Entrepreneurial decisions in a social context

Citation for published version (APA):

Kok, J. B. (2006). *Entrepreneurial decisions in a social context*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20070124jk>

Document status and date:

Published: 01/01/2006

DOI:

[10.26481/dis.20070124jk](https://doi.org/10.26481/dis.20070124jk)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

ENTREPRENEURIAL DECISIONS IN A SOCIAL CONTEXT

© 2006 J.B. Kok, Maastricht

All rights reserved. No part of this publication may be reproduced without the written permission of the author.

ISBN-10: 9052785996

ISBN-13: 9789052785998

Printed by Datawyse, Maastricht (NL)

ENTREPRENEURIAL DECISIONS IN A SOCIAL CONTEXT

PROEFSCHRIFT

ter verkrijging van de graad doctor

aan de Universiteit Maastricht

op gezag van de Rector Magnificus Prof. Mr. G.P.M.F. Mols

volgens het besluit van het College van Decanen

in het openbaar te verdedigen op

woensdag 24 januari 2007 om 16.00 uur door

Jasper Baldwin Kok

Promotor:

Prof. dr. G.A. Pfann

Beoordelingscommissie:

Prof. dr. M.A. Carree (voorzitter)

Dr. W.A. Letterie

Prof. dr. A. van Witteloostuijn (Universiteit van Groningen)

“He who knows does not speak;
he who speaks does not know.”

LAO TZU – *Tao Te Ching*

Acknowledgements

No man is an island. It is a quote that I remember from reading one of my first academic articles for my dissertation. I could have not imagined that this quote gained so much in meaning and depth over the years. I often consider myself a loner, appreciating these precious moments of silence, where the world stands still, time comes to a stop, and our true nature reveals itself. However, my life could not be so interesting, fascinating, gained so much depth, insight, wisdom, and love if it wasn't because of the other. The other with whom we create this marvel called we. And it is to this other that I would like to dedicate my dissertation, the other which I so often forgot to thank for all that they are and have done for me. Let these few words provide you with a hint of the enormous gratitude I feel towards you all. Without you this dissertation would have never unfolded.

At the first place, I would like to thank my promotor Gerard Pfann. Gerard, you are an excellent promotor. You have always stimulated my ideas adding your insight and your superior technical skills. You gave me the freedom to explore my own ideas. I remember you saying in the first year that I should not only look into dissertation related papers, but should also befriend myself with the broader literature of philosophy and science. Believe me, I did. I rediscovered my passion for reading. And I will never forget all the worlds that have opened up to me ever since. But first and foremost, you have been an amazing warm person. You always showed your interest in my personal life, especially when personal life became turbulent again. Thank you for all of that.

In the second place, I would like to thank my supervisor Boris Blumberg. Boris, you are such a nice person. It was you who got me interested in research. Thank you for seeing that potential in me. I believe that somewhere down the road we drifted a bit apart. Maybe, I had to go my own way, but I have always enjoyed our conversations.

Then, I would like to thank all these persons around me that accompanied me throughout my life. I know almost none of you have only the slightest idea what I have been doing these last four years. I am just not the guy to continue work in his social life. But I would like to thank you for all our stimulating discussions about life and our great moments when we experienced life. In particular I would like to express my heartfelt gratitude to the following persons.

My parents. Dear mom and dad, thanks for being such wonderful parents. You have always stimulated me to follow my way. Now, I know why I have chosen you as my parents.

My brother. Peter, we have always been more than brothers. I consider you as my friend. I remember all our stimulating discussions about art, science, and life. You have always remembered the Dionysian side in me. Although your life has been sometimes difficult and hard, I will always know your true face. I am proud to be your guide.

Lisette. Thanks Lisette, for being my soul mate. Thank you even more for opening me up to dimensions I was never even aware of. I believe you have had the right key.

Philip. Thanks for being an excellent friend. I could share with you both the dark and the light side of life.

Furthermore, I would like to thank all my friends and acquaintances from Luttelgeest and Xander in particular. You guys laid the basis for the rest of my life. I would also like to thank all my friends from Uros. Although I could not always catch up with you, I have always enjoyed the social side of running. Furthermore, I would like to thank all the wonderful people I have got to know over the years in Amsterdam and Sanne in Utrecht.

Also I would like to thank my colleagues from the department of Organization and Strategy. I always felt at ease with all of you. I would like to express my deep gratitude to the secretaries from our department. Dear Adela, Eva, en Kitty, thanks for all the work you have done for me. Furthermore, I would like to thank my former roommates Michael, Ben and Stefanie for being great company and my recent roommate Nicolle for withstanding my horrible humor.

And let me end my gratitude to the one that gazes through my eyes, that ever present witness, which is the true understanding that no man can be an island: *tat tvam asi*.

Table of contents

Acknowledgements

1	Introduction	1
1.1	Introduction	2
1.2	Contributions and relevance of the study.....	3
1.3	Research questions.....	5
1.4	Outline of the study.....	6
2	Literature review	8
2.1	Introduction.....	9
2.2	Entry and exit.....	11
2.2.1	Theoretical models of entry and exit.....	12
2.2.2	Entry.....	12
2.2.3	Exit.....	14
2.3	Empirical studies on self-employment selection.....	15
2.3.1	The statistical relationship between inflow, stock and duration.....	16
2.3.2	The bias in the determinants of the probability of becoming self-employed.....	18
2.3.3	Empirical studies on self-employment entry.....	19
2.3.4	Empirical studies on self-employment exit.....	22
2.4	Social capital.....	23
2.4.1	Theory: The social capital process.....	24
2.4.2	Social capital formation.....	27
2.4.3	Social capital access.....	29
2.4.3.1	Closed networks and social capital.....	29
2.4.3.2	Open networks and social capital.....	30
2.4.4	Mobilization of social capital.....	31
2.4.5	Social capital outcomes: Integrating self-employment and social capital.....	32
2.5	Conclusion.....	35
2.6	Appendix.....	37

3	Time aggregation bias in estimating the probability of entrepreneurship	39
3.1	Introduction.....	40
3.2	The process that generates the data.....	42
3.2.1	The information in the data	42
3.2.2	Modeling the transition to entrepreneurship as a Poisson process.....	43
3.3	Grouped duration models.....	45
3.3.1	Duration models.....	45
3.3.2	Different specifications for the failure function.....	46
3.3.3	Duration dependence.....	47
3.3.4	Unobserved heterogeneity.....	49
3.3.5	The proportionality assumption.....	49
3.4	An example of a study using dynamic data and a static estimator.....	50
3.5	Data analysis with dynamic techniques.....	51
3.6	Conclusion.....	62
4	The dynamics of business succession and de novo self-employment: A multiple risks approach	63
4.1	Introduction.....	64
4.2	Theory development.....	67
4.2.1	A rational choice model of the self-employment decision.....	67
4.3	Hypotheses development.....	71
4.3.1	Entrepreneurial capabilities and human capital.....	71
4.3.2	Social capital and family background.....	74
4.3.3	The dynamics of entering the self-employment selection over the lifecycle.....	77
4.4	Methodology.....	78
4.4.1	Data.....	78
4.4.2	Used methods and operationalization.....	79
4.4.3	Model.....	81
4.5	Results.....	83
4.5.1	The dynamics of real founders and family self-employment.....	87
4.6	Conclusions & Perspectives.....	88

5	The Limburg dataset	90
5.1	The survey.....	91
5.1.1	History.....	91
5.1.2	Aim.....	91
5.1.2.1	Research population.....	91
5.1.2.2	Time span.....	92
5.1.2.3	Choice for South Limburg.....	92
5.1.3	Development and execution of the research.....	92
5.1.3.1	Execution of research.....	92
5.1.3.2	Execution process.....	94
5.1.4	Conduct.....	95
5.1.4.1	Response and non-response.....	95
5.1.4.2	Non-response in the Limburg survey.....	96
5.2	Survey follow up: Matching Limburg data with Chamber of Commerce data	97
5.3	Some (descriptive) results from the data.....	99
5.3.1	The self-employment decision.....	99
5.3.2	Starting entrepreneurs versus non-starting entrepreneurs.....	99
5.3.2.1	Demographics.....	99
5.3.2.2	Social background.....	100
5.3.2.3	Human capital.....	101
5.3.2.4	Social capital.....	102
5.3.2.4.1	Social network size.....	103
5.3.2.4.2	Social network composition.....	104
5.3.2.5	Starting up: Motives.....	107
5.3.2.6	Satisfaction.....	107
5.3.3	Entry and exit timing.....	109
5.3.3.1	Non-parametric entry hazard.....	109
5.3.3.2	Non-parametric exit hazard.....	110
5.4	Conclusion.....	111
6	Non-response bias and business discontinuance rates	112
6.1	Introduction.....	113
6.2	The data.....	114
6.2.1	The Limburg survey.....	114
6.2.2	A taxonomy of non-response.....	114
6.3	The research problem and some initial statistics.....	118

6.3.1 Sample analogs discontinuance rate.....	119
6.4 A simple model of survey non-response and discontinuing the business idea.	119
6.4.1 Determination of business idea status.....	120
6.4.2 Determination of response status.....	121
6.4.3 Joint determination of continuance and response status.....	121
6.5 A sequential decision model.....	124
6.6 Conclusion.....	128
7 Reluctance to exit self-employment when a fraction of the set up cost is sunk	130
7.1 Introduction.....	131
7.2 Why do entrepreneurs enter on a different scale?.....	133
7.3 The entrepreneur's exit problem under heterogeneous irreversible set up costs.....	136
7.4 The econometric model and the data.....	139
7.4.1 Testing the predictions: Sunk set up costs, uncertainty, and productivity.....	143
7.5 Estimation results.....	145
7.6 Conclusion.....	148
7.7 Appendix.....	149
8 A discrete choice model for productive social relations	152
8.1 Introduction.....	153
8.2 Investing in productive social relations.....	155
8.2.1 Demand curves.....	156
8.2.2 Supply curves.....	157
8.2.3 Equilibrium.....	157
8.2.4 Productive social relations.....	158
8.3 Data.....	159
8.4 Estimation results.....	163
8.5 Conclusion.....	168
9 Summary & conclusion	169
9.1 Summary.....	170
9.2 Suggestions for future research.....	174

Bibliography	176
Nederlandstalige samenvatting	184
Curriculum Vitae	190

List of Tables

2.1	Predicted sign of β_{IF} given the sign of β_{SE} and γ	18
2.2	Predicted bias given the signs of β_{IF} and γ	19
2.3	Overview of empirical studies investigating the self-employment decisions...	20
2.4	Empirical results social capital and self-employment.....	34
3.1	Estimation results.....	52
3.2	Prediction accuracy.....	53
3.3	Test of proportional hazards assumption (based on Cox PH model).....	59
3.4	AIC values for several spline models.....	60
3.5	Estimation results – Parametric spline model.....	61
4.1	Descriptive statistics.....	80
4.2	Selection of parametric distributions – AIC values.....	83
4.3	Estimation results.....	85
5.1	Displacement starters and non-starters.....	97
5.2	Deregistration reasons.....	98
5.3	Cross-tabulation marital status and self-employment status.....	99
5.4	Distribution parental education.....	101
5.5	Distribution respondents over educational levels.....	101
5.6	Cross-tabulation leadership experience and number of employees.....	102
5.7	Start-up motives.....	107
6.1	Non-response categories.....	115
6.2	Distribution of business idea continuance.....	117
6.3	Business idea and response status: Cell counts in the Limburg telephone survey.....	117
6.4	Discontinuance rate at different phases of survey.....	120
6.5	Simple model parameter estimates.....	123
6.6	Estimation results sequential decision model.....	126
6.7	Partial effects.....	127

7.1	Explaining initial commitment in set up costs.....	137
7.2	Deregistration reasons for exit.....	142
7.3	Descriptive statistics of independent variables.....	144
7.4	Estimates of log-logistic duration model of waiting time out of self-employment.....	146
7.5	Summary statistics.....	149
7.6	Correlation matrix.....	150
8.1	Population and sampling details.....	160
8.2a	Descriptive statistics: Mean and standard deviations by industry.....	162
8.2b	Descriptive statistics: Mean and standard deviations by number of social relations.....	163
8.3	Order choice probabilities for number of social relations.....	164
8.4	Industry choice probabilities, equilibrium demand, prices, and profits.....	167

List of Figures

2.1	The social capital process.....	25
3.1	Distribution of entry years into self-employment.....	42
3.2	Derivatives of the log hazard function.....	55
3.3	The non-parametric baseline discrete hazard.....	56
3.4	Proportionate change in discrete hazard induced by the self-employed father: various specifications for F.....	57
3.5	Proportionate change in discrete hazard induced by the self-employed father: cloglog specification for F.....	59
3.6	The baseline hazard function.....	60
3.7	The non-proportional effect of self-employed father over time.....	62
4.1	Conceptual model of the self-employment process.....	68
4.2	Capital accumulation, thresholds, and waiting time.....	70
4.3	Human capital mechanism.....	72
4.4	Baseline hazards.....	87
5.1	Execution process.....	93
5.2	Interview dates.....	95
5.3	Age distribution.....	100
5.4	Satisfaction support social network.....	103
5.5	Social network size.....	103
5.6	Distribution of proportion of self-employed over starters and non-starters.....	104
5.7	Distribution of proportion of family members over starters and non-starters...	105
5.8	Distribution of proportion of strong ties over starters and non-starters.....	106
5.9	Participation in associational life.....	106
5.10	Distribution of satisfaction.....	108
5.11	Income distribution.....	109
5.12	Non-parametric entry hazard.....	110
5.13	Non-parametric exit hazard.....	111
6.1	Flow diagram of non-response taxonomy.....	116
6.2	Sequential decision model.....	124

7.1	Distribution of initial capital investments in firm (log-form).....	134
7.2	Non-parametric Kaplan-Meier exit hazard.....	140
7.3	Baseline exit hazard.....	141
8.1	Frequency distribution of productive social relations of entrepreneurs.....	154
8.2	Within industry equilibrium levels of investments in productive social relations.....	155
8.3	Predicted profit level.....	166

Chapter 1

Introduction

1 Introduction

“A fruitful analysis of human action requires us to avoid the atomization implicit in the theoretical extremes of under- and oversocialized conceptions. Actors do not behave or decide as atoms outside a social context, nor do they adhere slavishly to a script written for them by the particular intersection of social categories that they happen to occupy. Their attempts at purposive action are instead embedded in concrete, ongoing systems of social relations.” (Granovetter 1985: 487)

1.1 Introduction

The entrepreneur takes a remarkable place in society. For the public, the entrepreneur is the hardworking, creative, independent successful businessman¹. He is Bill Gates or Donald Trump. In its ultimate form, he is the paperboy, who started with nothing and who worked hard, with genius, to build his empire, sometimes relentless, but always forward looking, seizing opportunities, and transforming challenges into victories. For the government, the entrepreneur is the supercharger of the economy. He is creating jobs and boosting economic growth. But for the researcher, the entrepreneur presents mainly a problem. How do we define him? What are his functions? Is he the Sayian manager of labor and capital, or the Marshallian jack-of-all-trade with general and specific abilities to deal with the variety of task of entrepreneurship, or is he the Schumpeterian innovator, who disrupts economic equilibrium, or rather the Knightean bearer of uncertainty? Although researchers differ in their conceptualizations and definitions of the entrepreneur, they acknowledge the entrepreneur’s important function of changing the economic and social context by innovating new products, filling niches unattractive for multinational enterprises, and increasing competition (see e.g. Parker 2004).

As the opening quotes suggests, human action, like entrepreneurial action, operates never without a social context. Entrepreneurs do not behave or decide as atoms outside a social world. For the entrepreneur not only shapes the economic and social context with his products and actions, he is in turn dependent on and shaped by his social and economic context. For it is not without reason that an entrepreneur often stems from an entrepreneurial family, which does not only provide the entrepreneur with a pair of genes fit for the variety of entrepreneurial tasks, but also provides him with information, advice and support, and maybe even with the possibility to take over a company. And in every dialogue with an entrepreneur, he will ascertain you the value of his social network, in favoring him orders, assuring him product and service quality, and providing him timely and reliable information about suppliers and clients. During the last two decades, researchers have acknowledged the value of social structure and social networks and coined this phenomenon social capital, in close analogy to physical, financial and human capital.

But the social context can be more subtle. The entrepreneur’s history, his past choices, all influences his recent course through the entrepreneurial context. His investments in education, his savings on past earnings, his marriage, and even parental heritage,

¹ In this dissertation I will often speak of the entrepreneur as a man. Not because of gender preferences, but merely of the fact that still more men than women are entrepreneur today.

can help the entrepreneur to fulfill his dreams to start a business. But at the same time, the accumulation of job-specific capital, the feeding and nurturing of children can make an entrepreneur anxious or reluctant to become who he really is. Not only is his social context in continuing flux, also the effects the social context bear on the individual can change over time. For example, parental social and financial support are mainly important at the beginning of an individual's career, when he tries to get a foot on the ground, but lacks the necessary resources and experience to go on his own. And for the aging entrepreneur more time comes available when the demands of children fade away as they grow up and leave parental home.

The entrepreneur is so strongly embedded in his social context, both horizontally through space by the social and economic structure and vertically through time by his past actions and choices and future aspirations, that entrepreneurial decisions can not be researched from a static atomized perspective, but should be studied as a dynamic phenomenon mutually reinforcing social and economic context. Research questions should follow this empirical reality. Important research questions in self-employment and entrepreneurship literature need revision. The most common research questions "who becomes an entrepreneur and who exits entrepreneurship?" must be extended with research questions like "when does an individual becomes an entrepreneur and when does an individual decide to exit entrepreneurship?". Furthermore, researchers should question how these processes behave in a social context.

1.2 Contributions and relevance of the study

This dissertation explores the entry, exit, and capital accumulation decisions of entrepreneurs in a social context. In this dissertation, we are interested in the individual as a decision maker. Why do individuals decide to become an entrepreneur and why do individuals decide to exit entrepreneurship? Why do entrepreneurs differ in their timing of these decisions? And how can an individual's social environment affect these choices? Knowledge on the reasons why individuals enter self-employment is important for policy makers that want to promote the level of self-employment and unleash possible financial, human, and social capital constraints. However, the empirical evidence of the determinants of self-employment entry provides mixed and contradictory results. Part of the confusion is attributable to the different sampling schemes used to study entry. Flow samples provide genuine information on the determinants of entrepreneurship. Stock samples, however, provide only information about the determinants of entrepreneurship under very restrictive assumptions. We believe that much of the controversy can be overcome by understanding the process that generates the self-employment entry data. Therefore, we derive a duration model of self-employment entry that is consistent with the stochastic process that generates the data. By this, we contribute to the dynamic analysis of self-employment entry decisions, as envisaged and modeled by Carroll and Mosakowski (1987). To our knowledge, we are the first to explicitly model the timing of the entry decision.

There is another reason for the mixed empirical results. Entrepreneurs grow up in a social context. Growing up in an unproductive and backward social environment can constrain an individual's opportunity to become self-employed because he has no or

limited access to financial or human capital. On the other hand, growing up in a social environment can also be the source of the opportunity to become self-employed. Namely, a considerable part of self-employment is family self-employment (Carroll and Mosakowski 1987). People who take over a family business can be different from those who found an entirely new business. While real founders enter self-employment with the burden of financing the sunk nature of start up expenses, continuation of a family business is like entering self-employment with a valuable set of assets already existent. Therefore, we build a theory of the dynamics of business succession and real founding. This model explains how human, financial, and social capital accumulation decisions influence the timing of the entry decision for these alternative types of self-employment. Differentiating real founding and family business succession can provide insights how the different requirements of these types of self-employment influence the speed whereby individuals become self-employed. This is to our knowledge the first theoretical model that can differentiate between family self-employment and de novo self-employment. Thereby, we finally materialize the suggestions for future research by Carroll and Mosakowski (1987), who only addressed this distinction empirically.

We believe that knowledge of the timing of the entry decision is important for the following reasons. If self-employment and its derivative entrepreneurship are important for economic growth and innovation (e.g. Cosh et al. 1999; Geroski and Pomroy 1990) then it is not only desirable that individuals become self-employed but also that they become self-employed early in their careers. Self-employment experience is namely an important predictor of future self-employment (Carroll and Mosakowski 1987; Evans and Leighton 1989) and of self-employment success related to profits (Bosma et al. 2004) and related to business survival (Holmes and Schmitz 1996; Quadrini 1999; Taylor 1999). But the self-employment process does not stop after entry. An individual entering self-employment demarcates the end of the entry event; however, an individual entering entrepreneurship also marks the start of the self-employment spell. We believe that these processes are related. Determinants that influence the timing of the entry decision can also influence the timing of the exit decision. We start from the postulate that the timing of the entry decision and the timing of the exit decision are not unrelated. We build a theoretical framework that acknowledges the option nature of the self-employment decision. Pivotal to this theory is that starting up a business involves sunk costs and that future earnings from self-employment are uncertain. We follow recent models of investments under uncertainty that show how the presence of sunk costs introduces an option value of waiting for new information and delaying irreversible decisions (Dixit and Pindyck 1994). These models combine the question of whether to invest (the “q” value) with when to invest (the option value). The option value of setting up a self-employed business is largely ignored in the theoretical and empirical self-employment models. We follow recent work by Pfann (2006) in identifying idiosyncratic option values.

Finally, we investigate why and how entrepreneurs invest in social relations. We believe that knowledge of these social capital accumulation decisions is important to understand differences in the composition of social networks among entrepreneurs. Inspired by Gleaser, Laibson, and Sacerdote’s (2002) analysis of the formation of social capital in the form of club membership, we apply Becker’s (1964) human capital theory to investments in social capital in the form of productive social relations. Social relations can add to profitability by economizing on transaction costs, minimizing search costs,

enforcing contracts, stimulating trade, and providing timely access to valuable and reliable information. Entrepreneurs who face higher returns on investments in social relations may therefore be more successful entrepreneurs. We model the accomplishment of social relations as a rational choice. In our model, we analyze the choice of the number of productive relations in a profit maximizing framework under uncertainty. The main objective of this model is to show that profitability of productive social relations depends on the productivity of social relations within an industry and idiosyncratic accomplishment costs.

1.3 Research questions

In order to better understand entry, exit, and capital accumulation behavior of entrepreneurs in a social context, the empirical part of this dissertation scrutinizes individual decisions related to self-employment. A first subset of research questions relates to the determinants of the timing of the entry decision.

- How can the probability of entry into entrepreneurship be studied in a dynamic context?
- What impact do individual characteristics have on the timing of the entry decision?
- How does the social environment influence the self-employment decision?
- How do the dynamics of the entry decision differ between different types of self-employment?

Empirical research relies on the participation of respondents. A common problem in surveys is non-response. Since part of our research is based on a newly created dataset, the so-called Limburg dataset; we ask ourselves the following research questions.

- What are the sources of non-response in the Limburg survey?
- Is there a structural relationship between the outcome of the self-employment entry decision and non-response?

A second set of research questions focuses on the relationship between entry and exit decisions of individuals.

- How can the theory of real options be applied to self-employment decisions?
- What are the causes and consequences of sunk set up costs for self-employment entry and exit decisions?
- How are the entry and exit decisions related to each other?
- How do individual exit thresholds relate to the timing of the exit decision?

Finally, we shift our attention from the entry and exit decisions of individuals to the accumulation decisions of entrepreneurs. More specific, we investigate the following research questions:

- Why do social networks differ among entrepreneurs?
- How can the accomplishment of entrepreneurial social networks be researched?

