-0.439* (0.226)
t^2	0.068*** (0.014)	0.088*** (0.028)	0.055** (0.023)
GDP growth	-0.134** (0.064)	-0.233* (0.124)	-0.139 (0.098)
GDP growth (one-year lag)	0.278*** (0.077)	0.160 (0.133)	0.344*** (0.130)
Intercepts			
Constant	-8.784*** (1.506)	-5.915** (2.473)	
ω		1.674^{a} (0.259)	
ξ_1			-6.946*** (1.446)
ξ_2			-9.602*** (1.453)
Estimation info			
log-likelihood	-583.26	-434.79	-429.27
M (effective sample)	320	320	320

* significant at 10% level
** significant at 5% level
*** significant at 1% level
[a] The variance ω of the unobserved heterogeneity term significantly differs from zero according to the LR test. The LR test-statistic equals 296.94 is significant at 1% level.

The results of EIM's Research Programme on SMEs and Entrepreneurship are published in the following series: Research Reports and Publieksrapportages. The most recent publications of both series may be downloaded at: www.entrepreneurship-sme.eu.

Recent Research Reports and Scales Papers

H201017	25-5-2010	Entrepreneurship and role models
H201016	29-4-2010	The Nature and Prevalence of Inter-Organizational Project Ventures: Evidence from a large scale Field Study in the Netherlands 2006-2009
H201015	27-4-2010	New Firm Performance: Does the Age of Founders Affect Employment Creation?
H201014	1-4-2010	Van defensief MKB-beleid naar offensief ondernemerschapsbeleid
H201013	16-3-2010	Human capital and start-up succes of nascent entrepreneurs
H201012	16-3-2010	New business creation in the Netherlands
H201011	16-3-2010	Factors influencing the entrepreneurial engagement of opportunity and necessity entrepreneurs
H201010	16-3-2010	The more business owners the merrier? The role of tertiary education
H201009	2-3-2010	Open, distributed and user-centered: Towards a paradigm shift in innovation policy
H201008	1-2-2010	Geographical distance of innovation collaborations
H201006	21-1-2010	Family ownership, innovation and other context variables as determinants of sustainable entrepreneurship in SMEs: An empirical research study
H201005	12-1-2010	Intrapreneurship – An international study
H201004	12-1-2010	Schumpeter versus Kirzner: An empirical investigation of opportunity types
H201003	7-1-2010	Institutions and entrepreneurship: The role of the rule of law
H201002	5-1-2010	Patterns of innovation networking in Dutch small firms
H201001	4-1-2010	Administratieve lasten en ondernemerschap
H200911	29-1-2010	Ambitious entrepreneurship, high-growth firms and macroeconomic growth
H200910	4-2-2010	Entrepreneurial exit and entrepreneurial engagement
H200909	10-3-2009	Entrepreneurship Education Monitor (EEM)
H200908	3-3-2009	Internationale samenwerking door het Nederlandse MKB
H200907	2-3-2009	The Dynamics of Entry and Exit
H200906	2-3-2009	Bedrijfsgrootteverdelingen in Nederland
H200905	2-3-2009	Start-ups as drivers of incumbent firm mobility: An analysis at the region-sector level for the Netherlands
H200904	16-2-2009	Een reconstructie van het beleidsprogramma Ondernemerschap en Starters 1982-2003: een eclectische analyse
H200903	16-2-2009	Determinants and dimensions of firm growth
H200902	2-2-2009	The decision to innovate: Antecedents of opportunity exploitation in high tech small firms

H200901	7-1-2009	The Relationship between Successor, Planning Characteristics, and the Transfer Process on Post-Transfer Profitability in SMEs
H200825	19-12-2008	Isomorfie en het beloningspakket van werknemers in het MKB
H200824	16-12-2008	The relation between entrepreneurship and economic development: is it U-shaped?
H200823	11-12-2008	International Entrepreneurship: An Introduction, Framework and Research Agenda
H200822	11-12-2008	The two-way relationship between entrepreneurship and economic performance
H200821	5-12-2008	Spin-outs
H200820	27-11-2008	Innovative Work Behavior: Measurement and Validation
H200819	17-11-2008	Open innovation in SMEs: Trends, motives and management challenges
H200818	10-11-2008	High-Growth SMEs Evidence from the Netherlands
H200817	3-11-2008	Internationalization of European SMEs towards Emerging Markets
H200816	27-10-2008	Measuring business dynamics among incumbent firms in The Netherlands
H200815	20-10-2008	Vergrijzing van het arbeidsaanbod
H200814	16-10-2008	User Innovation in SMEs: Incidence and Transfer to Producers
H200813	30-9-2008	How Does Entrepreneurial Activity Affect the Supply of Business Angels?
H200812	16-9-2008	Science and technology-based regional entrepreneurship in the Netherlands: building support structures for business creation and growth entrepreneurship
H200811	8-9-2008	What Determines the Growth Ambition of Dutch Early-Stage Entrepreneurs?
H200810	6-8-2008	The Entrepreneurial Advantage of World Cities; Evidence from Global Entrepreneurship Monitor Data
H200809	25-7-2008	The Entrepreneurial Adjustment Process in Disequilibrium: Entry and Exit when Markets Under and Over Shoot
H200808	2-7-2008	Entrepreneurial Career Capital, Innovation and New Venture Export Orientation
H200807	24-6-2008	Twee decennia ondernemerschapsbeleid in beeld: een jong beleidsprogramma in sociaaleconomische context geplaatst
H200806	18-6-2008	Overcoming Resource-Constraints through Internationalization? An Empirical Analysis of European SMEs
H200805	9-6-2008	Whither a flat landscape? Regional differences in Entrepreneurship in the Netherlands
H200804	19-2-2008	Samenwerken op afstand
H200803	1-1-2008	Explaining Preferences and Actual Involvement in Self-Employment: New Insights into the Role of Gender
H200802	5-6-2008	Intrapreneurship; Conceptualizing entrepreneurial employee behaviour