1.4 Outline of the study

The dissertation is organized as follows. To clarify the scientific context in which the research questions are embedded, chapter 2 surveys the literature on entrepreneurship, self-employment, and social capital. Because of the rather broad scope of this dissertation, we only discuss parts of the literature, which are either canonical or which we deem important for answering our research questions. This literature survey is therefore necessary inconclusive and is certainly not questioning the worth of other researchers who approach these topics from other perspectives not included in this survey.

The empirical part of the dissertation starts in chapter 3 with the econometric modeling of the dynamic self-employment entry decision. We explicitly model the timing of the entry decision into entrepreneurship. We investigate the predictive power and parameter accuracy of different dynamic duration models versus static single period classification models and determine the bias of time aggregation implicit in static models applied to multi-period data.

Chapter 4 extends chapter 3 by investigating the different routes into self-employment. We build a theory of the dynamics of business succession and de novo self-employment taking into account the capital accumulation decisions of individuals. Hypotheses are developed on how human and social capital influences the timing of the two self-employment options. A multiple risks model is estimated to test the hypotheses. In chapter 5, we introduce the Limburg dataset. We will use this dataset in chapters 6, 7, and 8. The Limburg dataset consists of detailed information on labor market history, social background, and social networks of 1223 individuals that seriously considered becoming self-employed around 1998 and 1999. In 2004, this dataset was matched with a dataset of the Chamber of Commerce that provide information about the status of the self-employment decision. In this chapter, we provide some initial results of both datasets related to our research questions.

Chapter 6 assesses empirically the sources of non-response at various stages of development of the Limburg dataset. A structural model is estimated to reveal some important population characteristics, like the probability of attrition (e.g. death), the probability of non-response, and the entrepreneurial ability of the prospective self-employed. Adverse events, like discontinuing a business idea, can make individuals reluctant to cooperate with the survey. A structural link between business continuance rate and non-response can lead to biased results in self-employment choice probabilities. We derive a sequential decision model to investigate the link between non-response and business continuance. The model's empirical contents are investigated with a multinomial logit analysis.

Chapter 7 investigates the individual exit decisions of entrepreneurs. We present a theoretical framework that shows the real options nature of self-employment decisions. We scrutinize on the causes and consequences of the fact that part of the set up costs are sunk. We show how idiosyncratic sunk costs, idiosyncratic productivity, and idiosyncratic uncertainty affect idiosyncratic profit thresholds and the length of the self-

employment spell when business conditions are uncertain. The predictions of the model are tested with a duration analysis.

In chapter 8, we design a model to explain the accomplishment of entrepreneurs' productive social networks in concentrated markets. We analyze the choice of the number of productive social relations in a profit maximizing framework. A discrete choice formulation is used to draw inferences about the unobserved profitability of entrepreneurs' social networks under uncertainty. The predictions are tested with an ordered logit model.

Finally, chapter 9 concludes. It reviews the empirical results of the dissertation and provides answers to the research questions. At the end of the summary, suggestions for future research are given.

Chapter 2

Literature review

2 Literature review

2.1 Introduction

Entrepreneurship is the act of starting up and exploiting a business idea. Entrepreneurship often involves the creation of new business ventures. Entrepreneurship is characterised by risk taking and risk creation, creativity and innovation, drive and ambition, and need for independence and self-actualization. Entrepreneurship is often equated with self-employment and small business ownership. However, not all self-employed are entrepreneurs (Carland et al. 1984; Webster 1977). Self-employment mainly refers to an income-generating mechanism that is different from paid labour. Furthermore, not all entrepreneurs are self-employed. Entrepreneurship can also take place within an organization. This is often referred as intrapreneurship. In this literature review, we will apply the broad definition of entrepreneurship as self-employment and use the terms interchangeably².

Historically, the definition, characteristics, and function of an entrepreneur is defined in a plethora of ways³. Cantillon (1755) coined the term entrepreneur. He argued that the distinguishing feature of an entrepreneur is risk bearing. The entrepreneur buys certain factors of production at certain prices in order to create a product and sell it at an uncertain price in the future. The entrepreneur bears the risk of the uncertain future price of the product. With Jean-Baptiste Say (1803), the entrepreneur obtained its modern definition of an entrepreneur as a manager of a business. In his definition, an entrepreneur is the coordinator of production and the distributor of wage and interest among the productive inputs of labor and capital. John Stuart Mill (1848) brought the term entrepreneur into general use among economists. Like Cantillon, he argued that the distinguishing characteristic of an entrepreneur is risk bearing. Marshall (1890) provided a neoclassical point of view of the entrepreneur. He presented the entrepreneur as the supplier of commodities. The entrepreneur exercises control and bears all responsibilities of the firm. In order to fulfill the variety of tasks of an entrepreneur, the entrepreneur should possess some general abilities and intelligence. Marshall pointed towards family background, education, and innate abilities as important determinants of these general abilities. Furthermore, in order to become successful, the entrepreneur should also possess some specific abilities, like forecasting, spotting and exploiting opportunities, and knowledge of the trade. With Schumpeter (1934), the entrepreneur transformed from a manager to an innovator. Whereas for Marshall, innovation was merely a by-product of the production process, for Schumpeter, the innovative creative ability of the entrepreneur becomes pivotal and is seen as the driving force of change in the economic system. He acknowledged that the ability to innovate required psychological motives and thereby opened

² Simon Parker, one of the authorities on entrepreneurship and self-employment, explains this issue as follows: "at the conceptual level, the term 'entrepreneur' and 'entrepreneurship' will be used; at the practical level, where issues of measurement, estimation and policy are involved, we will use the closest approximation to the manifestation of entrepreneurship that appears to be suitable. That will usually be 'self-employment', though occasionally the term 'small firm' will be more relevant." (Parker 2004: 5).

³ See e.g. Carland et al. (1984) or Van Praag (1999) for more elaborate historical perspectives on entrepreneurship and Brockhaus (1982) for some excellent definitions.

entrepreneurship for psychological inquiry. We owe the distinction between risk and uncertainty to the dissertation of Frank Knight (1921). He argued that the true economic function of the entrepreneur is the bearing of uncertainty. Uncertainty differs from risk, in that for risk one can calculate the ex-ante probabilities of an outcome of an event while with uncertainty the outcome of an event cannot be known. Therefore, risk can be insured. With Knight, good judgement becomes an important characteristic of the entrepreneur as well as his ability to deal with this uncertainty. More able entrepreneurs can therefore reap above average profits.

During the last thirty years there has been a shift from theoretical descriptive research to more empirically oriented research of entrepreneurship and self-employment. Within psychology and organizational research, researchers investigated the (psychological) traits and motives of the entrepreneur (see e.g. Brockhaus 1982; McClelland 1961). At the same time within economics and business, more structural empirical research started. This research evolved around two research questions: (1) Who becomes an entrepreneur? and (2) What determines the success of an entrepreneur? The difference between these two disciplines, psychology and economics, is that psychological research searched for individual traits and motives related to entrepreneurship, while economic research focused on the choice for entrepreneurship. The economic approach assumes maximizing behaviour of the entrepreneur. The choice for entrepreneurship is therefore simultaneously determined by the preference of the entrepreneur, its relative productivity in self-employment versus paid labour, and the availability of liquidity constraints. The empirical determinants of this choice will be surveyed in this literature review.

Not all entrepreneurs are successful. Besides a set of determinants of self-employment selection, the empirical research investigated a set of factors that influenced the probability that an entrepreneur would exit self-employment. One of the possible factors for successful entrepreneurship that has gained much popularity in formal research and popular literature is the social network of the entrepreneur. The research on social networks is embedded in the larger research program on social capital that has gained much popularity throughout the nineties. The term social capital can be traced back in sociology to Bourdieu (1986) and Loury (1977) in economics, but the concept was picked up and formalized by Coleman (Coleman 1988; 1990) and finally popularized by Putnam (2000) in his book *Bowling Alone*. According to Putnam, social capital is

"...features of social life - networks, norms, and trust - that enable participants to act together more effectively to pursue shared objectives... Social capital, in short, refers to social connections and the attendant norms and trust". (Putnam 1995: 665)

This broad definition is reflected in the broad research on this topic. It is instructive to separate collective social capital from individual social capital. From a collective level, social capital is often perceived as the norms and trust embedded in groups, cities, societies and nations, which lead to shared objectives, like community programs and safety. From an individual perspective, social capital is often seen as individual's purposeful investments in social relations with expected returns in the marketplace. The investment is often disbursed in increased access to information, gained trust to reduce uncertainty, and control of behavior and resources. In this literature review, we will investigate mainly how individual social capital, in the form of social relationships, can

influence self-employment entry and exit and how the research on social capital, in the form of social networks, can be incorporated in the research on entrepreneurship.

In the last few years, we have witnessed a surge in survey literature in the field of entrepreneurship (see e.g. Le 1999; Parker 2004; Van Praag 2005). Surveying these books and research papers, we have tried to identify the remaining gaps and lack of clarities in the literature of entrepreneurship.

This literature review is written with the following objectives in mind. First, we try to shed light on the contradictory results of the empirical literature on the determinants of entry into self-employment. Second, we provide a basis for theory development on individual social capital, specifically related to self-employment. Third, we summarize the empirical literature of the relationship between social capital and entrepreneurship.

This literature review is organized as follows. Section 2.2 deals with self-employment entry and exit. First, we will discuss and explain the canonical theoretical models on self-employment entry and exit in section 2.3. Then, we will investigate the empirical determinants of self-employment entry and exit. We will do so by critically examining the different sampling schemes used by researchers. Section 2.4 deals with social capital. First, we will discuss the main theoretical and empirical contributions to the concept of social capital. We present social capital as a process, following its formation, access, mobilization, and the consequences of social capital. Then, we will discuss the main empirical papers that have explicitly incorporated social capital as a determinant for explaining self-employment success. Section 2.5 summarizes the main arguments and provides recommendations for research.

2.2 Entry and exit

The literature offers several frameworks to study entry and exit behavior of firms, depending on the unit of analysis. At the industry level, entry and exit is often studied in an integrated equilibrium framework, where firm entry and exit is determined simultaneously depending on the industry distribution of entrepreneurial abilities (Jovanovic 1982), active reinvestments and learning (Pakes and Ericson 1998), the changes in industry prices, demand, or some other characteristics of the process for firm shocks (Hopenhayn 1992). At the individual level, entry and exit decisions are seldom researched simultaneously because entry approaches differ from exit approaches. Entry is mainly modeled as a choice in an occupational choice framework. The individual decides to work for a boss or pursues the more risky alternative of making earnings for himself. Whereas entry is a choice, exit is often perceived as a function of natural selection (Friedman 1953; Williamson 1991). Well-performing firms survive while poorly performing firms disappear. However, Gimeno, Folta, Cooper, and Woo (1997) argue that the survival of a firm is not exclusively a function of economic performance, but also depends on a business owner's own threshold of performance. Business owners may disentangle economic viable firms if income from self-employment is below the alternative wage paid in the labor market. In the next section, we are mainly interested in the behavior of individuals when they choose to enter or exit from markets.

2.2.1 Theoretical models of entry and exit

2.2.2 Entry

In order to explain who becomes an entrepreneur, and so who enters the market of entrepreneurs, researchers often use some form of an occupational choice model. The mechanism of the occupational choice model is straightforward. Self-employment refers to an income generating mechanism that is different from employment in the labor market. Since the occupational choice model is key to understanding self-employment selection, we will discuss an adapted version of Dunn and Holtz-Eakin (2000) entrepreneurial choice model.

The income of an individual working in the labor market consists of his wage and the returns on his assets. Let his wage be a function of its earning ability in a wage job w_i , assets A_i , and the net rate of return r . Then the individual's income as a wage earner is $w_i + rA_i$.

The income of the self-employed is more complex. The main difference between a wage earner and a self-employed person is that the self-employed person needs to invest capital in order to earn income. For a self-employed entrepreneur, the income is a function of his gross entrepreneurial earnings and the returns on remaining assets after capital is invested in the firm. Let the individual's gross earnings as an entrepreneur be a function of his entrepreneurial ability θ_i , the amount of capital invested k_i , and some random component ε , summarizing the good fortune of the entrepreneur. Then, the individual's income as an entrepreneur is $\theta_i f(k_i) \varepsilon + r(A_i - k_i)$.

If $k_i > A_i$, then the entrepreneur must borrow the remaining assets in the capital market. However, due to uninsurable risk and possible uncertain entrepreneurial abilities, capital markets are imperfect and therefore the amount of borrowing is constrained. The size of the constraint depends on the individual assets. The size of the constraint is defined by $k_i \leq l_k(A_i)$, with $l'_k(A_i) > 0$ and where $l_k(A_i)$ is the amount that can be borrowed.

For an entrepreneur, the optimal amount of capital k_i maximizes his entrepreneurial income subject to this liquidity constraint. The solution to this constrained maximization problem implies that $k_i^* = g(A_i, \theta_i^*)$. Dunn and Holtz-Eakin (2000) extend this analysis by allowing capital asset transfer between parents and offspring. In this case, own capital A_i is extended with family capital A_p . Allowing intergenerational capital transfer does not alter the results below. If the entrepreneur is liquidity constrained, then $\partial k_i^* / \partial A_i = l'_k(A_i) > 0$.

Until now, we have discussed the differences in the income generating mechanism. However, both incomes are uncertain. Generally among researchers, it is assumed that entrepreneurial income is more volatile because (1) the uncertain evolution of business conditions, and (2) the uncertainty of the entrepreneurs ability or business idea. Most researchers summarize this information in the formulation of expected income

streams. So that the standard entrepreneurial choice models predict that an individual will opt for entrepreneurship if the expected utility is higher in the self-employment option, that is, if,

$$(2.1) \quad E\{U[\theta_i f(k_i^*)\varepsilon + r(A_i - k_i); Z_i]\} > E\{U[w_i + rA_i]; Z_i\}$$

Where Z_i is a vector of personal characteristics (e.g. marital status, children, education, and social background).

Naturally, these models assume that individuals switch back and forth if the value for the options changes over time. According to this model, switching can be rather frequent and this is mainly due to the fact that almost all entrepreneurial choice models assume costless switching.

The literature on entrepreneurial choice offers several extensions and variations to this model. The main difference between researchers is the treatment, specification, and interpretation of entrepreneurial ability θ_i . In Lucas' (1978) general equilibrium model and in Evans and Jovanovic (1989), θ_i is assumed to be known and stable over time. The most able entrepreneurs become self-employed and the least able become wage earner. Moreover, different entrepreneurs manage firms with different sizes. The most able entrepreneurs start large firms, and the marginal entrepreneur starts the smallest firm in this model.

At the other extreme, Jovanovic (1982) assumes that entrepreneurial ability is initially unknown. Only the population distribution of entrepreneurial abilities is known. True entrepreneurial ability can only be learned after entry in the market in a Bayesian way by comparing expected productivity by true productivity revealed through company performance. Calvo and Wellisz (1980) and Van Praag and Cramer (2001) take a different approach and assume that entrepreneurial ability can be learned over time. Calvo and Wellisz (1980) argue that an individual's entrepreneurial ability is a function of his age and his ability to learn from the growth in common technological knowledge. Van Praag and Cramer (2001) argue that entrepreneurial ability is determined by individual characteristics like educational attainment, gender and social background. Finally, Dunn and Holtz-Eakin (2000) argue that entrepreneurial ability can be transmitted from the self-employed parents to their offspring.

Besides the different interpretations and specifications of entrepreneurial ability, there are several extensions of the standard entrepreneurial choice model. Holmes and Schmitz (1990) offer a model that incorporate the Schumpeterian idea of the entrepreneur as someone who recognizes and exploits new business opportunities. These opportunities arise continually in an economy which is in a permanent state of disequilibrium. In their model, individuals can choose to continue managing a firm or to set up a new firm, and transfer the control of the firm to a less able entrepreneur. Their analysis shows that the least able entrepreneurs will manage existing firms, and the most able entrepreneurs will focus on starting up new ventures. Whereas Holmes and Schmitz exploit the ideas of Schumpeter on entrepreneurship, Kihlstrom and Laffont (1979) build their model on the ideas of Knight on trading off risk and returns. Their model is built on the general perception of the entrepreneur as a risk taker. In their model, there is a population distri-

bution of risk aversion. The least risk averse individuals become entrepreneurs and manage the largest firms.

Almost as often as individuals decide to start up and manage a business, entrepreneurs decide to sell the firm or go bankrupt and exit from the market.

2.2.3 Exit

Our understanding why firms or individuals exit from the market is mainly driven by the seminal learning model of Jovanovic (1982). His elegant model explains firm exit from an industry through learning about initially unknown abilities. Entrepreneurs learn about their true entrepreneurial abilities only after they have entered the industry. Entrepreneurs learn about their entrepreneurial abilities by observing their firms' performance, which is based on stochastic draws from the market. Based on this constant flow of information, entrepreneurs adjust their belief in a Bayesian way, and adjust their strategies based on their updated beliefs. In Jovanovic' model, entrepreneurial ability is reflected in the cost function and is therefore a parameter for efficiency. Only the most able and most efficient entrepreneurs survive, while the less able or unlucky exit the market. Many of the predictions from Jovanovic model are supported by empirical evidence (see Parker 2004 Chapter 4 for an extensive treatment).

For understanding why individuals exit, Jovanovic' model is mainly used to determine the shape of the exit hazard. According to this model, exit rates are initially low, because entrepreneurs have to learn about their true ability and adjust only slowly. However, if the entrepreneur is faced with more evidence of low entrepreneurial ability, then the exit rate increases. Once the low ability entrepreneurs are driven from the market, the exit rate will decrease. There is ample evidence for this hump-shaped hazard rates for new firms (see e.g. Audretsch and Mahmood 1995; Mata and Portugal 1994).

Another implication of Jovanovic model is that all entrepreneurs should enter at the same scale. The fact that entrants should have the same scale is the result of the assumption that entrepreneurs have unknown prior entry entrepreneurial ability. Frank's (1988) model of intertemporal exit starts from the assumption that firms enter at a different scale. Furthermore, unlike Jovanovic, Frank assumes that (part) of the entry costs are sunk. Before investing these sunk costs, the firm must expect that these expenditures are justified. Frank argues that entrants who enter at a large scale have more optimistic beliefs about their entrepreneurial ability. And because these entrepreneurs have more optimistic beliefs it takes a longer time period to convince them of their true and revealed entrepreneurial ability and to change their beliefs and behavior. A disputable point in Frank's model is that he assumes that optimism and not entrepreneurial ability dictates firm entry size [in contrast to Lucas model, where entrepreneurial ability determines firm size]. The entrepreneur in Frank's model is still uncertain about its true entrepreneurial ability and learns it only ex post. Therefore, Frank's theory is based on the assumption that the entrepreneur knows that he can not know his entrepreneurial ability before entry, however, he still perceives his chances subjectively, and bases his entry strategy on this subjective belief.

Jovanovic (1982) and Frank (1988) entry and exit model provide two options concerning entry. Either firms enter at the same scale, or firms differ in scale because

larger firms are more optimistic. At the aggregate level these models seem to produce satisfying results, however, at a disaggregate level they seem less convincing. The main dissatisfaction stems from the fact that these exit models do not match with reasoning from occupational choice models that were introduced to explain entry.

In Jovanovic' model, the outside option is the entrepreneurial ability employed in a different activity. Jovanovic implicitly assumes that the value of this outside option is always significantly lower than the value of the chosen option. Given the fact that all entrepreneurs enter at the same scale, there are no marginal entrepreneurs. This is in sharp contrast with the occupational choice models. These models assume that entrepreneurs are marginal⁴. If entrepreneurs in Jovanovic' model would be marginal, then theorizing along Jovanovic' model should not lead to a hump-shaped exit hazard, but to a strictly decreasing hazard. The hazard would be strictly decreasing because directly after entry, firms with a negative draw can earn more in the outside option and should therefore exit.

Further dissatisfaction with aggregate entry and exit models arises from the assumption that entry and exit thresholds are set by the industry. As a consequence, firms exit if their performance is below industry standards. While this result may hold at aggregate levels, at disaggregated levels, this is quite unsatisfactory. If we refer back to equation (2.1), we see that thresholds are individual. Gimeno et al. (1997) develop a model of exit on this observation. They argue that firms differ in their thresholds of performance, and exit is determined by whether firm performance fall below these individual thresholds. The question then is: where do these thresholds come from? Gimeno et al. find that exit thresholds are mainly determined by the value of the outside option, in most cases for entrepreneurship this value is determined by income as a wage earner and sunk costs. In their analysis, these sunk costs arise from switching occupations. In order to induce exit, returns from entrepreneurship must fall below the value of the outside option minus the switching costs. Although their model does not account for uncertain entrepreneurial abilities, introducing uncertain entrepreneurial abilities will lead to a hump-shaped exit hazard. In that case it takes time to learn that entrepreneurial abilities fall below individual thresholds. Therefore, the key to understanding the hump-shaped exit hazard resides not only in uncertain entrepreneurial abilities, but also in sunk costs. The presence of sunk costs in self-employment decisions is not a far-fetched assumption. Asplund (2000) shows that 50-80 percent of the investments in metalworking machinery used in Swedish manufacturing industries are sunk. The consequences of sunk costs within the occupational choice models have not been put in a proper perspective. Next, we will discuss the empirical evidence on the empirical determinants of entry and exit.

2.3 Empirical Studies on Self-employment Selection

The decision to become self-employed has been investigated by various researchers from different academic disciplines. Two types of quantitative empirical studies can be distin-

⁴ Not all entrepreneurs are marginal. An exception are entrepreneurs that become self-employed immediately after graduating. However, Carroll and Mosakowski (1987) find that only 5% of de novo entrepreneurs start immediately after graduating.

guished. One type of studies use explicative models based on a set of hypotheses derived from theoretical considerations about factors influencing the self-employment decisions. Other studies develop structural models explaining self-employment. These models rely on the rational that self-employment occurs if the expected rewards of self-employment exceed those of paid work. Hence, after estimating earnings functions for self-employed and employees, for each individual the earning difference is calculated and used in the structural equation. An additional property of structural modeling is that the earnings functions contain a selectivity term that is constant through time. Thus, possible biases from time independent self-selection are corrected for.

2.3.1 The statistical relationship between inflow, stock and duration

The decision to become self-employed is researched under different sampling frames: stock and flow sampling. However, the consequences of applying different sampling schemes are not properly understood⁵. When sampling the flow into self-employment we are interested in the sorts of people who become self-employed, and when sampling the stock of the self-employed we are interested in the sorts of people likely to be found self-employed (Chesher and Lancaster 1983). Our research interest is in finding determinants of the selection process into self-employment, so the relevant population is the flow of individuals. Many social scientists have interpreted inferences based on stock samples as if they were taken from flow samples. That is, they have perceived equally the factors that influence the probability of being found self-employed with the probability of becoming self-employed. Inferences based on stock samples provide no information on the probability of becoming self-employed, unless the economy is in a steady state and one has estimates of the effects of covariates that influence the length of stay as a self-employed. As none of the empirical studies we reviewed have simultaneously researched the probability of being found self-employed and the probability of leaving self-employment, we can only use results on stock samples cautiously with related evidence of duration studies from different samples.

Still, if the economy is in a steady state, with inflow rate constant and survival not dependent on calendar time, then the inflow, stock and duration of self-employed are related through the following expression (Ridder 1982: 21):

$$(2.2) \quad \Pr(SE) = \Pr(IF) * E(T)$$

where, $\Pr(IF)$ is the probability of becoming self-employed in a certain period, $\Pr(SE)$ is the probability of being part of the stock of self-employed at a certain moment, and $E(T)$ is the mean (completed) self-employment duration.

If we assume that:

$$(2.3a) \quad \Pr(SE) = f(x, \beta_{SE})$$

⁵ For instance, a recent survey paper on the empirical evidence of self-employment selection by Le (1999) does not even mention the sampling schemes and the consequences of the different sampling schemes on interpretation of the results.

and

$$(2.3b) \quad \Pr(IF) = g(x, \beta_{IF})$$

If we assume furthermore the following conditions: (1) $f = g$, (2) linear, and (3) additive and let $\phi = f = g$ then equation (2.2) can be written in log-form and rearranged as (van Praag 1996: 64):

$$(2.3c) \quad \log[E(T)] = \log[\Pr(IF)] - \log[\Pr(SE)]$$

Assume that equation (2.3a) and (2.3b) are explained by the same set of regressors. Represent the j -th element of the set of regressors $\bar{x}$ by x_j . Then, the partial derivatives of the predicted probabilities with respect to x_j are:

$$(2.4) \quad \frac{\partial}{\partial x_j} \Pr(SE) = \phi(\bar{x}' \beta_{SE}) * \beta_{SE}^j$$

and

$$(2.5) \quad \frac{\partial}{\partial x_j} \Pr(IF) = \phi(\bar{x}' \beta_{IF}) * \beta_{IF}^j$$

where β_{SE} and β_{IF} represent vectors of parameters, where β_{SE}^j and β_{IF}^j represent the j -th elements of the parameter vectors and are consistent with the effect of x_j on the respective probabilities.

Define:

$$(2.6a) \quad \gamma_j \equiv \frac{\partial}{\partial x_j} \log[E(T)]$$

or, from equation (2.3)

$$(2.6b) \quad \gamma_j = \frac{\partial}{\partial x_j} [\log[\Pr(SE)] - \log[\Pr(IF)]]$$

$$(2.6c) \quad \gamma_j = \frac{\partial}{\partial x_j} \log[\Pr(SE)] - \frac{\partial}{\partial x_j} \log[\Pr(IF)]$$

By partial differencing, we have (see 2.6 Appendix for the complete derivation):

$$(2.7) \quad \gamma_j = \frac{1}{\Pr(SE)} \left[\phi(\bar{x}' \beta_{SE}) * \beta_{SE}^j - \phi(\bar{x}' \beta_{IF}) * \beta_{IF}^j * E(T) \right]$$

Equation (2.7) reduces to a numerical expression for the predicted effect of each covariate on $\log[E(T)]$, the mean duration of self-employment.

2.3.2 The bias in the determinants of the probability of becoming self-employed

The relevant empirical research on the determinants of the probability of entry suffers from three biases: (1) interpreting β_{SE} as β_{IF} , (2) left-truncation, and (3) time aggregation. The effect and direction of the first two biases are both a function of the length of stay as entrepreneur, so γ_j . If we know γ_j , we can make inferences about the nature of the bias and its effect on the true β_{IF}^j .

In case of bias (1), we can make use of established empirical evidence about the probability of being self-employed and the mean duration, to give indirect evidence about the effect of the regressor on the probability of becoming self-employed. Imputing the empirical evidence about the coefficients β_{SE}^j and γ^j into equation (2.7) provides the expected effect of x_j on the probability of becoming self-employed (see Table 2.1). Since empirical evidence on β_{SE}^j and γ^j comes from different studies, we can not always determine β_{IF}^j precisely.

Table 2.1: Predicted sign of β_{IF} given the signs of β_{SE} and γ

		γ		
		+	0	-
β_{SE}	+	?	+	+
	0	-	0	+
	-	-	-	?

In case of bias (2), the effect of the bias is not so clear as bias (1). The magnitude of the bias depends on the direction of γ^j and the length of the interval in which left-truncation takes place. Table 2.2 provides the direction of the bias on β_{IF}^j , given knowledge on γ^j . It encircles the cases for which the bias may be problematic for qualitative inferences about β_{IF}^j .

Table 2.2: Predicted bias given the signs of β_{IF} and γ

		γ		
		+	0	-
β_{IF}	+	+	0	-
	0	+	0	-
	-	+	0	-

The third bias, time aggregation, is unrelated to mean duration. Time aggregation bias occurs if a static methodology, e.g. pooling observations into a single period logit or probit, is applied to dynamic, multiperiod data. The parameter estimates are always biased and inconsistent (Shumway 2001). We do discuss the bias in detail in chapter 3 in this dissertation but the coefficients should always be interpreted with care.

2.3.3 Empirical studies on self-employment entry

Table 2.3 gives an overview of empirical studies on self-employment.⁶ The factors that influence the chance of becoming self-employed are divided into significant negative or positive impact or no significant effect. Table 2.3 provides direct and indirect evidence on the factors that influence the self-employment decision. Direct evidence comes from studies that draw inferences on flow samples, while indirect evidence follows from stock samples combined with evidence on duration studies according to the scheme of Table 2.1⁷. According to the empirical result each study is put in one of the three categories. A coefficient is considered significant if $p < 0.05$. More important, Table 2.3 indicates the biases in the empirical studies.

The overall picture taken from the existing research on self-employment is very ambiguous. Most surprisingly, the standard economic argument that ‘people will choose the activity with the highest (lifetime) income’ is not straightforward. Most studies find no significant effect on income differences (De Wit and Van Winden 1989; Dolton and Makepeace 1990; Rees and Shah 1986). Moreover, studies that find significant effects are based on stock samples.

The effects of human capital are not straightforward either. The results of general human capital are contradictory. Evans and Leighton (1989), Blanchflower and Meyer (1994), and Dunn and Holtz-Eakin (2000) find a positive effect of general education on the probability to become self-employed. Lazear (2004) finds that general education, in terms of educational program content, also increases the probability of entry. Conversely, Evans and Leighton (1989) find a negative impact of education on the probability of becoming self-employed using another sample.

⁶ This overview does not claim completeness. We attempt to include the main references. The literature review differs from other overviews (e.g. Parker 2004; van Praag 2001), by addressing biases.

⁷ The outflow and duration studies are related. If a regressor increases the probability of exit, then it reduces the duration of being self-employed.

Table 2.3: Overview of Empirical Studies Investigating the Self-Employment Decisions

Variables		Inflow			Stock			Outflow			Duration		
		-	0	+	-	0	+	-	0	+	-	0	+
Human Capital	Income difference				(1) (8)* (15)	(2) (11)							
	Education general	(3)* (9)**f (16)	(4) f (10) f (16)	(5b)**f (13)f (4)f (23)*	(6)* (11) (14) (1) (2)	(7) (12) (15)		(22)* (24)*	(5b)**f	(5a)**f	(25) (27)	(25) (27)	(8)
	Education occupational		(16)										
	Work experience		(4)f (9)**f (16)	(10)f	(6)*	(4)f (1)		(24)*	(26)*		(17) (27)	(8) (25)	
	Work experience squared				(6)*						(8)		
	Self-employment experience										(27)	(25)	
	Self-employment relevant experience										(17)	(27)	
	Assets			(4) f (9)**f (13) (10) f	(19)	(19) (11) (2)		(21)			(17)		
	Assets squared	(4)f	(4)f (10) f (20)**f	(9)**f	(19)	(19)							
	Inheritance					(18) (19)							
Physical/ Financial Capital	Home		(20)**f	(9)**f				(21)					
	Income	(10) f	(20)**f										
	Start-up capital							(24)* (26)*					

Table 2.3 (continued)

Social capital	Self-employed parents	(13) ^f (16)			(18) (19) (6)*			(26)* (27)*			(25)	
		(4)* ^f (10) ^f (9) * ^f	(3)* (4)* ^f (20)* ^f (13) ^f	(5b)* ^f (4)* ^f	(11)	(12) (6)* (14) (15) (1)	(19) (2)	(5a) * ^f (5b) * ^f (21)			(27)	(25)
	<i>Married</i>											
	<i>Emotional support spouse</i>											
	<i>Information from contacts</i>										(27)	(27)
	<i>Number of children</i>	(20)* ^f				(14) (15) (2)	(11)	(21)				
	<i>Age cat.</i>											
	0-4						(14)					
	5-9					(14)						
	10-12					(14)						
	13-14						(14)					
	15-20					(14)						
	<i>Dummy</i>					(6)*	(1)					
Controls	<i>Age</i>		(4)* ^f (9) * ^f (20)* ^f	(3)* (5a)* ^f (5b)* ^f (13) ^f		(7) (11) (2) (1)	(12) (14) (15) (1)	(5b) * ^f (21) (26)*	(5a) * ^f	(22)*	(27) (27)	(17) (25)
	<i>Age squared</i>	(3)* (5a)* ^f (20)* ^f	(4) ^f (9) * ^f (13) ^f		(12) (14) (15)	(7) (11)		(22)* (26)*	(25)*	(21)		
	<i>Females</i>	(23)* 16			(18) (19)			(24)*			(27)	

References: (1) Dolton and Makepeace, 1990 (UK), (2) Taylor, 1996 (UK), (3) Alba-Ramirez, 1994 (Spain), (4) Evans and Leighton, 1989 (US), (5) Blanchflower and Meyer, 1994 (a=Australia; b=US), (6) De Wit and Van Winden, 1989 (Netherlands), (7) Blau, 1985, (8) Bruderl et al. 1992, (9) Van Praag and Van Ophem 1995, (10) Evans and Jovanovic, 1989 (US), (11) Gill, 1988, (12) Borjars and Bronars, 1989 (US), (13) Dunn and Holtz-Eakin, 2000 (US), (14) Kidd, 1993 (Australia), (15) Rees and Shah, 1986 (UK), (16) Carroll and Mosakowski, 1987 (West Germany), (17) Van Praag, 2003 (US), (18) Blanchflower and Oswald, 1998 (UK), (19) Lindh and Ohlsson, 1996 (Sweden), (20) Holtz-Eakin, Joulfaian, and Rosen, 1994a (US), (21) Holtz-Eakin, Joulfaian, and Rosen 1994b, (22) Bates, 1990, (23) Lazear, 2004 (US), (24) Bruderl and Preisendorfer, 1996 (West Germany), (25) Taylor, 1999 (UK), (26) Gimeno et al. 1997 (US), (27) Bosma, Van Praag, Thurik, and de Wit, (2004).

Note: ^f = left-truncation, * = time aggregation

*: Effects are counted as negative or positive if the respective coefficients is different from 0 at a significance level of $p < .05$.

Alba-Ramirez (1994) for Spain, and Carroll and Mosakowski (1987) for West-Germany confirm the negative impact of education on the probability of self-employment. No effect was found by Evans and Jovanovic (1989) and Carroll and Mosakowski (1987) in different model specifications. The results from the stock and duration studies in Table 2.3 confirm this ambiguous picture of education as a determinant of self-employment selection. The results about more specific human capital are also not convincing. Evans and Jovanovic (1989) find a positive effect, but Carroll and Mosakowski (1987), Evans and Leighton (1989), and Van Praag and Van Ophem (1995) find no significant direct effect. The effects are even more contaminated by the biases inherent in most studies.

No study rejects hypotheses on the positive relationship between financial capital and becoming self-employed (Evans and Jovanovic 1989; Evans and Leighton 1989; Holtz Eakin et al. 1994). But some studies do not find significant effects of financial capital (Lindh and Ohlsson 1996). One of the most consistent and prominent factors to explain self-employment is whether the father has been self-employed (see e.g. Blanchflower and Meyer 1994; Carroll and Mosakowski 1987; Lindh and Ohlsson 1996; Taylor 1999). While the influence of the father's economic activity is manifest, these studies do not provide a clear vision about other forms of social capital, such as partnerships and children. Marriage is found to have a negative (Evans and Jovanovic 1989; Evans and Leighton 1989; Van Praag and Van Ophem 1995), positive (Blanchflower and Meyer 1994; Evans and Leighton 1989) or insignificant effect on the probability of becoming self-employed (Alba-Ramirez 1994; Dunn and Holtz-Eakin 2000; Evans and Leighton 1989; Holtz-Eakin et al. 1994). The stock and duration studies in Table 2.3 find usually an insignificant effect.

We can conclude two things from the empirical literature: (1) the effects of the covariates are unstable across samples, and (2) data limitations and utilizations may cause significant biases. These ambiguities and biases can be overcome in the following ways. Utilize a flow sample, thereby directly measuring the probability of becoming self-employed. Apply a dynamic methodology to dynamic data, so that the estimates do not suffer from severe time aggregation. Besides these sampling issues, make variables time-varying where possible. In this way, the variables influence the hazard only when the effect should take place, and not allow it to be constant over life.

2.3.4 Empirical studies on self-employment exit and duration

The decision to exit self-employment has been investigated from several angles. These studies investigated the propensity to exit from an individual perspective or from a firm perspective. Both types of studies share their focus on finding determinants of survival in self-employment. The exit studies that take an individual perspective focus mainly on the individual's human capital, in terms of education, training, and experience, and the individual's financial capital, in terms of previous income or assets. Exit studies that take a firm perspective focus primarily on the size of the enterprise, the business strategy of the firm, and macro-economic conditions. As we are interested in individual determinants

of exit from self-employment, we will focus mainly on studies with a focus on the individual entrepreneur.

In contrast with the models on self-employment selection, empirical studies on self-employment exit do not suffer so much from the methodological flaws that plagued entry studies. There are several reasons. First of all, most studies use a flow sample. They follow a cohort of firms or individuals that enter at the same time for a certain time period. Of course, not all firms will exit within the observed time period. These firms are censored. Therefore, the natural econometric candidate is a duration model. Although some early studies use a binary discrete choice model to study who exit within a certain time frame⁸ (e.g. Bates 1990), most recent studies that focus on the determinants of the probability of exit and small business longevity use duration models (see e.g. Bosma et al. 2004).

The results concerning the relationship between human capital and the probability of exit are mixed. We see the same pattern as with the relationship between human capital and the probability of entry. Some find a positive relationship, some no relationship, while others find a negative relationship. Taylor (1999) argues for example that academic success is a poor indicator for entrepreneurial ability. He claims that, consistent with the signaling hypothesis, individuals who plan a career in self-employment have no need to acquire formal qualifications. However, more self-employment specific experience increases the probability of survival in self-employment (Bosma et al. 2004; Taylor 1999).

There is some evidence that individuals who have more access to financial capital have better survival chances. Bates (1990), for example, finds that individuals who have started up a business with above average levels of capital have a lower probability of exit. More evidence comes from the studies that find a positive relationship between firm size and survival (e.g. Brüderl and Preisendörfer 1998; Gimeno et al. 1997). If larger firms have had better access to financial capital, then a positive relationship between firm size and survival provides indirect evidence for a positive relationship between financial capital and success.

2.4 Social capital

In this literature review we will try to gain some understanding of the concept of social capital both from a structural and a rational choice (agency) perspective. The structural perspective distinguishes itself from agency by making macro assumptions of social structure, which provide context for analyzing human behavior. Gaining understanding of the concept of social capital allows us to understand why and how social capital works (or not) in general, and in specific for self-employment. Therefore, we have to understand the genealogy of social capital: its causes, its nature, and its consequences. In this review on social capital, we will render social capital as a process, starting with its formation and ending with its consequences.

Social capital was defined in section 2.1 as these features of social life, like networks, norms, and trust that enables individuals to reap benefits from interacting with

⁸ I will discuss the implications of time aggregation in chapter 3.

each other. Most of the research on social capital deals with the consequences of social capital. A review of the literature learns that social capital affects people, states, cities, and even nations.

For some, social capital is considered as the mix of sociology and economics (Frank 1992). As social capital is the bridge between these two distinct disciplines, they bring different assumptions and methodologies of defining, measuring and interpreting social capital⁹. Historically, economists have advocated methodological individualism. This principle implies that all collective activity can be traced back to the behavior of individuals. This view is prevalent in the utilitarian tradition of classical and neo-classical economics, in which economic actors behave as self-interested utility maximizers who are only minimally affected by social relations (Granovetter 1985).

In contrast, early sociologists rejected this individualism and advocated methodological collectivism or holism, inspired by sociologists as Durkheim and Weber. This principle implies that groups often have traits, behaviors and outcomes that cannot be understood by reducing them to their individual parts and that the relationships among individuals affect their behavior. These diverging assumptions about action within a social context bear a strong influence on theoretical and empirical work in both sciences. Theoretical work in the field of sociology emphasizes the social structure and how the social structure constraints individual choice and behavior. In contrast, economists have put more interest in the freedom of the individual and its subsequent purposeful and rational actions. The last two decades have witnessed a shift away from these pure views on individual behavior and action within a social context, as is reflected for example in the works of Becker (1996) *Accounting for Tastes* and Becker and Murphy (2000) *Social Economics*¹⁰. These works preach cautiousness in analyzing individual choices within a social framework.

2.4.1 Theory: The social capital process

Figure 2.1 summarizes the social capital process. The first phase deals with the determinants of social capital. How is social capital formed? The second phase shows how the formation of social capital brings into place opportunity structures for deriving social capital benefits. However, a beneficent opportunity structure is no guarantee for network benefits. In the third phase, the mobilization and use of social capital determines the productivity of the opportunity structure. Finally, we arrive at the individual effects (these can be positive or negative – see Portes (1998) and Portes and Landolt (1996) for downside effects of social capital). Figure 2.1 also schematically reveals the different methodologies used in social capital research. The figure presents the pure form of agency and structure and how they form different paths to social capital outcomes. We differentiate three paths: the structural path, the agency path, and the instrumental path.

The structural perspective distinguishes itself from agency by making macro assumptions of social structure, which provide context for analyzing human behavior (Lin 1990). The first assumption states that most societies are structured hierarchically at the

⁹ These problems are nicely illustrated by a famous quote of Duessenberry that “economics is all about how people make choices; sociology is all about how they don’t have any choices at all” (1960: 233).

¹⁰ The fact that Gary Becker has contributed significantly in mixing sociology with economics can be attributed to the fact that he originally had a chair in sociology before he moved to economics.

Figure 2.1: The Social Capital Process

macro level. Every level of the hierarchy reflects different demands for valued goods like status, wealth, and power. The second assumption states that these hierarchies of valued goods have the tendency towards congruence and transferability. Individuals who value wealth also value status and power. Valued goods are traded within hierarchies to reproduce class. The last assumption states that most hierarchies tend to be pyramidal shaped in the sense that there are more people with less valued goods and less people with more valued goods. These assumptions are guided by the micro-assumption of homophilous interaction. This principle states that social interactions takes place among people at similar hierarchical levels, also known as the like-me hypothesis (Homans 1950; Laumann 1966).

These assumptions define the structural path of social capital. The structural path is reflected in one of the first definitions of social capital by the French sociologist Bourdieu:

“Social capital is the aggregate of the actual or potential resources that are linked to a possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition...a ‘credential’, which entitles them to credit, in the various senses of the word,” (1986: 249).

The network of relationships is perceived as an asset that can be used to the advantage of individuals, groups, and families. Access is defined by origin, which is reflected in the correlation between the resources of parents and resources available for the child. Individual outcomes from a structural view are contingent on their structural position. The social resources available come from people of equal structural positions.

The agency path starts with the individual. The individual is a rational actor, which purposefully creates social organization (Coleman 1990). The individual has only one goal, promoting its own interests. This perspective allows the individual to make its own choices, based on its idiosyncratic preferences. Like investments in human capital, individuals invest in social capital by maximizing the utility of today’s costs of investments with its discounted future benefits.

This agency path into social capital is probably best reflected by Glaeser and his colleagues’ economic approach to social capital based on rational choice (Glaeser 2001; Glaeser et al. 2002). From their view point, social capital is defined as:

“A person’s social characteristics – including social skills, charisma, and the size of his Rolodex – which enables him to reap market and non-market returns from interactions with others.” (Glaeser et al. 2002: 438)

They assume that an individual’s social capital comprises both the intrinsic abilities (e.g. being extroverted and charismatic) and the results of social capital investments (e.g. a large social network). This view on social capital put aside all structural constraints and focuses solely on the rational choices of individuals. The individualistic focus in this definition of social capital reveals the rational choice approach. Besides the structural path and agency path, there is a third path of social capital.

The instrumental path of social capital combines elements from the structural and the agency perspective (Lin 1990; Lin 1999). It takes as starting point the structural perspective, however adds the micro-assumption of social action. In contrast to expressive actions initiated to maintenance and perseverance of valued resources, there are

instrumental actions. Instrumental actions are undertaken by individuals in order to obtain valued resources that are not at the disposal of the individual. Those actions are initiated towards individuals higher up in the hierarchy who possess valuable resources. It is often assumed that the ties that bridge different levels of hierarchies are weak ties (Granovetter 1973). These weak bridging ties are in conflict with structural assumptions and the principle of homophilous interactions and are sometimes assumed subordinate to structure (Lin 1990). Nonetheless, these weak bridging ties loosen the individual from its stringent structural constraints and allow him upward mobility in social structure beyond its position.

To summarize, we see that there are different paths in the social capital process that lead to different outcomes. The structural path is merely focused on the reproduction of class and its resources. The instrumental path deals with upward mobility within a structural framework, while the agency path shows how individual outcomes are the result of rational self-interested choices. These paths through the stages of the social capital process give rise to different empirical contributions to the concept of social capital.

2.4.2 Social capital formation

What drives the formation of social capital? Let us claim for simplicity that an individual's social capital comprises his social network and the resources available in the relationships of his network. From a structural perspective, the drivers of an individual's network and the resources available in this network stem from an individual's original socio-economic position. Blau and Duncan (1967) pioneered the research on social origin. They claim that individual attainment does not only depend on education, but also on the socio-economic background of the parents. Lin (1990) adapts this claim into his strength-of-position hypothesis. This hypothesis states that the level of original position is positively associated with access and use of social resources. Given the principle of homophilous interactions, the better the original position of an individual, the better are the resources he can access and mobilize through its social network. An individual's initial position is often indicated by parental resources or attained resources. Evidence for this claim comes from a random job search survey of 399 individuals conducted in 1975 in the New York area (Lai, Ling, & Leung, 1998). Parental resources, measured by the respondent's father's occupational status and educational level, show a positive association with an individual's attained resources.

These structural parameters are clearly out of an individual's control. Even though its origins stem from sociological research, the claim that original position affects individual outcomes is widely recognized in economic research (Loury 1977; Mulligan 1999). Mulligan investigates the determinants of earnings differentials between parents and offspring comparing the income of 3,247 individuals in 1991 with parental income generated from the 1970 Census Public Use Microsample. He finds that parental permanent income accounts for 40% of the association between parental success and offspring success, even after controlling for a rich set of control variables, including cognitive test results. Bowles et al. (2001) suggest that the transfer of non-cognitive skills not captured by conventional cognitive tests might account for the intergenerational transfer of suc-

cess. These skills include attitude towards risk, ability to adapt to new economic conditions, hard work, and the rate of time preference.

The agency perspective has a different approach in establishing the determinants of social capital. In this perspective, not the structure determines the formation of social capital, but the rational actions of an individual. Coleman's pioneering works (1986; 1988; 1990) claim that rational actors create their own social networks for achieving certain ends. Glaeser (2001) and Glaeser et al. (2002) provide a pure economic approach to the formation of social capital. In their view social capital is the result of optimal individual investment decisions. They extend the definition of social capital as social networks with their determinants, namely a person's social characteristics that allow him to develop a social network and reap market and non-market returns.

The main empirical evidence comes from the study of Glaeser et al. (2002). They used cross sectional data from 1972 to 1998 of the General Social Survey in the United States. To measure the stock of social capital the number of memberships in types of organizations is used. Age shows an inverted u-shape relation with organization membership. This supports the claim that investments in social capital decline when the number of periods to recoup investment costs decreases. Mobility shows a negative relation with membership, supporting the claim that people invest more if their benefits from investing in social capital rise. Mobility decreases the returns on investments in social capital, because mobility reduces also the number of periods to recoup investment costs. The benefits for investing in social capital are assumed to be positively related with the sociability of and individual's occupation. The data confirms this prediction. Individuals with benefits from social interaction invest more in social relations. The data fails to support the model's prediction that membership falls with the opportunity cost of time. The opportunity cost of time is indicated by income or educational level¹¹. Both income and educational level have a positive impact on membership. The relation between education and social capital can result from the possibility that social skills are learned at school. The positive relation between human and social capital is supported in a cross-sectional study of 1402 Dutch top managers (Boxman et al. 1991). This study shows that human capital, measured by formal education, work experience, and number of formal jobs, is positively associated with social capital, measured by work contacts and memberships of clubs.

From a theoretical stance, there is one theory that attempts to integrate structural determinism with an agency perspective of rational choice. Becker's (1996) economic theory on social capital accounts for tastes of related individuals on a person's utility through the stock of social capital. He defines social capital as:

"The influences of past actions by peers and others in an individual's social network and control system." (Becker 1996: 4).

¹¹ Although the opportunity cost of time increases, higher educated individuals might also use time more efficiently, reflected in the ability of high educated individuals to generate a greater variety of non-routine activities (Gronau and Hamermesh 2001). It should be noted for completion that these economic explanations focus more on the cost of civic engagement, while the argument by Glaeser et al. (2002) stress more the importance of high returns e.g. the self-employed pharmacist membership in all clubs to have contact with its customer.

In Becker's opinion, the social network does not determine directly the behavior and decisions of individuals, rather indirectly through the influence on the relative utility of goods (Becker and Murphy 2000). An individual's social network determines his preferences and these in turn determine his choices. Once the social network has formed the actual formation of social capital is mostly outside the control of the individual. Therefore, the real moment of choosing is choosing the social network. Rational individuals will choose the networks that support their way of life. However, structure comes in ones more, when Becker explains:

"Still, the choices available may be limited by market prices and other circumstances. A teenager enrolled in a particular school may have little choice over the peer pressure she feels. The equilibrium degree of segregation between blacks and whites, rich and poor, and other groups is determined by market and other forces that are partly beyond the control of the individual." (Becker 1996: 13).

This quote makes us agnostic in judging the value of social structure and agency. The tension between structure and agency seems to be more the combination of an individual's freedom to make rational decisions with whom to enter into social networks, given the opportunities and constraints provided by social structure. Evidence for the structural formation and consequences of social capital in terms of social networks and resources is well developed. However, the rational choice approach to the formation of social capital in terms of an individual's investment decisions, based on his social characteristics, is still in its infancy.

2.4.3 Social capital access

In this section we shift attention from the formation to the form and functions of network structures. From an outward perspective, networks can be closed or open, based on the embeddedness of the ties. However, a thorough network analysis should not only focus on the forest, the network, but should also take an inward perspective on the trees, its ties. Different perspectives on the form of networks are best understood by their function. Open networks with weak bridging ties strive for structural autonomy and opportunity (Burt 1992), while closed networks with strong bonding ties strive for social closure and support (Coleman 1990). These perspectives enclose different theoretical and empirical contributions to the concept of social capital.

2.4.3.1 Closed networks and social capital

Understanding the differences between open and closed networks is best understood by looking at its contributors and their methodological assumptions. The work on closed networks and its functions can be traced back to Coleman's (1988; 1990) seminal works on social theory and social capital. Coleman applies methodological individualism in the form of rational choice to social organization. Individuals purposefully promote their interests by forming social relations. Without these purposeful interactions of individuals no social organization would exist (the core argument of methodological individualism).

These assumptions lead to the most well known definition of social capital:

"I will conceive of ... social-structural resources as a capital asset for the individual, that is, as social capital. Social capital is defined by its function. It is not a single entity, but a variety of different entities having two characteristics in common: they all consists of some aspects of social structure, and they facilitate certain actions of individuals who are within the structure." (Coleman 1990: 302)

Although Coleman conceives social capital as an individual asset, it is a property of the social structure and inheres in relations. This relational property of social capital is reflected in the forms of social capital (Coleman 1990). The first form of social capital is obligations and expectations. This form of reciprocity allows extending the time horizon for exchange. Favors done today might be repaid in the future. Social capital in this form reduces transaction cost and makes social exchange more efficient, as financial capital made economic exchange more efficient (Frank 1992). The second form of social capital is information. This form of social capital is important for timely access to new knowledge. Membership in a social structure may provide information about finding jobs (Granovetter 1974). The third form of social capital that emerges out of social structure is norms and legitimate authority. Norms are an effective way to regulate behavior. These norms come close to Becker's (1996) definition of social capital. Norms can be interpreted as the influences of past actions of individuals in an individual's social network and control system. The last form of social capital is the appropriability of social organization. Social relations and social organization serve often-other purposes for which they are originally created¹². The appropriability of social capital reduces the set-up cost for forming new networks for new purposes.

Coleman stresses the importance of social closure in social organization in order to increase the effectiveness and trustworthiness of social organization. The individual has little control over its functions, once social organizations are formed (cf. Becker 1996). This definition of social capital as a property of collective social structures has boomed studies on social capital (Fukuyama 1995; Guiso et al. 2000; Putnam 1993). Most of these studies use survey-based measures of country-level trust that affects collective level outcomes as economic growth (Knack and Keefer 1997) and government corruption (LaPorta et al. 1997). Moreover, most of the economic literature of the post-Coleman era use these group effects at the state, regional, and country-level as the unit of analysis, reflecting the notion that the formation of social capital is a group-level attribute and should be studied at some high level of aggregation (Glaeser et al. 2002). However, the interest in this literature review is on social capital as either an asset or liability for the individual. For a more individually oriented view on social capital we must shift to the open network view. Paradoxically, where Coleman's theory on closed networks starts with rational choice, Burt's theory (1992) on open networks starts with structure.

2.4.3.2 Open networks and social capital

Burt's (1992) view on social capital starts with social structure. Burt transforms the social structure in a capital in its own right. His interest is not on whom we reach (Lin 1990), but how we reach. Burt transforms the function of social structure from constraint to

¹² This could have problematic implications for discounting future utilities of social relations.

opportunity, from freedom to power and from absolute to negotiated control. He defines social capital:

“Relations with other players, friends, colleagues, and more general contacts through whom you receive opportunities to use your financial and human capital.” (Burt 1992: 9).

The individual actor becomes an entrepreneur in a competitive arena where everybody is bargaining for resources. The entrepreneur should position itself in the social structure in such a way that he receives information and control advantages. Efficiency is achieved by selecting contacts that provide non-redundant information. Non-redundant contacts serve as bridges to new information and opportunities. The principle of non-redundancy shows strong resemblance with the distinction between weak and strong ties. Based on the classic work of Granovetter (1973), weak ties are assumed to be bridges to new and fresh information, while strong ties often create redundancy. Burt acknowledges that bridges are often weak ties, but does not a priori rule out that strong ties can be bridges too. Burt proposes the concept of structural holes in order to complement the information benefits with control benefits for being the broker between two unconnected alters. An individual's optimal strategy should be based on structural autonomy where at his end of the relationship he is free of structural holes, while at the other end he is rich in structural holes.

If we compare Burt's (1992) concept of social capital with Coleman's (1990) concept of social capital, two things become clear. Social capital in the form of opportunities is individually created and controlled, while social capital as trust and norms is created by the collective. Both forms of social capital can be a resource for the individual. The parameters to be optimized for information and opportunity are network size and diversity, while social support demands strong ties and closure. Therefore, networks of different configurations should be judged according to their function. Open network functions would benefit from capital widening investments, while closed network functions would benefit from capital-intensive investments.

2.4.4 Mobilization of social capital

The last moderating phase in the social capital process deals with the instrumental actions from the opportunity structure. Mobilized social capital refers to the use of specific contacts for certain purposes. The appropriability of social relations (Coleman 1990) allows multiple instrumental actions from ties. The status of the contact used in instrumental actions is negatively related with tie strength and the status of the contact is positively related with an individual's education (Lin 1999). It is assumed that the value of the contact used in instrumental action is positively associated with the outcome of the action. This last phase in the social capital process gives insight in the productivity of the opportunity structure.

The access and use of social capital is often not separated in empirical literature. The only exception is the literature on job search and status attainment. The study of Lai, Lin, and Lueng (1998) find support for the negative relation between tie strength and the status and resources of the contact used in a job search survey of 1975. Results of this study indicate that a contact's resources have a positive impact on the status outcome of

the individual. Furthermore, De Graaf and Flap (1988) find a positive relation between the value of the contact used and the outcome of the instrumental action in a Dutch job search survey.

Next to the job search literature, there is implicit evidence of the importance of separating access from use of social capital. Coleman's (1988) pioneering work points in a similar direction. He uses the *High School and Beyond* dataset of 4,000 children to show the importance of social capital in the creation of human capital. The results show that the presence of both parents reduces a child's chance of dropping out of school with 6.0 percentage points. This indicator of social capital measures a child's access to social capital in the form of the presence of both parents. In the same analysis Coleman finds that the mother's expectation for child's education (no college versus college) reduces the chances of dropping out of school with 8.6 percentage points, controlling for the presence of both parents. The mother's expectation captures the mobilization of social capital and provides a better predictor than social capital access.

There is similar evidence in the study on 1,710 business founders in Munich and Upper Bavaria (Brüderl and Preisendörfer 1998). Their analysis differentiates between levels of support from weak ties, indicated by business partners, acquaintances, former employers, and former co-workers and the level of support from strong ties, indicated by spouse, parents, friends, and relatives. The results of this study indicate that medium to high support from weak ties compared to low support is more positively related to firm growth measured by employment or sales, while equally positively related to survival. In addition, medium and high support from strong ties compared to low support is more positively associated with survival and sales growth, and equally positively associated with employment growth. The results seem to indicate that it is not so much the presence or the strength of the tie that determines its instrumental effects, but more the actual use and motivation of the tie.

The best illustration for stressing the necessity to separate access from mobilization comes from a panel study of 1100 Dutch business founders surveyed between 1994 and 1997 (Bosma et al. 2004). This study shows that the presence of a spouse is marginally negatively related with survival time and not related with profit, while the actual emotional support is positively associated with both survival time and profit. Differentiating access from use of social relationships in this case seems to transform a social liability into social capital. Results have shown that social capital outcomes differentiate with the specificity of its measurement. In the next session, we will discuss the consequences and benefits of social capital formation for entrepreneurship.

2.4.5 Social capital outcomes – Integrating self-employment and social capital

The role of social capital in economic, political, and social life has been investigated in great detail. But how can social capital benefit entrepreneurship and self-employment? There are several theoretical reasons why relationship-specific investments can influence productivity and survival. Information conveyed in social relations reduces the uncertainty about competitor and supplier prices in imperfect markets (Stigler 1961). Investments which are specific to a certain relationship can create trust and economize on

transaction costs and can lead to a concentration of trade between identified partners (Ben-Porath 1980). And social relations provide information that reduces search costs when markets show frictions (Kranton 1996). Furthermore, social relations can provide access to financial capital when entrepreneurs are liquidity constraint. Also, the spouse may provide cheap labour or income. However, there are only a few empirical studies that have tried to substantiate these theoretical claims. These studies focus mostly on the effect of access and mobilization of social capital on the success or survival chances of newly founded firms.

Table 2.4 summarizes the most relevant empirical evidence of the relationship between social capital and self-employment. In their study on self-employment, Brüderl and Preisendorfer (1998) pose the network success hypothesis. This hypothesis claims that the social network is related to organizational performance through better access to information, customers and suppliers, and informal credits. They argue that the family network often provides unpaid family work and may provide emotional support. Their results indicate that support from strong ties, like family and friends, is more important for success of new businesses than support from weak ties, like business partners and acquaintances. It should be noted that these authors did not measure the structure or size of the social network, but merely how the entrepreneur mobilized their social network to support business survival.

Pennings, Lee, and Van Witteloostuijn (1998) examined the effect of social capital upon firm dissolution of Dutch accounting firms for the period 1880-1990. They focused on specific social capital namely the firm member's connectedness with potential clients. The information conveyed in social relations is especially important in this industry, where the information of the quality of the professionals is imperfect. Similar contacts may offer redundant information and to test for this possibility they included network heterogeneity measures. Their results showed that the size of the network increased the survival chances of the firm, but network heterogeneity did not alter firm survival.

Quite recently, Bosma et al. (2004) researched the survival chances of a sample of Dutch start ups. Building on the social capital investment model of Gleaser et al. (2002), they tested how investments in social capital affect the success of a newly founded business in terms of survival, profit, and employment. They argue along human capital theory that more specific investments in social capital (and human capital) should increase success more than general investments because general investments have a broader scope and longer time frame to recoup investments. They find some support that more specific investments in human capital and social capital, like gathering information from commercial and business channels, increase success more than general investments. All three studies use some crude measures and some loose theoretical arguments why access and mobilization of social relations can increase the success chances of entrepreneurs. In the last years, empirical studies started to investigate the connection between social relations and firm performance including social networks into the production function of firms. Barr (2000) predicts that social networks affect firm productivity indirectly through knowledge flows between firms. Fafchamp and Minten (2002) argue that social networks provide information and trust to overcome or exploit market imperfections and therefore raise profitability. Profitability is increased either directly through price discrimination or indirectly through an increase in the productivity of factor inputs. Both

Table 2.4: Empirical Results Social Capital and Self-Employment

Study	Sample	Social Capital Indicators	Results
Bosma et al., 2004	Flow sample 11 00 Dutch entrepreneurs in 1997	Contact entrepreneurs in network	Contact with entrepreneurs increases the amount of employment generated by the firm
		Way of information gathering:	Gathering information via general channels increases survival and employment; via direct business relations has no effect; via commercial relations increases survival, profit, and employment; and via fellow entrepreneurs increases profit.
		- general channels	
		- direct business relations	
		- commercial relations	
		- fellow entrepreneurs	
		Emotional Support from spouse	Emotional support from spouse increases survival and profit.
		Presence of spouse	Presence of spouse reduces survival, but increases employment.
Pennings et al., 1998	Sample of 1851 Dutch accounting firms	Partners from client environments	Having partners from client environments reduces firm dissolution
		Associates from client environments	No effect
		Partners to client environment	Having partners to client environments reduces firm dissolution.
		Associates to client environment	No effect
		Heterogeneity of partners origin	No effect
		Heterogeneity of associates origin	No effect
		Heterogeneity in associates origin	No effect
		Heterogeneity of departed associates origin	No effect
Brüderl & Preisendorfer, 1998	1700 newly founded ventures in Upper Bavaria (Germany) in 1990	Support from strong ties	More support from strong ties is associated with increased survival chances and increased sales growth.
		Support from weak ties	More support from weak ties is associated with increased sales growth
		Active help from spouse	Active support from spouse is associated with an increase in survival chance and sales growth
		Emotional support from spouse	Emotional support from spouse is associated with increased survival chances.
Barr, 2000	236 Ghanaian entrepreneurs in 1992 and 1993	Total number of contacts	An increase in the total number of contacts increases the value added of entrepreneur.
		Heterogeneity of contacts	More heterogeneous contacts increase value added of entrepreneur.
		Mean number of contacts	An increase in the total number of contacts integrated in wider network increases the value added of entrepreneur.
		Integrated in wider network; Heterogeneity of contacts integrated in wider network	More heterogeneous contacts integrated in wider network increases value added of entrepreneur.
Fafchamps & Minten, 2002	739 agricultural traders from Madagascar in 1997	Number of relatives in agricultural trade	An increase in the number of relatives in agricultural trade reduces total sales.
		Number of traders known	An increase in the number of traders known increases value added and total sales.
		Number of potential lenders	An increase in the number of potential lenders increases value added and total sales.

studies find strong evidence for a positive relationship between social networks and firm productivity.

2.5 Conclusion

In this chapter we reviewed the literature on entrepreneurship and social capital. We surveyed the canonical theoretical models on self-employment entry and exit. Self-employment entry is mainly investigated from an occupational choice perspective, while the theory of self-employment exit is mainly dominated by Jovanovic (1982) learning model. Then, we surveyed the empirical evidence of determinants of the probability of entry and the probability of exit. Especially the results of the empirical studies that investigated the determinants of self-employment selection were contradictory. We have found several reasons for these contradictory results. The main problem is the sampling frame used. Researchers have both used stock and flow sampling techniques to investigate the determinants of self-employment. Inferences from stock samples do not provide any information on determinants of the probability of entry, unless simultaneously information is known about the probability of exit. Even if this information is available, these results are only valid under steady state conditions. Empirical research on the determinants of the probability of exit is less methodological flawed. Most researchers use a flow sample and respect the time dimension by utilizing duration analysis.

Besides the theoretical and empirical research on self-employment, we investigated if and how social capital can be used for entrepreneurship. We restricted our focus on individual social capital, which we defined as the purposeful investment in social relationships with expected returns. These expected returns can be as broad as resources, information, trust, and control embedded within social relationships and social networks. We surveyed the origin of social capital from sociology and economics. These approaches differ mainly in their perception of the formation of social capital. From an economic perspective, investments in social capital are a rational decision trading off cost and returns. From a sociological perspective, investments in social capital are constrained by the social structure the person is embedded in. From an integral perspective, both approaches have their merits, and economists like Gary Becker have introduced many sociological concepts, as group behaviour and norms, in economic analyses.

Finally, we have surveyed the empirical literature that simultaneously analysed social capital and self-employment. This literature is scarce and underdeveloped. Most studies merely research the effect of the size of the social network on self-employment performance measures, like sales, growth, and survival. Although these studies confirm that social capital may be an important predictor of self-employment success, the reason why and how social capital leads to success has not been put in a proper theoretical perspective.

Based on the survey of the literature, we propose the following recommendations for research. (1) Although the theory on self-employment entry and exit is well developed, different theoretical models are used to explain self-employment entry and exit; occupational choice models for entry and learning models for exit. Both theories have their merits. The occupational choice model performs better in predicting individual thresholds, while the learning model is better able to explain the behavior of firm after

entry in the market. However, both models have their limitations as well. Both models ignore the sunkness of the set up costs. Introducing sunk costs in theorizing on individual entry and exit decisions can explain why individuals are reluctant to enter or exit self-employment. Focusing on the sunkness of the set up costs, would fit nicely in the research on investment decisions under uncertainty (see e.g. Dixit and Pindyck 1994). We apply the research on investments under uncertainty to the self-employment decision in chapter 7. (2) The theory on self-employment entry should differentiate between different types of self-employment. This suggestion was already made in Carroll and Mosakowski (1987) but has not been fulfilled to date. Actually, Carroll and Mosakowski are the only ones who have differentiated family self-employment from de novo self-employment. In chapter 4 of this dissertation, we will build a model that differentiates between family self-employment and de novo self-employment. (3) The research on self-employment selection has mainly been studied, theoretically and empirically, from a static perspective. This has lead to biased parameter estimates and neglected the inherent timing of the self-employment decision. In chapter 3 and 4, we investigate the timing of the entry decision. (4) The literature review has shown that social relations can increase self-employment success. However, these studies have not explicitly modelled why and how social relations are productive. In chapter 8, we will investigate the formation of productive social relations in an entrepreneurial context.

2.6 Appendix

From the text we have:

$$(2.8) \quad \Pr(SE) = \Pr(IF) * E(T)$$

and

$$(2.9) \quad \frac{\partial}{\partial x_j} \Pr(SE) = \phi(\bar{x}' \beta_{SE}) * \beta_{SE}^j$$

$$(2.10) \quad \frac{\partial}{\partial x_j} \Pr(IF) = \phi(\bar{x}' \beta_{IF}) * \beta_{IF}^j$$

and

$$(2.11) \quad \gamma_j = \frac{\partial}{\partial x_j} \log[\Pr(SE)] - \frac{\partial}{\partial x_j} \log[\Pr(IF)]$$

Partial differencing using the rules:

$$(2.12a) \quad \frac{\partial F(G(x))}{\partial x} = \frac{\partial F}{\partial G} * \frac{\partial G}{\partial x}$$

where (2.12b) $F = \log(G)$, and (2.12c) $G = \Pr(x)$, with (2.12d) $\frac{\partial F}{\partial G} = \frac{1}{G}$, and (2.12e) $\frac{\partial G}{\partial x} =$ equation (2.9) or equation (2.10).

Fill in (2.12a) into (2.11) renders:

$$(2.13) \quad \gamma_j = \frac{1}{\Pr(SE)} * \frac{\partial \Pr(SE)}{\partial x_j} - \frac{1}{\Pr(IF)} * \frac{\partial \Pr(IF)}{\partial x_j}$$

Rewrite (2.8) into

$$(2.14a) \quad \Pr(IF) = \frac{\Pr(SE)}{E(T)}$$

or equivalently

$$(2.14b) \quad \frac{1}{\Pr(IF)} = \frac{E(T)}{\Pr(SE)}$$

Fill in (2.14b) into (2.13) renders:

$$(2.15a) \gamma_j = \frac{1}{\Pr(SE)} * \frac{\partial \Pr(SE)}{\partial x_j} - \frac{E(T)}{\Pr(SE)} * \frac{\partial \Pr(IF)}{\partial x_j}$$

or equivalently

$$(2.15b) \gamma_j = \frac{1}{\Pr(SE)} \left[\frac{\partial \Pr(SE)}{\partial x_j} - \frac{\partial \Pr(IF)}{\partial x_j} * E(T) \right]$$

which gives equation (2.7) in the text.

Chapter 3

Time aggregation bias in estimating the probability of entrepreneurship

3 Time aggregation bias in estimating the probability of entrepreneurship

3.1 Introduction

The application of a static single period classification method to multiperiod data, like the simple probit and logit, implicitly assumes that the probabilities of choice are stable over time. The temporal process that generates change has reached an equilibrium state (Carroll 1983: 425), and the probability to become an entrepreneur has no duration dependence. In order to understand the choices of individuals for entrepreneurship, we need to think about the stochastic processes that are consistent with individuals making choices. Then, the data-generating process should dictate model choice (Lancaster 1990). Failure to do so will guarantee misspecification and lead to incorrect inferences.

This paper examines the empirical effects of time aggregation bias when researchers apply a static single period classification methodology to multiperiod decisions. When social researchers study qualitative choices, they sometimes use static research methods which ignore the temporal dimensions of the research problem. By estimating single-period classification models with multiple-period data, they produce choice probabilities that are biased and inconsistent estimates of the true population probabilities (Shumway 2001: 87).

Studies on time-aggregation can be classified in two ways. This first type of studies discuss the consequences of applying static research methods to multiperiod data (Allison 1984; Carroll 1983; Cox and Oakes 1984; Shumway 2001). Utilizing a static methodology by aggregating the time dimension and dichotomizing the dependent choice variable is both arbitrary and inefficient (Allison 1984; Cox and Oakes 1984). It is arbitrary because the end date of the time dimension is often determined by the researcher and it is inefficient because it ignores the information contained in the variation through time. Shumway (2001) provides the econometric properties of static models and dynamic hazard, or duration, models. He shows theoretically and empirically the inconsistency of the static estimator. The hazard approach is theoretically preferable, because it corrects for periods at risk and allows for time-varying covariates.

The second type of studies focuses on the effects of time-aggregation within duration models (Bergström and Edin 1992; Petersen 1991; Petersen and Koput 1992; Roed and Zhang 2002). These studies investigate on the effects of grouping data into intervals, for example daily information into weekly data. Petersen (1991) and Petersen and Koput (1992) show that the bias of time-aggregation in estimates increases with the size of the time interval.

Another purpose of this study is to investigate the sensitivity of time to entrepreneurship regressor estimates to distributional assumptions of the discrete hazard. Previous research on the sensitivity of duration estimates focused on the sensitivity of the estimates to distributional assumptions of the parametric baseline hazard (Addison and Portugal 1987; Bergström and Edin 1992). Others have focused on the sensitivity of parameter estimates to different specifications of the cumulative density function of the discrete

hazard (Royston and Parmar 2002; Sueyoshi 1995). Bergstrom and Edin (1992) provide empirical evidence of the effect of time aggregation on parameter estimates of different distributional assumptions of the parametric baseline hazard in continuous duration models. We follow Sueyoshi's (1995) methodology and focus on the sensitivity of parameter estimates in grouped duration models. We add to this literature by addressing the sensitivity of the estimates to different specifications of the cumulative density functions (CDF) of the discrete hazard, like the logistic, normal, and type I extreme value (cloglog), to time aggregation.

We use the research of Cramer, Hartog, Jonker and Van Praag (2002), on the effect of empirical relevant variables on the probability of entrepreneurship in the Netherlands in a retrospective survey of labor market histories between 1953 and 1993 of approximately 1700 individuals, as an example of a study using dynamic data and a static estimation technique. We examine and test the implicit assumptions of their static estimator and its empirical consequences in terms of precision and predictive accuracy, and show the information that can be gained by employing dynamic estimators.

We apply a variety of dynamic estimation techniques. For grouped duration models, we use semi-parametric estimation techniques and compare it to alternative approaches. We follow Prentice and Gloeckler (1978), Han and Hausman (1990), and Meyer (1990) by non-parametrically estimating the effect of time on the probability of entrepreneurship. We apply this technique to different functional forms of the discrete hazard of which the cloglog specification is the discrete variant of the continuous proportional hazard model. Lancaster (1979) and Heckman and Singer (1984) among others, discuss the importance of allowing for unobservable differences between individuals (unobserved heterogeneity). Unobserved heterogeneity may stem from omitted variables, but also from misspecification of the parametric baseline hazard. We allow and test for this unobserved heterogeneity in this dataset. We also address another possible source of unobserved heterogeneity in the proportional hazards model: time-varying coefficients¹³. Non-proportional hazards may result from an interaction between time and the regressors. We test for non-proportional hazards in a Cox partial likelihood framework (Cox 1972). Deviations from proportionality in covariates can be estimated with flexible spline models (Royston and Parmar 2002). This method of estimation leads to interesting conclusions in our dataset. We find that the positive effect of the self-employed parent, one of the most consistent predictors of entrepreneurship selection within the self-employment literature, is only significant within the first decade after compulsory schooling.

The structure of the paper is as follows. Section 3.2 addresses the process that generates the data. It presents the information in the data and models the transition to entrepreneurship as a stochastic point process. Section 3.3 derives the grouped duration models that will be estimated. Section 3.4 discusses the dataset used by Cramer et al. (2002). Section 3.5 presents the results. Section 3.6 concludes.

¹³ See Hastie and Tibshirani (1993) for a good reference on time-varying coefficients models.

3.2 The process that generates the data

3.2.1 The information in the data

After compulsory schooling, individuals face several career options: continue schooling, obtain social security, work in the labor market, or become self-employed, which we label entrepreneurship. What are the chances that an individual becomes an entrepreneur after compulsory schooling? What are the factors influencing the chance to become an entrepreneur? To answer these research questions, we have to research the transitions into entrepreneurship. The process of movement from a non-entrepreneurship career to entrepreneurship generates a sequence of points over time – the times at which a transition to entrepreneurship is made. Figure 3.1 provides the first transition times to entrepreneurship for 1981 compulsory school leavers from the North of Brabant in 1953.

Figure 3.1: Distribution of entry years into self-employment

Figure 3.1 shows that utilizing a static approach by aggregating the time dimension and dichotomizing the dependent variable (1 if individual became self-employed between 1953 and 1993, 0 otherwise) is both arbitrary and wastes information. It is arbitrary because there is nothing special about the 40-year dividing line except that the study ended at that point. It wastes information because it ignores the variation on either side of the dividing line. One might suspect, for example, that someone started as an entrepreneur immediately after compulsory school had a different initiate motivation to become entrepreneur than someone who started at the age of 50.

Moreover, static models assume that the probability of becoming self-employed is exponentially distributed and Markov independent or memoryless. This means that the probability of becoming an entrepreneur in any period after time t is always the same, conditional on everything that happens up to time t , but not on t itself. Figure 3.1 shows that this assumption is at least suspicious. Markov independence is violated in the pres-

ence of duration dependence: the probability of becoming self-employed changes over time.

The Poisson process is a stochastic point process where the probability to become entrepreneur is exponentially distributed. Since the static models assume Markov independence, we will start with the Poisson process. The next section shows that these static models are just censored and aggregated versions of a dynamic hazard model. It shows that the probability that someone becomes ever self-employed (static model) in life is a function of the rate of event occurrence (hazard rate) and time.

3.2.2 Modeling the transition to entrepreneurship as a Poisson process

What is it that generates the different entry years into entrepreneurship? We begin by describing the data which constitute our information about the transfers towards self-employment. Let y_{it} be a variable representing binary data, where

$$(3.1) \quad y_{it} = \begin{cases} 1 & \text{if a transition to self - employment occurs during } t \\ 0 & \text{otherwise} \end{cases}$$

This is the most disaggregated form of data used and enables one to model explicitly both the cross-sectional and time-series processes underlying the data. The only information loss is when the event occurs during the time period t .

Listing the duration between the state of never self-employed to ever self-employed is a different way of coding these events. Duration may be coded as a discrete or continuous variable. Since we have only yearly information we use a discrete variable, where we count the number of time units (years) t between the events. We denote this variable y_i :

$$(3.2) \quad y_i = \sum_{t=1}^{T_i} 1 = T_i$$

The only information lost by this coding of the transition is the precision of when within a year the transition occurred.

Given Markov independence, we can express the static approach in terms of the hazard rate: the rate of event occurrence. The hazard rate is better understood as the expected number of events in a time interval that is one unit long. Let y_{it}^* be the true number of times an individual becomes self-employed. The binary-censored static data are then obtained by transforming y_{it}^* into

$$(3.3) \quad y_{it} = \begin{cases} 0 & \text{if } y_{it}^* = 0 \\ 1 & \text{if } y_{it}^* > 0 \end{cases}$$

The distribution of the data y_{it} is then given by

$$(3.4) \quad \Pr[Y_{it} = y_{it} | y_{it}^*] = \begin{cases} \Pr[y_{it} = 0 | y_{it}^* \leq 0] \\ \Pr[y_{it} = 1 | y_{it}^* > 0] \end{cases}$$

Since y_{it}^* is distributed Poisson with mean λ (Lancaster 1990; Ross 1993), it follows that for some interval Δt (Alt et al. 2000):

$$(3.5) \quad \begin{aligned} \Pr[y_{it} = 0] &= \exp(-\lambda \Delta t) \\ \Pr[y_{it} = 1] &= 1 - \exp(-\lambda \Delta t) \end{aligned}$$

or equivalently

$$(3.6) \quad f(Y_{it} = y_{it} | \lambda) = (1 - \exp(-\lambda \Delta t))^{y_{it}} (\exp(-\lambda \Delta t))^{1-y_{it}}$$

Equation (3.6) gives the relation between the probability of an event occurring and the rate of event occurrence λ . Let λ , the rate of event occurrence, be a log-linear function of a set of explanatory variables X

$$(3.7) \quad \lambda = \exp(X\beta)$$

Filling in equation (3.7) into (3.6) gives

$$(3.8) \quad f(Y_{it} = y_{it} | X, \beta) = (1 - \exp(-\exp(X\beta)\Delta t))^{y_{it}} (\exp(-\exp(X\beta)\Delta t))^{1-y_{it}}$$

Which is a simple binary choice model with a type I extreme value error distribution (also known as the cloglog distribution) (Alt et al. 2000). Note that this static estimator becomes a dynamic estimator if the time interval is split into smaller time units.

Cramer et al. (2002) assumed a normal distribution of the error term. More importantly, they aggregated 40 years into 1 interval. Petersen (1991) shows that the larger the time interval the higher the bias of the aggregation. Shumway (2001) provides an econometric specification of the aggregation bias. The static approach does not take into account the number of periods at risk. For the static approach it does not matter whether one becomes self-employed after 1 or 40 years. Shumway (2001) shows that the estimated parameters of a static approach are always upward biased. Furthermore, some personal characteristics, like age, vary through time and their effect on the probability of entrepreneurship may also vary through time. These so-called time-varying covariates and time-varying coefficients cannot be included in static estimators, but are easily handled with a dynamic estimator. What should be evidently clear from this section is that the static estimator is only a censored (aggregated) version of the dynamic estimator. The only difference is that the dynamic estimator utilizes all information available in the data. Both estimators answer who becomes an entrepreneur, however the dynamic estimator also provides information about when that individual becomes an entrepreneur.

3.3 Grouped duration models

In this section, we discuss the use of duration (dynamic) models as a more flexible and robust alternative to static models. Since the underlying data generating process is continuous, we model durations first in continuous time (Heckman and Singer 1984, Lancaster 1990). Since our data is grouped into years, we derive the grouped (discrete) analogue of the continuous duration model and show that the discrete hazard model is equal to a dynamic binary choice model (Allison 1982; Jenkins 1995; Singer and Willett 1993; Sueyoshi 1995). Then we discuss the implications of different specifications of the functional form of the discrete hazard. Next, we discuss the discrete way to handle deviations from Markov independence by allowing non-parametric duration dependence. Besides duration dependence, the discrete hazard model allows for testing for unobserved heterogeneity. Then, we deal with the implicit proportionality assumption in discrete hazard models and offer a flexible alternative based on splines to deal with non-proportional hazards (Royston and Parmar 2002).

3.3.1 Duration models

We start with the information of equation (3.2) to derive a more flexible duration model. Let T be a positive, continuous random variable for the time to entry into entrepreneurship after compulsory schooling. The hazard rate, or the rate of event occurrence, at time t is defined as the conditional probability of entry into self-employment, given that the individual has not been self-employed before t (survival up to t):

$$(3.9) \quad \lambda(t) = \lim_{\partial t \rightarrow 0} \frac{\Pr(t \leq T < t + \partial t \mid T \geq t)}{\partial t}$$

The next step is to allow for population heterogeneity. If we allow the hazard to differ in terms of observed covariates X with associated parameters β , then (3.9) can be written as:

$$(3.10) \quad \lambda(t, X, \beta) = \lim_{\partial t \rightarrow 0} \frac{\Pr(t \leq T < t + \partial t \mid T \geq t, X, \beta)}{\partial t}$$

The associated survivor function is equal to the integrated hazard:

$$(3.11) \quad S(t, X, \beta) = \Pr(T \geq t) = \exp \left[- \int_0^t \lambda(s, X, \beta) ds \right]$$

The associated probability density function of the survivor function yields:

$$(3.12) \quad f(t, X, \beta) = \lambda(t, X, \beta) S(t, X, \beta)$$

In our empirical example of entrepreneurs, the continuous realisations of T are grouped in years, and defined as durations t_j , for $j=1, \dots, J$. Survival to time t_j is the same as surviving each of the intervals $[t_{k-1}, t_k)$ for $k=1, \dots, j$. Therefore, we can express the overall survivor function in terms of interval-specific conditional survivor functions α defined as (Sueyoshi 1995):

$$(3.13) \quad \alpha_k(X, \beta) = S(t_k, X, \beta | T \geq t_{k-1}) = \exp \left[- \int_{t_{k-1}}^{t_k} \lambda(s, X, \beta) ds \right]$$

By the definition of conditional probabilities, the survivor function at arbitrary duration t_j , can be expressed as:

$$(3.14) \quad S(t_j, X, \beta) = \prod_{k=1}^j \alpha_k(X, \beta)$$

The likelihood contributions for grouped data are based solely on the survivor function (or equivalently on its mirror function, the Failure function) $S(t)$ and hence upon α_j evaluated at various t_j . Note that the probability of an observed exit in the j -th interval is given by:

$$(3.15) \quad S(t_{j-1}, X, \beta) - S(t_j, X, \beta) = [1 - \alpha_j(X, \beta)] \prod_{k=1}^{j-1} \alpha_k(X, \beta)$$

where the left-part of the equation is the probability of surviving the first $j-1$ intervals, but not surviving the j -th (again following the rules of conditional probability). Also note that $1 - \alpha_j$ is the interval specific failure function.

We can rewrite equation (3.15) to arrive at the probability of failure (entry into self-employment) in the j -th interval:

$$(3.16)$$

$$\Pr(t_{j-1} \leq T < t_j | X, \beta) = S(t_{j-1}, X, \beta) - S(t_j, X, \beta) = F(t_j, X, \beta) \prod_{k=1}^{j-1} [1 - F(t_k, X, \beta)]$$

Equation (3.16) implies that a discrete (grouped) duration model is equivalent to a sequential discrete choice model with error distribution functions F (Sueyoshi 1995). Note that the likelihood function depends on the specification of F . Equation (3.16) equals equation (3.8) in the case of a cloglog specification of F .

3.3.2 Different specifications for the failure function

Although applying the type I extreme value link to equation (3.8) and (3.16) is convenient and leads to the easy interpretable proportional hazards model, other specifications for F are possible (Jenkins 1995, Royston and Parmar 2002, Sueyoshi 1995). Since the within behaviour of the hazard cannot be identified from grouped duration data,

evaluation of the alternative duration specifications should focus on the implications of various functional forms of the discrete hazard specification:

$$(3.17) \quad \Psi_j = 1 - \alpha_j(X, \beta) = F(X\beta)$$

In contrast with the continuous hazard rate, the discrete hazard rate is a probability. We investigate the three most frequently used functional forms: the normal, logistic, and type I extreme value. We are particularly interested in how the estimated parameters differ in interpretation and how different specifications lead to different estimated durations and probabilities of entry into self-employment (goodness-of-fit). Moreover, we test for the effect of using different specifications of the discrete hazard on the bias from aggregating time.

The previous sections have shown that the type I extreme value CDF is the discrete analogue to the proportional hazards model. A one unit change in the covariate leads to a proportionate change in the hazard. The logistic specification leads to a proportional logit model. A one unit change leads to a proportional effect in the log of the odds. The normal specification leads to a less well-known specification and non-proportional effect. Sueyoshi (1995: 419) has shown that a one unit change in a covariate leads to a disproportionate effect because the effect also depends on time.

3.3.3 Duration dependence

So far, we have assumed that the probability of becoming self-employed is stable over time. This has been the assumption in most of the empirical studies of entrepreneurship/self-employment selection. However, these studies did not study explicitly the duration to become ever entrepreneur, but more the chance that someone becomes an entrepreneur in some arbitrary time interval. None of these studies followed school leavers over time (inflow sample). Rees and Shah (1986) found that the probability of being an entrepreneur is convex in age in the UK. However, Evans and Leighton (1989) found in the US that the probability of entering self-employment is independent of age. Still, there are various reasons why the probability to become self-employed may differ over time. In the first years after compulsory school, individuals need to learn the skills or earn the necessary financial capital to start their own business (Evans and Jovanovic 1989). When learning these skills and earning a threshold level of financial capital, an individual's probability to become self-employed may be lower than after this first period. At the end of an individual's working career, the probability to become self-employed may level off again. Starting a business is an investment, and the time to capitalize on the investment decreases. In these hypothetical cases, the chance of becoming self-employed is clearly duration dependent.

The most common way to introduce duration dependence is to assume a proportional hazards model of the form:

$$(3.18) \quad \lambda(t) = \lambda_0(t) \exp(X\beta)$$

where $\lambda_0(t)$ is the baseline hazard at t , which can take several forms and is possibly unknown. From equation (3.11) we know that we can write the probability that a spell lasts until time t , given that it has lasted until $t-1$ as a function of the hazard:

$$(3.19a) \Pr(T > t_k | T \geq t_{k-1}) = S(t_k, X, \beta | T \geq t_{k-1}) = \exp \left[- \int_{t_{k-1}}^{t_k} \lambda(s, X, \beta) ds \right]$$

equals

$$(3.19b) \Pr(T > t_k | T \geq t_{k-1}) = \exp \left[- \exp(X, \beta) \int_{t_{k-1}}^{t_k} \lambda_0(s) ds \right]$$

which can be written as:

$$(3.19c) \Pr(T > t_k | T \geq t_{k-1}) = \exp[-\exp(X, \beta) + \gamma(t)]$$

where

$$(3.20) \quad \gamma(t) = \ln \int_{t_{k-1}}^{t_k} \lambda_0(s) ds$$

Equation (3.20) is equivalent to extending equation (3.16) with time dummies and estimating the sequential discrete choice model with a type I extreme value CDF.

The log-likelihood contribution for the whole sample of N individuals can be written in terms of equation (3.19c) following the result of equation (3.16) as:

$$(3.21) \quad \log L(\gamma, \beta) = \sum_{i=1}^N \left[\partial_i \log[1 - \exp\{-\exp[\gamma(k_i) + \beta X]\}] - \sum_{i=1}^{k_i-1} \exp[\gamma(k_i) + \beta X] \right]$$

where $\partial_i = 1$ if T_i is smaller or equal than the censoring time and $\gamma = [\gamma(0)\gamma(1)..\gamma(T)]$, a set of time dummies capturing the baseline hazard (semi-parametric) (Prentice and Gloeckler 1978). Based on this specification of the likelihood we can easily test for duration dependence with a likelihood ratio test. Duration dependence is existent if the grouping as a whole is significant. In this proportional hazards specification, one might also capture the duration dependence parametrically, by assuming a parametric distribution like the Weibull or Gompertz, for the baseline hazard. We will return to parametric specifications of the baseline hazard, when we discuss non-proportional effects.

3.3.4 Unobserved heterogeneity

Duration dependence may also be a result of unobserved heterogeneity (Elbers and Ridder 1982). Unobserved heterogeneity affects duration dependence by a so-called weeding out effect. The most entrepreneurship prone individuals tend to enter self-employment faster. Therefore, after controlling for unobserved heterogeneity, the baseline hazard may become smaller. Unobserved heterogeneity has been found more important in parametric formulations than in non-parametric (Jenkins 2004: 87). Wrongly specified parametric distributions inflate unobserved heterogeneity. In our semi-parametric discrete (grouped) duration model we control for unobserved heterogeneity. Following Meyer (1990), we assume that the unobserved heterogeneity affects the hazard multiplicatively. Controlling for unobserved equation (3.18) can be written as:

$$(3.22) \quad \lambda_i(t) = \theta_i \lambda_0(t) \exp(X_i \beta)$$

where θ is a random variable and is assumed to be independent of X .

The resulting log-likelihood is achieved by integrating over its distribution (Meyer 1990: 770).

(3.23)

$$\log L(\gamma, \beta, \mu) = \sum_{i=1}^N \log \left[-\partial_i \int \exp \left[-\theta \sum_{t=0}^{k_i} \exp[\gamma(k_i) + \beta X] d\mu(\theta) \right] + \int \exp \left[-\theta \sum_{t=1}^{k_i-1} \exp[\gamma(k_i) + \beta X] d\mu(\theta) \right] \right]$$

The significance of unobserved heterogeneity can be tested by a likelihood-ratio test under the null-hypothesis of no unobserved heterogeneity.

3.3.5 The proportionality assumption

With the exception of the normal specification, the dynamic discrete choice models based on the type I extreme value and the logistic specification assume a proportional relationship between the predictors and the hazard (Jenkins 1995; Singer and Willett 1993; Sueyoshi 1995). Different values for the predictors shift the baseline hazard. Sometimes, the predictors do not only shift the baseline hazard, but they may also alter the baseline hazard. For instance, the influence of parents on the probability of becoming self-employed may be higher in the early ages of an individual, when the person is more dependent on parental financial assistance and advice. In this case, the effect of self-employed parents may increase the hazard of becoming self-employed less when the individual ages. If the predictor also changes the shape of the hazard, then there is a non-proportional relationship between predictors and the hazard. Sueyoshi's suggests (1995) that more general models should allow for interactions between the time-dependence in the hazards and the explanatory variables. Non-proportional relationships can be included in the model by exploring statistical interactions between the predictors and the hazard.

We test for proportional hazards. We use Stata's specification test of the proportionality assumption based on Schoenfeld (1982) residuals, which tests for a nonzero slope of the scaled Schoenfeld residuals on functions of time (Grambsch and Therneau

1994). Under the null hypothesis of proportional hazards, the parameter estimates modelled as a function of time should have a zero slope. A rejection of the null hypothesis indicates non-proportional effects.

A rather ad-hoc approach in a discrete time setting to deal with non-proportional hazards is to include an interaction between each time-dummy and the covariates (Singer & Willett 1993). This approach can be implemented easily in a discrete time formulation, but at the cost of a substantial reduction in the degrees of freedom. A more flexible approach to allow for non-proportional hazards rests on parametric splines (Royston & Parmar 2002). The parametric splines model extends the non-parametric approach based on time dummies with a parametric baseline hazard based on splines. More important, the splines can be interacted with the coefficients of the non-proportional predictors leading to parametric time-varying coefficients for the predictors. More general models with time-varying coefficients are advocated by the seminal paper of Hastie and Tibshirani (1993). They have been applied in the self-employment exit literature by Kauermann, Tutz, and Brüderl (2005). If we test for non-proportional hazards, then we can estimate the following flexible parametric spline model based on one to six splines and on three specifications of the error distribution: logistic, normal, type I extreme value. In the case of only minor deviations of the proportional hazard assumption, Royston and Parmar (2002) advice to follow the proportional hazards assumption. The parametric spline model with possible time dependent covariate effects is based on the integrated proportional hazards model and formulated as follows (Parmar & Royston 2002: 2184):

$$(3.24) \quad \ln H(t; X) = \gamma_0 + \gamma^T v(x) + \beta^T X$$

where $v(z) = (z, v_1(z), \dots, v_m(z))^T$, $\dim(\gamma) = m+1$ and the j -th component of γ ($1 \leq j \leq m+1$) is

$$(3.25) \quad \gamma_j = \begin{cases} \gamma_{j0} & k = 0 \text{ (proportional hazards model)} \\ \gamma_{j0} + \sum_{l=1}^k \gamma_{jl} x_l & k \geq 1 \text{ (non - proportional hazards for } x_1, \dots, x_k) \end{cases}$$

where $z = \ln(t)$ and $H(t; X)$ is the integrated hazard function.

3.4 An example of a study using dynamic data and a static estimation technique

For our analysis we use one of the models estimated on the North Brabant Dataset in Cramer et al. (2002). The North Brabant longitudinal data set contains information about 3167 individuals who were born 1939/1940 and went to school in the Dutch province North Brabant in 1952 (Hartog 1989; Jonker 1995). Then researchers took a random sample of pupils who attended the sixth grade of primary school. 2874 individuals provided information about their school performance and family background. In 1983 all 3167 individuals were contacted again. 2588 people responded and were asked about their later education, labor market status, earnings, and household composition. In 1993

at the age of approximately 54, the 3167 school children of 1952 were approached for the third time. The questionnaire mailed in 1993 contained questions on the labor market status since 1983 and especially on self-employment experience. 2099 individuals answered in 1993. Due to attrition and item non-response the final sample consists of 1657 individuals.

Cramer et al. (2002) applied a static estimator to multi-period data. They modeled the choice for entrepreneurship as respondents report self-employment at any time between 1953 and 1993. We are interested in the exact timing of this choice. In the original dataset, each respondent has one record of data. Before we can apply our dynamic estimation techniques, this person-oriented dataset must be converted into a person-period dataset in which every person has multiple records, one per year.

For the explanatory variables we use Cramer et al.'s second model. The explanatory variables included are intelligence, gender, vocational schooling, self-employed parent, education of parent, and transformed price. The transformed price is a measure of risk aversion extracted from a lottery question in the survey.

3.5 Data analysis with dynamic techniques

Table 3.1 shows the estimation results. Model 1 – 3 show the results of the binary choice models. Model 4 – 6 provide the results of the grouped duration data. All models assume a constant probability of becoming entrepreneur over time (Markov independence). The differences in results stem only from accounting for the variability over time. Model 7 – 9 show the results of the semi-parametric grouped duration models when accounted for non-parametric duration dependence by including time dummies. And model 10 shows the result of a grouped duration model with time dummies accounting for unobserved heterogeneity by gamma frailty¹⁴. The number of observations differs between the models. Model 1-3 have 1657 observations, the number of individuals in our sample. Model 4-6 have 62035, the number of person-period observations. Model 7-10 have only 58729 person-period observations. In period 3 and 4 there were no transfers to self-employment. The two periods (13304 person-period observations) were dropped because the outcome was perfectly predicted (multicollinearity). In order to compare the coefficients of the probit, logit, and cloglog, we have to take into account the different variances of these distributions, being respectively $\sigma=1$, $\sigma=\pi/\sqrt{3}$, and $\sigma=\pi/\sqrt{6}$. Therefore we standardize the coefficients with respect to probit estimates by dividing the logit and cloglog estimates by their respective variances.

Table 3.2 compares the within sample predictive accuracy of the models described above. To construct the table, we sort all observations (persons or person-periods) into deciles based on their predicted probability values¹⁵. The table reports the percentage of entrepreneurs that are classified into each of the five highest probability deciles. It also lists the percentage of entrepreneurs classified among the least 50% of individuals (observations) to fail. There are 260 entrepreneurs to evaluate the prediction accuracy.

¹⁴ This model is estimated by PGMHAZ8 (Jenkins 2004)

¹⁵ Shumway (2001) uses a similar approach to assess the out-of-sample forecasting accuracy of static and dynamic models.

Table 3.1: Estimation Results

	Static Models						Dynamic Models			
	Duration dependence: none						Duration dependence: non-parametric			
							Unobserved heterogeneity: none		Unobserved heterogeneity: gamma frailty	
	Probit	Logit	Cloglog	Dynamic Probit	Proportional Odds	Proportional Hazard	Dynamic Probit	Proportional Odds	Proportional Hazard	Proportional Hazard
Intelligence	0.700 (0.365)-	0.652 (0.360)-	0.741 (0.438)-	0.358 (0.193)-	0.525 (0.300)-	0.738 (0.423)-	0.377 (0.194)-	0.521 (0.299)-	0.728 (0.420)-	0.861 (0.482)-
Gender	-0.481 (0.087)**	-0.509 (0.091)**	-0.660 (0.119)**	-0.286 (0.053)**	-0.477 (0.088)**	-0.672 (0.124)**	-0.294 (0.053)**	-0.472 (0.087)**	-0.665 (0.123)**	-0.624 (0.121)**
Vocational schooling	0.234 (0.103)*	0.203 (0.101)*	0.235 (0.122)-	0.121 (0.058)*	0.177 (0.090)-	0.249 (0.117)-	0.124 (0.059)*	0.189 (0.090)*	0.253 (0.127)*	0.316 (0.144)*
Father independent	0.692 (0.081)**	0.696 (0.081)**	0.883 (0.102)**	0.399 (0.046)**	0.644 (0.074)**	0.909 (0.104)**	0.403 (0.046)**	0.644 (0.073)**	0.907 (0.103)**	0.940 (0.156)**
Father educated	0.513 (0.205)*	0.537 (0.197)**	0.706 (0.243)**	0.281 (0.103)**	0.471 (0.162)**	0.665 (0.229)**	0.287 (0.102)**	0.458 (0.158)**	0.644 (0.222)**	0.454 (0.246)-
Transformed price	-0.315 (0.092)**	-0.290 (0.087)**	-0.323 (0.094)**	-0.148 (0.043)**	-0.216 (0.052)**	-0.304 (0.087)**	-0.156 (0.043)**	-0.215 (0.061)**	-0.300 (0.085)**	-0.325 (0.122)**
Constant	-1.703 (0.394)**	-1.600 (0.389)**	-2.185 (0.473)**	-3.018 (0.205)**	-3.601 (0.319)**	-5.091 (0.449)**				
Gamma										0.596 (.342)-
LR: gamma=0										4.198*
Observations	1657	1657	1657	62035	62035	62035	58729	58729	58729	58729
Parameters	7	7	7	7	7	7	45	45	45	46
Likelihood	-654.26	-653.97	-654.28	-1614.82	-1614.83	-1614.84	-1537.07	-1535.89	-1535.85	-1533.75
AIC	1322.51	1321.94	1322.57	3243.64	3243.67	3243.68	3164.15	3161.78	3161.70	3159.51

- significant at 10%; * significant at 5%; ** significant at 1%.

Table 3.2: Prediction Accuracy

Static Model:		Dynamic Model:							
		Duration dependence: none		Duration dependence: non-parametric					
		Unobserved heterogeneity: none		Unobserved heterogeneity: gamma fully					
Probability ranking vs. Actual entrants: number classified out of 260 possible (ranking based on person-period observations):									
Date	Probs	Logit	Cloglog	Dynamic Probit	Proportional Hazard	Proportional Odds	Dynamic Probit	Proportional Hazard	Proportional Odds
1	58	57	57	64	64	64	104	111	72
2	62	61	61	70	70	70	43	37	61
3	55	56	56	26	25	25	23	21	22
4	26	25	25	22	24	24	26	23	22
5	19	18	19	22	21	21	21	21	27
6-10	63	63	62	36	36	36	43	41	47

Model 1 is the same model as the second model of Table 3 in Cramer et al. (2002: 35). Although we were not able to replicate their results entirely, we do find similar effects and come close in significance levels. Note that Cramer et al. provide elasticities while we provide coefficients.

Model 4 equals model 1, but respects the time dimension of the data. Differences are striking. Conform to the theoretical predictions (Shumway 2001); applying a static methodology to dynamic data, not accounting for periods at risk leads to overestimating the parameters of interest. This result corroborate our prediction that applying a static methodology to multi-period data leads to biased and inconsistent parameter estimates. Within the probit framework, the parameter estimates of the static methodology are almost twice as large as the estimates of the dynamic methodology. The dynamic models (4 – 6) are also consistently more accurate than the static models. Table 3.2 shows that the dynamic probit model classifies almost 52% of all entrepreneurs in the highest two deciles, compared with 45% of the static model. It also classifies less entrepreneurs below the median probability.

In line with standard knowledge of static binary choice models, the goodness of fit is not greatly affected by applying a different functional form in the static methodology. Based on the log-likelihood and the AIC, we see that the logit model performs only marginally better than the probit, which in turn performs marginally better than the cloglog. Goodness of fit among the dynamic models (model 4 – 6) are also very similar. Interestingly, the parameter estimates of the logit and the cloglog model are less sensitive for neglecting time variation. With the exception of the constant, the results of the cloglog model are remarkably stable between the dynamic and the static methodology. Especially the parameter estimates of the cloglog model seem insensitive to time aggregation.

This (in)sensitivity of the parameter estimates to time aggregation can be theoretically explained by the difference in exit probability of the static and the dynamic models and the behavior of the derivatives of the log discrete hazard function for different distributional assumptions. The expected empirical exit probability reduces strongly between the static and the dynamic approach: from 15.69% (260/1657) in the static models to 0.42% (260/62035) in the dynamic models. Sueyoshi (1995: 421) has shown that a change in an explanatory variable cause a proportionate change in the discrete hazard of the form:

$$(3.26) \quad \frac{\partial \log \Psi(X\beta)}{\partial X_m} = \frac{\partial \log F(X\beta)}{\partial X_m} = \beta_m \frac{f(X\beta)}{F(X\beta)}$$

using the fact that $\partial \ln[g(x)]/\partial x = g'(x)/g(x)$.

In order to see the cause and the effect of time aggregation on the parameter estimates, we plot the derivatives of the log cumulative distribution functions, $f(X\beta)/F(X\beta)$, for the probit, logit, and cloglog specifications against various values of $F(X\beta)$. Given that X is stable in the static and dynamic approach, the change in β must come from the sensitivity of $f(X\beta)/F(X\beta)$ against the change in F . Figure 3.2 plots the derivatives of the log discrete hazard function for the various specifications at values of

the empirical exit probabilities $F(X\beta)^{16}$. Figure 3.2 shows that the probit specification is very sensitive for exit probabilities lower than 0.5. Note that in our empirical example of time aggregation, the exit probability reduces from 0.157 to 0.004. According to Figure 3.2, we would expect that time aggregation results in a sharp change in the parameter estimates of the probit model and hardly any change in the cloglog model and minor changes in the logit model. Table 3.1 confirms these predictions. Parameter estimates of the static probit model are twice as large as the dynamic probit model, while the estimates of the cloglog remain stable. The effects of time aggregation on the parameter estimates differ significantly between the various specifications of the cumulative density functions (discrete hazard).

Figure 3.2: Derivatives of the log hazard function, $f(z)/F(z)$, evaluated at values of the hazard probabilities $F(z)$; various specifications for the discrete hazard function F

Models 7-9 allow non-parametrically for duration dependence. Duration dependence significantly increases the fit of the model. Duration dependence is significant for all specifications being: model 4 versus 7 ($\chi^2(38)=156$ with a p-value of 0.000), model 5 versus 8 ($\chi^2(38)=158$ with a p-value of 0.000), model 5 versus 9 ($\chi^2(38)=158$ with a p-value of 0.000). Figure 3.3 plots the non-parametric baseline hazard. The baseline hazard can be easily retrieved in discrete setting from substituting the estimated time dummies back into their respective distribution functions. We observe clear violations of Markov independence.

¹⁶ Sueyoshi (1995: 422) provides a similar figure to discuss the behavior of the different specifications of the log discrete hazard. He does not mention time aggregation.

Figure 3.3: The non-parametric baseline discrete hazard; various specifications for the discrete hazard function F

The models with duration dependence have a better fit than the models without duration dependence. Also the predictive accuracy of the models with duration dependence is strikingly better. The models with duration dependence classify more entrepreneurs in the highest decile and less individuals below the median probability.

For the models with duration dependence (model 6-9), Table 3.1 shows that the proportional odds and the proportional hazard model perform better than the dynamic probit. Although the AIC and likelihood criteria can hardly discriminate between the proportional hazard and proportional odds model, the predictive accuracy of entrepreneurship is highest for the proportional odds model. According to Table 3.1, the proportional hazards model performs marginally better than the proportional logit. However according to Table 3.2, the proportional logit model predicts entrepreneurship more accurately.

Next, we consider the interpretation of the covariates for the different distributional assumptions and model specifications. Parameter estimates in non-linear models do not have a clear interpretation. This problem is enhanced through the dynamics in the duration models. We use the conventional approach to consider the behaviour of the models evaluated at mean values of the explanatory variables. The discrete hazard is computed by evaluating the appropriate cumulative distribution function at the sum of an aggregator representing $X\beta$ and, for the static and the Markov model, the constant. Note that both models assume independence of the discrete hazard with respect to time. The discrete hazard for the models with duration dependence (7-9) is computed for each period by evaluating the CDF at the sum of the aggregator and the period-specific intercept term.

Figure 3.4: Proportionate change in discrete hazard induced by the self-employed Father: various specifications for F

In Figure 3.4 we analyse the change in the log discrete hazard associated with our strongest predictor, the self-employed father. The discrete change is computed as the difference in F when the indicator is first set to 0 and then to 1 (keeping the aggregator constant at its mean), and evaluating the proportionate change in F : $F(1)/F(0) - 1$ (Sueyoshi 1995). Figure 3.4 shows the results of this exercise.

We first consider the effects of time aggregation. For all specifications, Figure 3.4 shows that the static models underestimate the effect of the self-employed father on the discrete hazard. For example, the marginal effect of the self-employed father on the hazard increases the discrete self-employment hazard by around 190 % for the static models, but almost 230 % for the dynamic Markov models. This result contradicts the findings in Table 3.1. The results in Table 3.1 showed that the parameter estimates for the static logit and probit decreased drastically, and the cloglog remained almost the same. Unlike the estimated parameters, the behaviour of the discrete hazard is extremely consistent for different specifications of the CDF (a result of Markov Independence).

The consistency between the various specifications disappears when we deviate from Markov independence and allow for duration dependence. Figure 3.4(d) plots the

effect of the self-employed father on the discrete hazard for various CDF. The logit and the cloglog specifications show clear proportionality. However, the probit specification deviates from proportionality. For example, after 7 years after compulsory school, the self-employed father increases the self-employment hazard by almost 280% and after 12 years only by 180%. We conclude from these results, that the various CDF perform equally well under Markov Independence. However, when the data inhibits duration dependence, the probit specification deviates strongly from proportionality. The likelihood criterion for the models with duration dependence indicates that a model with proportional hazards performs better than a model without proportional hazards. However, a more formal test for proportionality is conducted later. Before we discuss the proportionality assumption, we discuss unobserved heterogeneity.

Model 10 in Table 3.1 provides the Prentice and Gloeckler (1978) model when adjusting for unobserved heterogeneity as in Meyer (1990) estimated by PGMHAZ8 (Jenkins 2004). The results show that we cannot reject the null hypothesis of no unobserved heterogeneity. The likelihood ratio test and the AIC indicate that model 10 is the best fitting model. Unfortunately, the goodness of fit is not reflected in the predictive accuracy of the model. Interpretation of the coefficients in models with unobserved heterogeneity is complicated. As expected from evidence with unobserved heterogeneity models (Jenkins 2004), most of the estimated parameters in the models with unobserved heterogeneity (model 10) are larger than the in the models without unobserved heterogeneity (model 9): a result of the so-called weeding out effect.

How does controlling for unobserved heterogeneity influence the effect of the self-employed father on the discrete hazard? Figure 3.5 shows that the model without unobserved heterogeneity underestimates the true effects of the self-employed father significantly¹⁷. The effect rises from 220% to 370%. Figure 3.5 shows clearly that the bias for not controlling for unobserved heterogeneity is more severe than the time-aggregation bias and the bias of not controlling for duration dependence.

Overall, the static models provide biased and inconsistent estimates. Reasons are threefold. First the static models do not account for the periods at risk. Cost is overestimating parameters of interest. Secondly, static models assume that there is no duration dependence. We test the null-hypothesis of no duration dependence and reject the assumption of no duration dependence. Thirdly, there is unobserved heterogeneity in the data. For prediction accuracy, the picture is less clear. Although dynamic models are superior, they differ in accuracy. A possible reason is that our goodness-of-fit criterion focuses on the predictive accuracy of entrepreneurship entry and not how well the models discriminate between entrants and non-entrants. Unfortunately, the calibration test statistics of static discrete choice models can not be easily transferred to dynamic discrete choice models.

Note that our best fitting models, the logit and cloglog specification assumes proportional hazards. Although our proportional models outperform our non-proportional model based on the probit specification, the assumption of proportional hazards may not be true. To test the proportionality assumption, we estimate a semi-parametric Cox model based on Partial Likelihood (Cox, 1972)¹⁸. This model treats time as a nuisance parameter and only concerns rankings. Table 3.3 presents the results of the test. The overall test of

¹⁷ Note, that the aggregator is evaluated at mean unobserved heterogeneity.

¹⁸ Results available on request.

proportional hazards is rejected. The main driver of this non-proportionality is the self-employed father and to a lesser extent the education of the father.

Figure 3.5: Proportionate change in discrete hazard induced by the self-employed father: cloglog specification for F

Table 3.3: Test of proportional hazards assumption (based on Cox PH model)

	rho	Chi2	df	Prob>Chi2
Intelligence	0.04934	0.55	1	0.4568
Gender	0.00604	0.01	1	0.9197
Vocational schooling	-0.07490	1.51	1	0.2187
Father independent	-0.19863	10.14	1	0.0014
Father educated	0.13386	4.72	1	0.0299
Transformed price	-0.10971	2.94	1	0.0863
Global test		27.58	6	0.0001

Most approaches to deal with non-proportionality rely on some ad-hoc methods, like interactions between time dummies and covariates. We use a more consistent approach based on splines (Royston and Parmar 2002). This approach combines the flexibility of functional forms in the link functions (logit, probit, or cloglog) of the grouped (discrete) approach with the flexibility of the continuous approach in the parametric baseline hazard.

We tested several spline models. We tested models without time-varying coefficients and spline models with time-varying coefficients for self-employed father and the

education of the father. All fitted models with only the self-employed father performed better than the fitted models with the self-employed father and the education of the father. Table 3.4 compares different specifications of the CDF for the spline with the self-employed father as time-varying coefficient. AIC criteria points toward a parametric spline model with five knots and a time-varying coefficient for the self-employed father.

Table 3.4: AIC values for several spline duration models

d.f.	No time-varying coefficients			Time-varying coefficient: self-employed father		
	PH	PO	PP	PH	PO	PP
1*	1800.04	1788.55	1767.90	1797.29	1787.89	1769.76
2	1708.85	1706.52	1722.33	1694.96	1696.88	1720.88
3	1710.96	1707.86	1709.43	1699.45	1700.71	1707.81
4	1705.24	1701.53	1698.65	1676.96	1677.07	1679.06
5	1692.29	1689.37	1689.35	1665.65	1667.41	1674.68
6	1692.79	1690.35	1691.93	1670.39	1672.49	1680.13

* PH model with d.f.=1 is the Weibull model, PO model with d.f.=1 is the log-logistic model

Table 3.5 shows the estimation results of this model. Note that this model, except for the self-employed father, provides almost identical results as the non-parametric proportional hazards model. Figure 3.6 provide the baseline hazard. If we compare the parametric model with the non-parametric approach, we do not find striking differences. We still observe a convex relationship between age and entrepreneurship.

Figure 3.6: The baseline hazard function

Table 3.5: Estimation results – Parametric spline model

		Spline 1	Spline 2	Spline 3	Spline 4	Spline 5
Intelligence	0.938 (0.558)+					
Gender	-0.842 (0.152)**					
Vocational schooling	0.320 (0.155)*					
Father independent	8.353 (3.743)*	-2.253 (1.202)+	-1.424 (0.786)+	0.614 (0.206)**	-0.150 (0.044)**	-0.013 (0.019)
Father educated	0.842 (0.310)**					
Transformed price	-0.389 (0.117)**					
Constant	-16.262 (3.734)**	4.086 (1.185)**	1.594 (0.780)*	-0.252 (0.199)	0.058 (0.038)	0.017 (0.017)
Observations	62035					
Parameters	17					
Log-likelihood	-815.83					
AIC	1665.65					

The difference stems from the time-varying effect of the self-employed father. This result is striking but not surprising. Figure 3.7 plots the time-varying coefficient as a function of time. The effect of the self-employed father is very strong initially and decreases in strength within the first ten years after compulsory schooling. After the first ten years, the effects of the self-employed father remains stable (a proportional effect). From a theoretical perspective the finding makes sense. Social capital from the family is more important in the early years. The young individual is more dependent on its parents, can learn from his self-employed parents, and the parents may serve as a role model. Time-varying coefficients may also be another reason for the contradicting results in the empirical literature on determinants of self-employment selection.

Finally, we would like to address some limitations. Besides the limitations addressed in Cramer et al (2002), an implicit limitation in their study becomes explicit in our dynamic modelling. From the data, it is not clear when the individuals finished their vocational schooling. The other variables are stable over the lifetime; however it is not certain when vocational schooling takes effect. In our approach, we considered it to take effect immediately after compulsory schooling. Since this is an arbitrary approach, we ran several models that assumed different times at which vocational schooling took effect. We did additional analysis with vocational schooling taking effect after 5 and 10

years. We transformed vocational schooling in a time-varying covariate. That did not influence the results significantly¹⁹. The effect of vocational schooling became a bit stronger and more significant, while leaving the other results almost unaltered.

Figure 3.7: The non-proportional effect of self-employed father over time

3.6 Conclusion

We have modeled the process if and when an individual becomes an entrepreneur. We have modeled the probability to become self-employed as a stochastic point process with an absorbing state (ever an entrepreneur). We have utilized the simple Poisson process to explain that the probability that an individual becomes ever self-employed (which can be estimated by a static method) is a function of the hazard rate, the rate of event occurrence, and time. The static estimator is merely an aggregated version of a dynamic estimator and therefore less efficient than our dynamic estimators. We have shown that if the process that generates the data is the continuous Poisson process and the data is grouped into intervals, then the natural candidate for an empirical dynamic model is a grouped duration model with a type I extreme value error distribution.

We have shown that modeling the probability that an individual becomes self-employed directly in terms of the hazard rate (continuous or discrete) offers several advantages over the static approach. The dynamic approach correctly specifies the periods at risk. The static approach may result in serious time aggregation, which leads to biased and inconsistent estimates. The static approach implicitly assumes that the probability of becoming self-employed is independent of time (Markov independence). The

¹⁹ Results available on request of the authors.

dynamic approach can handle easily deviations from this assumption by allowing non-parametrically or parametrically for duration dependence. Duration dependence may be a result of unobserved heterogeneity. Our dynamic approach can easily handle unobserved heterogeneity. Finally, the dynamic approach respects two possible extensions that influence the process that generates the data: time-varying covariates and time-varying coefficients. Time varying coefficients can be nicely handled in a parametric way, which allows clear interpretation of the results.

We have studied the implications on a study on self-employment selection in the Netherlands between 1953 and 1993. We have replicated the findings of a static approach (Cramer et al. 2002) and also redone the analysis by a dynamic approach. Although, we do confirm their main findings, we have shown the clear superiority of the dynamic model in terms of parameter estimate precision and prediction accuracy. We have shown that the self-employed father increases the probability to become self-employed in the dynamic model with 40 up to 200 percentage points. Next, we reject the implicit assumption that the probability to become self-employed is independent over time. We do find clear duration dependence. The probability to become self-employed rises at young ages but decrease at old ages. Also, our dynamic models are far more accurate in predicting entrepreneurship. The dynamic model with duration dependence classifies 95% more individuals in the highest probability decile than the static model. A dynamic approach may also include time-varying coefficients to deal with deviations from implicit assumptions of proportionality. In this data, we did find a time-varying effect of the self-employed father. Having a self-employed father seems especially important during the initial stage of an individual's entrepreneurial career.

Our results regarding time aggregation in grouped duration models can be easily generalized. We have shown the sensitivity of parameter estimates to different specifications of the distribution functions in the presence of time-aggregation. We have found that parameter estimates are not reliable to consider the effects of time aggregation. We propose that a plot of the effect of a predictor on the proportionate change in the discrete hazard evaluated at the mean of the aggregator $X\beta$ is a better way to assess the effect of time aggregation for different specifications of the error distribution.

Chapter 4

The dynamics of business succession and de novo self-employment: A multiple risks approach

4 The dynamics of business succession and de novo self-employment: A multiple risks approach²⁰

4.1 Introduction

Most empirical studies on self-employment focus on the determinants of self-employment selection or on how liquidity constraints inhibit new venture creation²¹. However, Carroll and Mosakowski (1987) and our study find that more than one third of self-employment is family self-employment and not de novo self-employment. More important, we find that people who take over a family business are different from those who found an entirely new business. While real founders enter self-employment with the burden of financing the sunkness of start up expenses, continuation of a family business is like entering self-employment with a valuable set of assets already existent. People who take over control in a family business are less constrained and may enter self-employment faster than real founders. In this paper, we build a theoretical model that explains the dynamics of family business succession and de novo self-employment. From this model we derive hypotheses that explain the timing of the entry decision for these alternative types of self-employment.

We believe that knowledge on the timing of the self-employment decision and its determinants are important for the following reasons. If self-employment and its derivative entrepreneurship are important for economic growth and innovation (e.g. Cosh et al. 1999; Geroski and Pomroy 1990) then it is not only desirable that individuals become self-employed but also that they become self-employed early in their careers. Self-employment experience is namely an important predictor of future self-employment (Carroll and Mosakowski 1987; Evans and Leighton 1989) and of self-employment success related to profits (Bosma et al. 2004) and related to survival (Holmes and Schmitz 1996; Quadrini 1999; Taylor 1999).

Carroll and Mosakowski (1987) indicate that any complete theory of the self-employment process should at least do two things. First, the modelling framework should take a dynamic perspective. A static approach cannot account for the fact that self-employment is episodic, individuals attributes change over time, and that the self-employment propensity may vary over the life-cycle. Second, the modelling framework should differentiate between the various ways in which one may become self-employed. The factors that account for each route into self-employment are likely to differ from each other. A survey of the theoretical and empirical literature until today indicates that these recommendations have not been incorporated.

²⁰ This chapter is based on the paper “The dynamics of business succession and de novo self-employment: A multiple risks approach” by Blumberg, Kok, and Pfann (2006).

²¹ For example, Carroll and Mosakowski (1987) investigate the determinants of transition rates into self-employment for West-Germany, Evans and Leighton (1989) examine the determinants and the process of selection into self-employment in the US, and Evans and Jovanovic (1989) build a structural model of how liquidity constraints affect the probability of becoming self-employed in the US.

Therefore, our model of the self-employment process advances the literature on self-employment in the following ways. (1) We formally link and test the theoretical concepts of human, financial and social capital, to the timing of the decision to become self-employed in a dynamic framework. (2) We differentiate between family self-employment and de novo self-employment. Unlike Dunn and Holtz-Eakin (2000), we can test for the hidden possibility that the existence of self-employed parents merely reflect family self-employment instead of the intergenerational transfer of skills or capital.

Previous research that investigated the dynamics of the self-employment decision found that the probability of self-employment is independent of age (Evans and Leighton 1989). Our duration analysis with time varying covariates over a time period of 40 years of a sample of Dutch individuals from the North of Brabant, a province in the Netherlands, shows that self-employment may indeed be independent of age for a considerable period of our time frame. However, this picture changes dramatically when we differentiate between family business succession and de novo self-employment. Our results indicate that the family business succession and de novo self-employment have opposing and offsetting dynamics and the age independence of self-employment is the result of aggregating these types of self-employment. The chances of taking over a family business are initially high and eventually decreasing, while the chances of starting up a new venture are initially low and eventually increasing. Our dynamic theory accounts for these differences.

Not only do the dynamics of self-employment differ between real founders and family business successors, also the determinants of self-employment differ between these self-employment types. Our results indicate that higher educated and possibly wealthier fathers are especially relevant for real founders who face more binding liquidity constraints. Self-employed parents are more relevant for family business successors, because they do not only transfer important self-employment specific skills and wealth to their offspring, but they also provide their offspring an opportunity to take over the family business. However, this opportunity may be constrained. Our results indicate that siblings may be a resource for real founders, but they are more a liability for family business successors, as they compete for taking over the family business.

The structure of the paper is as follows. Section 4.2 introduces a rational choice model of the self-employment process. This dynamic model differentiates between de novo self-employment and family business succession and links the theoretical concepts of human, financial, and social capital to the timing of these self-employment decisions. Section 4.3 derives hypotheses from our model with respect to human and social capital and the dynamics of family business succession and de novo self-employment. The data and methods to test these hypotheses are described in section 4.4. Section 4.5 presents the outcome of the duration analysis with multiple risks. Section 4.6 concludes.

4.2 Theory development

4.2.1 A rational choice model of the self-employment decision

We build our theoretical model with the following objectives in mind. First, we wish to link the theoretical concepts human, financial, and social capital, directly to the waiting time into self-employment. Second, we wish to account for the different types of entrepreneurship, their effect on human capital and liquidity constraints and thresholds, and their dynamics over time. Third, we wish to focus on the intertemporal capital accumulation decisions of individuals.

Figure 4.1 provides an economic model of the self-employment decision process over the life course of an individual. This rational choice model assumes that individuals collect information on the alternatives, assess the value of the attributes of the alternatives, and choose the alternative with the highest perceived utility given the available constraints. This maximizing behavior is the building block of most choice models in economics. We choose the occupational choice model as our starting point. Within the occupational choice framework, the individual chooses the occupation with the highest perceived income or utility (see e.g. Dunn and Holtz-Eakin 2000). For example, the individual decides to work for a pay in the labor market or decides to earn its income with self-employment. The individual will opt for self-employment if the expected utility is higher in that option.

Self-employment refers to a particular income generating mechanism that is different from paid labor. The utility derived from paid labor depends on income and a vector of personal characteristics (e.g. education, race, marital status, and number of children) (Dunn and Holtz-Eakin 2000; Evans and Jovanovic 1989; Holtz-Eakin et al. 1994). For a self-employed entrepreneur, the individual's gross earnings are a function of its entrepreneurial ability, the capital invested in the business and an individual random component summarizing the unknown good fortune of the entrepreneur.

Lucas (1978) general equilibrium model of entrepreneurial selection argues that individuals with high entrepreneurial ability should enter entrepreneurship directly when entering working life. However, most entrepreneurs do not become immediately self-employed but work first in the labor market (Carroll and Mosakowski 1987). During the time that individuals spend in the labor market, they accumulate non-transferable job-specific human capital. This labor market specific human capital raises the income in the labor market, making wage earners more reluctant to exit their job and enter self-employment. This is consistent with the finding of Evans and Leighton (1989) that the return to wage experience in self-employment (2.1%) is lower than the return to wage experience in wage work (5.6%).

Therefore, in theory, the existence of non-transferable job specific human capital should induce would-be entrepreneurs to enter self-employment as early in their career as possible. Furthermore, job shopping theory (Johnson 1978) argues that individuals should take more risky job alternatives early in the career. By pursuing risky alternatives, individuals learn whether they possess abilities that are highly valued and paid accordingly.

Figure 4.1: Conceptual model of the Self-Employment Process

In the light of these theoretical arguments and echoing Evans and Leighton (1989), why is it then that empirical studies do not find high self-employment rates or young individuals?

The answer must not be searched exclusively in the difference of expected earnings in both income options, but in the availability of the options. Even for individuals that are willing to become self-employed and who can earn more as self-employed, self-employment may be not an option. Self-employment differs in some important aspects from employment in the labour market. Self-employment demands a variety of skills (Lazear 2005), involves considerable and partially sunk start up costs, while the success changes are uncertain (Jovanovic 1982) and difficult to predict and therefore hard to finance in the capital market (Zeira 1987). Therefore, capital must be raised in the capital market and skills and capabilities must be learned through schooling or work experience in order to start a new venture. Capital market constraints limit the ability of entrepreneurs to finance new ventures (Evans & Jovanovic, 1989). And human capital constraints limit the ability of entrepreneurs to manage a business (Calvo and Wellisz 1980). Therefore, willing young individuals lack often financial and human capital to start up a venture.

Liquidity constraints are a stylized fact in the self-employment literature. Evans and Jovanovic (1989), Evans and Leighton (1989), Lindh and Ohlsson (1996), Holtz-Eakin, Joulfaian, and Rosen (1994) and Blanchflower and Oswald (1998) provide evidence that greater personal wealth, winning lotteries, or inheritance relaxes liquidity constraints and eases the transition to self-employment. With perfect capital markets, entrepreneurs would not face liquidity constraints. However, starting entrepreneurs face liquidity constraints because the exact profitability of the business idea is not measurable (Zeira 1987). The expected profitability hinges on the business owner's perception of his entrepreneurial ability (Frank 1988; Jovanovic 1982) and he often finds it "impossible to persuade potential suppliers of equity capital to share his subjective belief" (Steigum 1978: 637). The capital market is not the only means through which an entrepreneur can raise capital. Basu and Parker (2001) point that family loans are the largest source of funds after bank loans. Family members share more easily the beliefs of their offspring and have more trust in the propensity of the offspring to pay back the loans. In the absence of external financing possibilities, the prospective entrepreneur might finance the business itself by working and saving money to overcome the financial capital thresholds. However raising capital through saving is time consuming and therefore the absence of external financing options increases the waiting time into self-employment.

Besides financing a business, a business must be managed. The entrepreneur must be a jack-of-all-trades (Lazear 2005). Searching and evaluating business opportunities, building and fulfilling contracts, hiring and firing employees, and managing and book-keeping are part of the entrepreneur's tasks. Some researchers claim that these mentioned capabilities are inborn (Lucas 1978; Jovanovic 1982), while Calvo and Wellisz (1980) argue that these capabilities are acquired through a learning process. Dunn and Holtz-Eakin (2000) find evidence for intergenerational transfer of these skills. The offspring of self-employed parents are more likely to enter self-employment and children of more successful entrepreneurs are more likely to enter self-employment than children of less successful entrepreneurs. While differing on the source of these capabilities, all researchers argue that individuals that acquire sufficient knowledge of these entrepreneurial

processes become entrepreneur. The speed of learning and accumulating of these capabilities determine the waiting time into self-employment (see Figure 4.2).

The entrepreneur not only chooses self-employment, he also chooses for a certain type of self-employment. There are various types of self-employment (Cooper and Dunkelberg 1986; Webster 1977). The two most important ones are the take over of the family business and the founding of an entirely new business (Carroll and Mosakowski, 1987). Certainly, the human and financial capital thresholds will depend on the route followed. People who take over a family business cannot be compared to those who found an entirely new business. Newly founded firms suffer from liability of newness, i.e. they face lower early survival chances because they miss stable exchange relations, do not have established a reputation yet, or must learn by doing (Hannan and Freeman 1989; Jovanovic and Lach 1989). Furthermore, while real founders enter self-employment with the burden of financing the sunkness of start up expenses, continuation

Figure 4.2: Capital Accumulation, Thresholds, and Waiting Time

of a family business is like entering self-employment with a valuable set of assets already existent. Taking over a family business evolves often gradually where the successor learns the tricks of the trade internally inside the business. Because of this internal learning and financing, the human and financial capital thresholds are presumably lower for individuals entering family business and will therefore require less time to enter self-

employment. Figure 4.2 depicts the effect of disentangling self-employment into real founders and family self-employment on the waiting into self-employment. Real founders have to accumulate more financial and human capital to overcome liquidity and capabilities thresholds.

We summarize these theoretical arguments revisiting Figure 4.1. Figure 4.1 shows the relationships between human, financial, and social capital and their relationship to the different types of self-employment, starting up a new venture and taking over a family business. Most theorizing focuses on the direct link between human capital and self-employment. We argue that the theory is more complex. A complete theory must introduce the intervening human capital, liquidity, time, and availability constraints to explain the relationship between the time dependent capital endowments and self-employment types. These constraints can be perceived as attributes of the alternative self-employment types. We argue that these constraints differ between types. We assume that real founders face higher financial and human capital constraints than those who take over a family business.

The concepts of human and social capital are umbrella terms. In the next section, we derive hypotheses for these concepts embedded in our rational choice theory of self-employment.

4.3 Hypotheses development

4.3.1 Entrepreneurial capabilities and human capital

Penrose (1959) first notes what defines entrepreneurial capabilities. She identified four qualities an entrepreneur should have: (1) ambition, (2) fund raising ingenuity, (3) versatility²², and (4) good judgment. In order to become successfully self-employed one may learn to improve these capabilities through time. These entrepreneurial capabilities drive selection into entrepreneurship in Lucas' (1978) general equilibrium model. Inspired by Lucas model, Calvo and Wellisz (1980) explain the impact of educational attainment on the probability of selection into an entrepreneurial position through managerial ability. Investment in education enhances managerial ability, which would relax the human capital constraint and therefore increase the choice for self-employment. This is depicted in Figure 4.2. Human capital accumulation may be endogenous with financial capital accumulation. The higher educated attract higher wages and have more possibilities for saving or external lending (Bates 1990). Therefore, investment in education relaxes both human capital and liquidity constraints.

However, there is also a direct link between educational attainment and self-employment selection. The direct link between human capital and self-employment in Figure 4.1 represents the effect of human capital on the different types of income generating mechanisms. The direct link between educational attainment and self-employment selection represents the value trade-off between being self-employed and some outside option like working in the labor market. Le (1999) argues that higher levels of education

²² Note the close resemblance with Lazear's (2005) jack-of-all-trade hypothesis.

might generate better outside options (i.e. better working conditions or better pay). Therefore, better outside options through education might decrease the probability of self-employment. Thus, investment in education relaxes human capital constraints, while at the same time raising better outside options. These offsetting forces can account for the fact that some studies find a positive effect of education on the probability of self-employment (e.g. Alba-Ramirez 1994; e.g. Van Praag and Van Ophem 1995) while others find a negative (e.g. Evans & Leighton, 1989; Evans & Jovanovic, 1989) or no effect (e.g. Blanchflower and Meyer 1994; Dunn and Holtz-Eakin 2000).

Figure 4.3 provides a graphic representation of this mechanism. If an individual invests in education, he relaxes his human capital constraint (line B in Figure 4.3 and the arrow between human capital and human capital constraint in Figure 4.1). At the same time, the investment in education generates better outside options (line A in Figure 4.3). These better outside options act as opportunity costs for becoming self-employed. Logically, there is a point (point Z in Figure 4.3) where an increase in the level of education raises opportunity costs to such a level that the outside option is the preferred choice over self-employment.

Figure 4.3: Human capital mechanism

Figure 4.2 has shown that human capital thresholds determine the waiting time into self-employment. We have also reasoned that thresholds are larger for real founders (line D in Figure 4.3) than for family business (line C in Figure 4.3). Given equal outside options for individuals, these two results indicate that an increase from lower to vocational schooling would induce more individuals into family business (the triangle XYZ in Figure 4.3) and less individuals into founding new ventures (the triangle VWZ in Figure 4.3).

Lazear (2005) provides a related argument with his jack-off-all-trades hypothesis of entrepreneurship. He states that self-employment requires a variety of skills. Entrepreneurs must have sufficient knowledge in a variety of fields to put together the many tasks needed for survival and success in business, while for wage earners it suffices and pays to be a specialist in the field demanded. Higher or specialist education, like university studies, often implies a higher level of specialization than vocational studies. While on the other hand, primary schooling and lower vocational schooling might not produce enough human capital to overcome the required human capital constraint for self-employment. Both arguments show that it pays for specialized and low educated to work in the labor market, while a vocational education is preferable for self-employment. Therefore, when an individual finishes a vocational schooling he will have a higher chance to become self-employed and thus:

Hypothesis 1a: Vocational schooling will reduce the waiting time into self-employment.

If starting up a business requires more human capital then taking over a family business then the pioneering founder must face a higher human capital constraint and it must follow that (see Figure 4.3):

Hypothesis 1b: The role of vocational schooling is larger for family self-employment.

While on the other hand, increased specialization instead of or after vocational schooling, will increase the availability and pay of an outside option and thus:

Hypothesis 2a: Specialized education will increase the waiting time into self-employment.

Both the jack-off-all-trade hypothesis and the mechanism in Figure 4.3 are inconclusive of the effect of specialized education on the type of self-employment. However, Laband and Lentz's (1983) economic theory of occupational following provide some useful hints. Their rational choice theory of occupational following states that the likelihood that the offspring follows their parents occupation increases in the interface between the home and the workplace and the presence of undissipated monopoly-type rents, like for example family business reputation. They state that if most followers acquire their human capital skills directly from their parents, indices of later acquisition such as formal education and training programs should be negatively related to the incidence of following (Laband & Lentz, 1983: 476). As the incidence of following is higher in family self-employment then in de novo self-employment, the negative effect of formal education on self-employment must be more pronounced for family self-employment. This is consis-

tent with the empirical finding of Carroll and Mosakowski (1987) that higher general education increases the probability of entering self-employment while it lowers the probability of entering family self-employment after entry into the labor force. And therefore,

Hypothesis 2b: The role of specialized education is larger for self-employment via a family business.

4.3.2 Social capital and family background

Next to human capital we introduce social capital as another important input for self-employment. Social capital refers to resources that individuals have at their disposal, or which they can access by means of their social relations. Social capital differs between people and can form a considerable advantage in learning how to focus on becoming self-employed (for more general notions on social capital see Burt 1992; Coleman 1988; Coleman 1990; for more general notions on social capital see Flap 1990). Although the concept of social capital originates from sociology, it has caught increasing interest in economics (see e.g. Becker 1996). The importance of social capital for the success on the labor market, i.e. finding a job, has been studied by Granovetter (1973). He distinguishes two kinds of relations: strong and weak ties. Strong ties are characterized by frequent and intense communication. Loose contacts to other people, like former colleagues and college friends, represent weak ties. Empirical evidence suggests that strong ties are the dominant factor for successful business founding (Brüderl and Preisendörfer 1998). In our study, we focus on strong ties, namely one's parents, one's siblings, one's partner and one's children.

The social ties considered here offer various forms of resources that may be helpful to succeed in starting a business. First, social relations can offer access to financial capital. Research on ethnic entrepreneurship reveals for example that within the closely knitted communities of Chinese and Korean immigrants, business starters obtain financial start capital from their social relations (Bates 1997; Light and Bonacich 1988). Another tie providing income security is a working partner, whose earnings can cover private expenses for housing, food etc. when the new business does not yet generate sufficient initial benefits. Second, social relations form an inexhaustible source of knowledge, know-how and experiences. In general, network ties can form an essential complement to one's own human capital and those network ties accumulate important information through time as well. Family members with specific human capital, for example a mother with a degree in accounting or law, are valuable contacts for fiscal and juridical problems. Very special ties are self-employed parents. Children from self-employed parents learn what it needs to be self-employed early in their childhood. Their parents are their entrepreneurial role model. And, parents with self-employment experience are more likely to support the self-employment decision of their children because their own experience enables them to assess such a decision more accurately.

To assess an individual's strong ties, i.e. the family relationships, we focus on the parents and the spouse through time. Experience refers mainly to specific entrepreneurial capabilities that have been transferred from the parents to the child during the up bring-

ing. The educational level of an individual's parents is an indicator capturing the transfer of more general human capital as well as a proxy for the amount of financial resources accessible through family ties. Thereby, we assume that higher educated parents are wealthier relative to lower educated parents.

Parental self-employment experience and educational level characterize the parents' background. Having a self-employed parent has been found as one of the most consistent predictors for off-spring self-employment (e.g. Carroll & Mosakowski 1987; Dunn & Holtz-Eakin 2000). However, there are many reasons why having self-employed parents increases offspring self-employment. Dunn and Holtz-Eakin (2000) study in particular these intergenerational links. They argue and find evidence for two possible mechanisms through which self-employed parents increase offspring self-employment. Successful self-employed may be more willing and able to provide offspring with financial capital to relax liquidity constraints (the link between SC and FC and FC and FCC in Figure 4.1). The second reason why self-employment may be correlated among generations is because parents may transfer entrepreneurial capabilities and skills through valuable work experience, reputation, or other entrepreneurial human capital (the link between SC and HC in Figure 4.1). The direct transfer of entrepreneurial skills is also consistent with the rational choice of occupational following, which states that parents transfer job specific human capital for intergenerational welfare maximization (see e.g. Becker and Tomes 1979; Laband and Lentz 1983).

The educational level of the parents is an indicator capturing the transfer of more general human capital as well as a proxy for the amount of financial resources accessible through family ties. Thereby, we assume that higher educated parents are wealthier relative to lower educated parents. The intergenerational transfer of wealth relaxes the financial constraints (see Figure 4.2), thereby increasing the chances that the offspring may overcome initial set up costs and reduce the time into self-employment.

These arguments lead to the following hypotheses:

Hypothesis 3a: Having a self-employed parent shortens the offspring's waiting time into self-employment.

Hypothesis 4a: Having a highly educated parent and possible wealthier parent shortens the offspring's waiting time into self-employment.

Is it possible that the presence of self-employed parents merely reflect family self-employment instead of the transfer of entrepreneurial skills? This question was legitimately raised by Dunn and Holtz-Eakin (2000) in their research on intergenerational transfers of skills and wealth. Although they indirectly tested for this possibility by comparing father and son's occupation and industry classification, a formal test of this possibility has not been undertaken up to today. The possibility of family self-employment is indicated in Figure 4.1 by the link between self-employed parents and the availability of family business constraint. Although individuals may take over a family business from aunts and uncles, the offspring from self-employed parents are more likely to have the possibility to take over a family business. The existence of a direct link between self-employed parents and the probability that a family business is available should be reflected in a stronger effect of self-employed parents on taking over a family business as opposed to starting up a business from scratch. Therefore,

Hypothesis 3b: The role of a self-employed parent is larger for entering self-employment via a family business.

The reverse can be said for the role of the educational level of the parent. Although some general skills might be transferred from parent to children, the main effect of the educational level of the parent is the transfer of wealth. Here we assume that the higher the educational level of the parents the more financial resources are accessible. We expect that social capital associated with access to financial capital is more relevant for real founders than for self-employed who take over a family business.

Hypothesis 4b: The role of the parent's education is larger for real founders.

However, an individual's chance of taking over the family business depends not only on his skills but also on the number of competitors willing to take over the family business. The more siblings the possible business successor possesses, the more competitors he faces that might take over the family business. These siblings would increase the family business availability constraint, thereby lowering the chance to take over the family business. On the other hand, social capital in the form of siblings may offer time and support to the business when necessary (the link between SC and time constraint in Figure 4.1). These opposing forces make it difficult to predict the role of siblings on the probability of self-employment. However, if we differentiate the types of self-employment we might disentangle these opposing effects. So that, siblings reduce the chance to take over or enter family self-employment but increase the probability of starting up a business from scratch.

Hypothesis 5a: Having siblings reduces the likelihood that self-employment is entered via a family business.

Hypothesis 5b: Having siblings increases the likelihood that self-employment is entered via a de novo firm.

Besides one's parents and siblings, one's spouse might be an important determinant of self-employment as well. Next to her human capital, a spouse helps a self-employed partner providing monetary, physical as well as emotional support. Furthermore, the spouse is often an unpaid reliable employee stepping in when work mounts up (Sanders and Nee 1996). We have argued that social capital associated with access to financial capital is more relevant for real founders than for self-employed who take over a family business. Similarly, we assume that social relations providing stable income sources are also more important for real founders who face higher levels of uncertainty and failure risks. Applying these arguments to the dimensions of social capital used here, we hypothesize that the effect is stronger for the real founders.

Hypothesis 6a: Marriage shortens the waiting time into self-employment.

Hypothesis 6b: The role of marriage is larger for real founders.

Up to now we emphasized the supportive features of social relations. However, social ties are two sided and based on some kind of reciprocity. Hence, from time to time social relations will make a claim on one's own resources. Such social demands can restrict the pursuit of other goals. For example, children have the right to claim affection and time that can not be invested in building up the business. Furthermore, the larger risk associated with an income from self-employment is more problematic for someone who is responsible for dependent children than for someone without such responsibilities. Therefore,

Hypothesis 7: Having children lengthens the time (duration) to self-employment.

4.3.3 The dynamics of entering self-employment over the lifecycle

Until now, we have investigated the effects of human and social capital determinants on the propensity to become self-employed and the strength and direction of the effects on different types of self-employment. In this section, we will investigate the route into self-employment. More specific, we will investigate the shape of the different routes into self-employment.

Unfortunately, the time dimension has been largely ignored in the econometric modeling of the self-employment process. This is partly due to the absence of an empirical time dimension in the highly influential studies based on the National Longitudinal Surveys of Labor Market Experience (NLS) dataset (see e.g. Evans & Leighton 1989; Dunn & Holtz-Eakin 2000). We do not refute the quality of this research, however we will pose the hypothesis that the lack of duration dependence found in these studies may be partly attributable to the inability of these studies to separate the different types of self-employment.

In the less often quoted study of Carroll and Mosakowski (1987), there are some very definite hints towards duration dependence among the divergent routes of de novo self-employment and family self-employment. They find that from the individuals that end up in family self-employment 42% have family self-employment as their first job. While only 5% of the real founders have self-employment as their first job. These figures show that the timing of the self-employment decision differs greatly among individuals who take over family self-employment and real founders. More important, these findings show that these routes may be opposing. Individuals that take over a family business do so most likely in the beginning of their career, while pioneering founders become self-employed later in their careers. The fact that these routes seem opposing and that family self-employment may account for one third of total self-employment raises the possibility that at the aggregate these opposing routes may offset duration dependence.

The question remains what accounts for these empirical findings? We refer back to our conceptual model to account for the difference in early self-employment. Consistent with the theory of occupational following (Laband & Lentz 1983), job shopping theory (Johnson 1978), and the accumulation of labor market job-specific human capital, individuals that have the intention to become self-employed should become self-employed early in their career. However, unlike family self-employment which has relatively low human and financial capital thresholds, starting up a business from scratch

requires large amounts of financial and human capital. In order to reach these thresholds, skills must be learned either through working experience or prolonged learning and wealth must be accumulated through savings. These accumulation processes take time, and therefore, a much lower fraction of de novo entrepreneurs will enter self-employment early in their life than family self-employed.

Only if sufficient amounts of human and financial capital have been accumulated, then de novo self-employment will rise. However, prolonged working in the labor market to accumulate human and financial capital will simultaneously increase job-market specific human capital which reduces self-employment (Evans & Leighton 1989). The same effect occurs for family self-employment. The longer the potential family self-employed is employed in the labor market and accumulates labor market job-specific human capital the lower will be chance that he will shift to self-employment in the future. The difference in early self-employment for de novo self-employed and family self-employed is reflected in the mean age of the self-employed in the study of Carroll and Mosakowski (1987). The mean age for pioneering founders is thirty and a half years while the mean age for family self-employed is 23 years²³.

What will be the dynamics of self-employment for these two types of self-employment? What are their baseline hazards? Our dynamic theory dictates that the baseline hazards must have at least the following features. The baseline hazard for family self-employment must be initially high and then decreasing afterwards. And the baseline hazard for de novo self-employment should be initially low, then increasing and possibly decreasing at later stages of the lifecycle. There are several parametric distributions of the baseline hazard consistent with our theory. For family self-employment, the possible candidates are the Weibull (decreasing) or Gompertz (decreasing) distribution. While for de novo self-employment the candidates are the Log-logistic, Log-normal, or the Weibull (increasing) distribution.

4.4 Methodology

4.4.1 Data

The hypotheses developed above are tested on a data set consisting of 1339 people. This sample is part of the North Brabant longitudinal data set which contains information about 3167 individuals who were born 1939/1940 and went to school in the Dutch province North Brabant in 1952 (Hartog 1989; Jonker 1995). Then researchers took a random sample of pupils who attended the sixth grade of primary school. 2874 individuals provided information about their school performance and family background. In 1983 all 3167 individuals were contacted again. 2588 people responded and were asked about their later education, labor market status, earnings, and household composition. In 1993 at the age of approximately 54, the 3167 school children of 1952 were approached for the

²³ A notable exception forms the de novo entrepreneur who might transfer his labour market job specific capital into a new business idea. This often happens in so-called spin offs from large corporations.

third time. The questionnaire mailed in 1993 contained questions on the labor market status since 1983 and especially on self-employment experience. 2099 individuals answered in 1993. Due to untraceable individuals and non-response 1339 individual were interviewed at all three points of time, i.e. in 1952, 1983 and 1993. Taking the 1952 data as a population for the data collection in 1983 non response is not systematic for family background and school performance (see Hartog 1989). 253 individuals provided incomplete information, which leads to a final sample of 1086 individuals. 186 of these respondents have ever been self-employed and provided information on the year of founding or acquiring business (see van Praag 1996 for more detailed information on the dataset). 64 took over a family business and 122 were real founders.

4.4.2 Used methods and operationalization

Contrary to earlier empirical research presented above, we apply duration analysis to test our hypotheses. Rather than conditioning on a time independent inverse Mill's ratio, as often used in traditional probit studies, we model a time varying hazard function. The basic principle of hazard models is to estimate the time it takes until an event occurs given that the event did not occur earlier. We estimate the time it takes to acquire the necessary resources to become self-employed. Hence, the dependent variable is the difference between year of entry into self-employment and the year of finishing compulsory education.

The chance of becoming self-employed is related to the available human and social capital (see Table 4.1 for the descriptives of the variables). Two variables serve as indicators for human capital: (1) VOCATIONAL SCHOOLING and (2) SPECIALIZED EDUCATION. VOCATIONAL SCHOOLING is a time varying dummy taking the value of 1 at the time the individual completed MBO (Middelbaar Beroeps Onderwijs). MBO is a general vocational education in the Netherlands. The variable SPECIALIZED EDUCATION is a time varying dummy taking the value of 1 at the time the individual completes his higher education. Higher education in the Netherlands is HBO (higher vocational schooling) or a university degree.

The social capital endowment is focused on strong ties, i.e. only close family ties are considered. Four ties are distinguished: (1) parents, (2) siblings, (3) marital status, and (4) own children, of which the first two are constant and the latter two vary through time. The variable SELF-EMPLOYED FATHER is a dummy that equals one if the respondent had a self-employed father and hence grew up in a self-employed environment. Access to parent's general human capital and possibly financial wealth is captured by the EDUCATIONAL LEVEL OF THE FATHER. The variable SIBLINGS is a continuous variable that indicates the number of brothers and sisters of the respondents.

The independent variables on social capital presented above are constant for the observation time 1952 to 1993. The acquired education of the respondent, the father's educational level and economic status are not likely to change in this period. Each respondent's value for these factors is fixed before the self-employment decision is made. This chronological order is not predetermined for the two other social capital variables MARRIED and CHILDREN. A person who is married at the second interview wave in 1983 could be married before as well as after becoming self-employed. The same holds true for

Table 4.1: Descriptive statistics

	Mean	Standard Deviation
Vocational Schooling*	.21	.41
Specialized Education*	.27	.44
Self-Employed Parents	.39	.49
Educational Level Father	2.52	.86
Siblings	5.90	2.71
Married*	.83	.38
Children*	.11	.31
IQ	104.73	10.27
Female	.36	.48
Commercial Training*	.43	.50
Rank	3.34	2.54

* in 1993

having children. Observing that respondents have children in 1983 does not tell us anything about parenthood at the time the respondent became self-employed. Since the year is known that a person got married and eventually divorced and remarried again the variable MARRIED can vary over time for each respondent. It takes the value one in the years a respondent is married and the value zero otherwise. Using the birth year of respondent's children we are able to construct a time varying co-variate CHILDREN. This variable turns one in the year the first child is born and turns zero again when the youngest child becomes 16 where we assume the child is not dependent on his parents anymore. Defining 16 as the age a child is not dependent on its parents anymore is certainly a little bit arbitrary. However, changing the independence year between 16 and 21 did not substantially alter the results.

We followed earlier empirical studies in the considered controls. All our respondents were born in 1939/40, so that age is not included. We control for gender with the dummy variable FEMALE. The variable IQ is an additive scale of the results of three IQ-tests that were taken while the respondents attended primary school in 1952/3. Cronbach's α , an indicator for the reliability of additive scales, for the three kinds of IQ scores is .759 and sufficiently high. IQ and IQ SQUARED might be an indicator for general human capital. Entrepreneurial human capital is measured by a bi-variate variable constructed from the kind of education. The time-varying control variable COMMERCIAL TRAINING gets the value one if the attended education incorporated some commercial training and the value zero otherwise. Finally, we add the variable RANK to control for the rank of the respondent within his brothers and sisters.

In the original data set each respondent was represented by one record. Introducing the time-dependent co-variables VOCATIONAL SCHOOLING, SPECIALIZED EDUCATION, COMMERCIAL TRAINING, MARRIED and CHILDREN enlarged the number of records for respondents. Thus, records are nested within persons, our unit of analysis. We controlled for this respondent nesting using robust estimations of the standard error in all analyses.

4.4.3 Model

In our statistical model we estimate the time it takes to become self-employed after an individual finishes compulsory school (indicating the start date at which individuals are at risk of becoming self-employed). The sampling frame that generated our data can be considered a random inflow sample with right censoring. The beginning of the sample period corresponds to the start of a spell. Therefore, our sampling frame does not suffer from the widely neglected initial conditions problem (Heckman and Singer 1984). Consequently, we do not have to assume a time-homogenous (stationary) environment. More formally, this ‘ideal’ sampling scheme (flow data) is said to be uninformative of duration t (Lancaster 1990: 163). Although our data is grouped into years, we model durations in a continuous time framework, because we believe that ratio of the length of the grouping interval to the typical spell length is small enough (Jenkins 2004). The continuous framework is mathematically more tractable and the parameter estimates are more comparable across different studies (Lancaster 1990: 12). Modeling durations in continuous time, in stead of discrete time, even in the case of aggregation (grouping), is advocated by Heckman and Singer (1984) and Lancaster (1990), at least if the process that generates the duration data is not intrinsically discrete.

The appropriate method to study durations is estimating a survival model. Let t be a continuous duration random variable with density function $f(t)$, cumulative density function (Failure function) $F(t)$ and hazard function $\theta(t)$. Note that $\theta(t) = f(t)/S(t)$. We first model the duration of a spell t , the time it takes before an individual becomes self-employed after finishing compulsory school, without distinguishing the entry routes. Let $t_1, t_2, \dots, t_n$ be the observed durations in our sample, where an indicator c_i equals 1 if the spell is completed and 0 if the spell is censored. Then the contribution of the i -th individual to the log likelihood is given by (cf. Lancaster 1990; Jenkins 2004):

$$(4.1) \quad \ln L_i = c_i \ln \theta_i(t_i) + \ln S_i(t_i)$$

where $S_i(t_i)$ is the survivor function, which is equivalent to $1-F(t)$. The survivor function is related to the hazard function by:

$$(4.2) \quad S_i(t_i) = \exp \left[- \int_0^{t_i} \theta_i(u) du \right]$$

So that the log-likelihood contributions in (4.1) can be written in terms of the hazard as:

$$(4.3) \quad \ln L_i = c_i \ln \theta_i(t_i) - \int_0^{t_i} \theta_i(u) du$$

where the continuous time hazard rate, $\theta(t)$, can take several functional forms. Since, this is the first paper to consider the functional form of the hazard, we estimate several functional forms. We choose the appropriate functional form on three criteria: (1) theoretical background: are the functional forms consistent with our theory, (2) statistical

tests: because most of the models are non-nested, we use Akaike Information Criterion (AIC) for model selection (Kennedy 2003: 117), (3) appropriate for testing our hypotheses. The last point is especially important in our case. Thomas (1996) and Booth and Satchell (1995) show that parameter estimates are only directly comparable if the parametric distribution of each risk take the proportional hazards form. In order to test our hypotheses, we restrict our analyses for the multiple risk models to the proportional hazards distributions.

Now we consider the different entry routes, taking over a family business and starting from scratch (pioneering founder). Therefore, we model duration t jointly with the entry route. Let E be a discrete random variable indicating the entry route or cause of exit from the not self-employed state. So that we consider the joint distribution of t and E . Denote the j -th ($j = 1$ (family business), 2 (real founder) cause-specific hazard for individual i by $\theta_{ji}(t)$.

If we assume independent failure types (independent multiple risks), the overall hazard function is given by:

$$(4.4) \quad \theta_i(t) = \theta_{1i} + \theta_{2i} = \sum_{j=1}^2 \theta_{ji}$$

and the overall survivor function by:

$$(4.5) \quad S_i(t_i) = \exp \left[- \int_0^{t_i} \sum_{j=1}^2 \theta_{ji}(u) du \right]$$

Let c_{ji} be an indicator which equals 1 if i enters via route j , 0 otherwise (censored or entry via other route). Then the log-likelihood contributions are given by:

$$(4.6a) \quad \ln L_i = \sum_{j=1}^2 c_{ji} \ln \theta_{ji}(t_i) + \ln S_i(t_i)$$

$$(4.6b) \quad \ln L_i = \sum_{j=1}^2 c_{ji} \ln \theta_{ji}(t_i) - \int_0^{t_i} \sum_{j=1}^2 \theta_{ji}(u) du$$

$$(4.6c) \quad \ln L_i = \sum_{j=1}^2 \left[c_{ji} \ln \theta_{ji}(t_i) - \int_0^{t_i} \theta_{ji}(u) du \right]$$

So that the overall log-likelihood can be separated into a sum of independent single cause-specific hazards, which can be estimated separately by treating the durations ending for other reasons as censored (Jenkins 2004; Narendrenathan and Stewart 1991; Thomas 1996). From equation (4.6c) it becomes clear that each cause-specific hazard can have its own parametric functional form. The appropriateness of the multiple risks specification (versus the single risk model) can be tested by a non-nested likelihood-ratio

test (Lindeboom and Theeuwes 1991; Narendrenathan and Stewart 1991). Under the null hypothesis of equal risks ($\theta_{1i} = \theta_{2i} = (0.5)\theta_i$), the maximum log-likelihood value of the multiple risk model equals the maximum log-likelihood of the single risk model plus the logarithm of a 0.5 multiplied by the total number of uncensored observations (the total number of entrants into self-employment irrespective of the route).

4.5 Results

Before we discuss the results of our duration analysis, we need to choose an appropriate parametric distribution for each model. Table 4.2 provides the log-likelihood and AIC values for the following parametric distributions: (1) proportional hazards: Exponential, Weibull and Gompertz; and (2) accelerated failure time: Log-Logistic and Log-Normal (see Cleves et al. 2004 for a detailed description of the distributions). We only estimate proportional hazards distribution for our multiple risk models in order to be able to test our hypotheses. The AIC values point toward the Log-Normal distribution for the single risk model, the Weibull distribution for de novo self-employment and the Gompertz distribution for family self-employment. These distributions are consistent with the shapes predicted by our theory (more about the interpretation of these distributions later). The interpretation of the coefficients differs between accelerated failure time models and proportional hazards models. A positive coefficient in the accelerated failure time model increases the waiting time into self-employment and a negative coefficient reduces the waiting time. A positive coefficient in the proportional hazards model increases the hazard of becoming self-employed, or equivalently, reduces the waiting time into self-employment and for a negative coefficient vice versa.

Table 4.2: Selection of parametric distributions – AIC values

		Exponential	Weibull	Gompertz	Log-Normal	Log-Logistic
Single Risk	Self-Employment	1250.6	1251.5	1240.5	1221.3	1239.3
	Family Business	554.5	555.2	523.7	-	-
Multiple Risk	Real Founder	914.6	910.6	916.6	-	-

Next, we discuss the issue of unobserved heterogeneity. Unobserved heterogeneity poses the risk that the duration dependence might be the result of some omitted variable. The inclusion of unobserved heterogeneity in our model is complicated. First, as noted by Van Praag (2001), the inclusion of unobserved heterogeneity with time-varying regressors raises many complications. The inclusion of unobserved heterogeneity in a multiple risk framework poses even extra challenges. Multiple risk models can be estimated from computer programs designed for single risk models (Thomas 1996). How-

ever, the inclusion of unobserved heterogeneity cannot be estimated in a single risk model, but should be programmed separately. Even if this would be possible, the interpretations of these coefficients are not straightforward (Jenkins 2004). Alternatively, we might gain some understanding of the presence and severity of unobserved heterogeneity by estimating two separate single risk models with individual unobserved heterogeneity specifications. The results from these estimations found no unobserved heterogeneity for family self-employment and some unobserved heterogeneity for de novo self-employment. Fortunately, the direction of the effects of the parameter estimates remained largely consistent, with some reductions in significance levels in the single risk of de novo self-employment²⁴. More importantly, the shape of the baseline hazard proved consistent for both specifications with and without unobserved heterogeneity.

Table 4.3 shows the results of the duration analysis for the whole sample. The table differentiates the single risk of becoming self-employed from the different routes toward self-employment: taking over a family business and starting from scratch. Overall, the results show that social capital is a better predictor of self-employment selection than human capital. Furthermore, the result of our likelihood ratio test confirms the appropriateness of the multiple risks specification versus the single risk specification ($\chi^2(14) = 35.25$ with a p-value < 0.01). Separating family business succession from real founding is a successful innovation.

In particular, we find that vocational schooling increases the probability of self-employment. The positive sign of the variable VOCATIONAL SCHOOLING indicates that the variety of skills offered at vocational schooling reduces the human capital constraint and reduces the waiting time into self-employment. Hypothesis 1a is supported. We also find support for hypothesis 1b. Although, the effect of Vocational Schooling is not significant for both routes into self-employment separately, the effect of vocational schooling is larger for family self-employment.

We also find a negative effect for SPECIALIZED EDUCATION on the probability of self-employment. This seems to indicate that SPECIALIZED EDUCATION increases the availability and pay of the outside option more than that it reduces the human capital constraint. This effect is not significant and therefore we cannot support hypothesis 2a. However, if we differentiate the different types of self-employment this picture changes. SPECIALIZED EDUCATION decreases the probability of family self-employment. This finding is consistent with the argument of Laband and Lentz (1983) that later acquisition of human capital such as formal education is negatively related to the incidence of following. Since the negative effect of SPECIALIZED EDUCATION on family self-employment is larger than the effect on de novo self-employment, we do support hypothesis 2b. The role of Specialized Education is larger for family self-employment.

The results of our human capital indicators pose some interesting new light on the effect of human capital on self-employment. Entrepreneurs more often follow vocational schooling, being presumably more jack-of-all-trades. Furthermore, the stronger effect of vocational schooling on family self-employment than on de novo self-employment provides new information on the mechanism through which human capital affects self-employment selection. First of all, this effect suggests that individuals that take over a family business are less human capital constrained. As they are less human capital

²⁴ Reduction in significance is often observed (and critiqued) when including unobserved heterogeneity.

Table 4.3: Estimation results

	Single Risk Self-Employed	Multiple Risk Family Business	Real Founder
Human Capital			
Vocational Schooling	-.485 [.200]*	.532 [.330]	.308 [.254]
Specialized Education	.203 [.188]	-.546 [.415]	-.100 [.241]
Social Capital			
Self-Employed Father	-.973 [.164]**	2.102 [.346]**	.590 [.200]**
Educational Level Father	-.175 [.074]*	.038 [.158]	.259 [.088]**
Siblings	-.001 [.035]	-.113 [.065]+	.057 [.043]
Married	-.502 [.251]*	2.216 [.486]*	.082 [.319]
Children	.257 [.178]	-1.060 [.323]**	.212 [.211]
Controls			
IQ	-.094 [.088]	-.015 [.163]	.126 [.123]
IQ Squared	.000 [.000]	.000 [.000]	-.001 [.001]
Female	-1.074 [.203]**	-1.201 [.345]**	-.969 [.247]**
Commercial Training	.320 [.161]*	-.690 [.318]*	-.221 [.217]
Rank	.010 [.038]	.082 [.068]	-.060 [.051]
Shape Parameters	Log-Normal	Gompertz	Weibull
σ	1.502 [.113]**		
γ		-.087 [.015]**	1.372 [.126]**
Specification Tests			
$\chi^2(12)$	84.76**	76.95**	44.59**
$\theta_{1i} = \theta_{2i} = 1/2\theta_i$	32.25**		

+ significant at 10%; * significant at 5%; ** significant at 1%
standard errors in parentheses

constrained, they may enter self-employment earlier, and do not accumulate unnecessary large amounts of labor market job-specific human capital. Moreover, we find evidence that formal education reduces self-employment in general and occupational following in particular.

Next to human capital, we put forward social capital as an important determinant of self-employment selection. The strong and positive effect of the variable SELF-EMPLOYED FATHER supports hypothesis 3a. Having a self-employed father reduces the time into self-employment. The presence of a significant effect of SELF-EMPLOYED FATHER on the probability of de novo self-employment annihilates the possibility that a self-employed father means only family self-employment. These results indicate that self-employed fathers transmit entrepreneurial specific skills to their offspring. Because the effect of the self-employed father is stronger for family self-employment, we do find strong support for our hypothesis that self-employed fathers reduce the family business availability constraint and increase the chance of taking over a family business. So, we find support for hypothesis 3b.

The existence of this family business availability constraint is nicely illustrated by the significant and negative effect of the variable SIBLINGS on the probability of family self-employment. Keeping the rank of the child constant, we find that an increase in the number of siblings reduces the chances of taking over a family business. This supports hypothesis 5a and provides a clear example that social capital is not always beneficial but can have negative effects too. The results show no significant support to our hypothesis that more siblings increase the chances of starting a firm from scratch. Therefore we do not find support for hypothesis 5b.

The effect of the variable EDUCATIONAL LEVEL FATHER is positive and significant. Since we assume that a highly educated father is also probable wealthier, a wealthy father may transfer wealth to his offspring and relax the liquidity constraint. This supports hypothesis 4a. The assumed relation between father's education and wealth is strengthened by the stronger effect of the father's education on de novo self-employment. As the de novo self-employed is more liquidity constrained, the effect of a wealthy father should be stronger for de novo self-employed. This finding supports hypothesis 4b.

Besides inherited family and family resources, an individual obtains additional resources and liabilities through commitments over time, like marriage and children. We expected that marriage increases the probability to become self-employed. This hypothesis (6) is confirmed in our single risk model, where the time-varying variable MARRIED reduces the time it takes to become self-employed. In our multiple risk framework, it is shown that the effect of marriage is stronger for family self-employment. This contradicts our expectations and therefore hypothesis 6b is not supported.

The existence of a time constraint may also account for the negative effect of the time varying variable CHILDREN on family self-employment. The single risk model shows no significant effect of children on self-employment and thereby providing no support for hypothesis 7. However, these children do increase the waiting time of family self-employment.

Our control variables provide the following information. Women are less likely to become self-employed and less likely to take over a family business than starting a new venture. De novo self-employment can be better structured around raising a child. Commercial training reduces the likelihood of self-employment, especially among family self-

employment. This indicates that all specialization, even commercial, reduces occupational following.

4.5.1 The dynamics of real founders and family self-employment

What can we say about the dynamics of self-employment? Table 4.3 provides the estimated shape parameters and Figure 4.4 shows graphically the estimated baseline hazards for the single risk model of becoming self-employed and the multiple risk models of real founding and taking over a family business. Figure 4.4a plots the baseline hazard of becoming self-employed. The hazard of becoming self-employed increases strongly over the first ten years and then remains constant. The strong rise in the first ten years is partly due to our sampling. We follow the individuals from the age of 13 up to 53. Although at risk of becoming self-employed, most individuals pursue some additional schooling after primary schooling. Remarkably, we find that the probability of becoming self-employed remains basically the same over a period of thirty years. This is consistent with the absence of duration dependence in the NLS dataset, whose starting age is on average 19 years (e.g. Evans and Leighton 1989; Dunn and Holtz-Eakin 2000).

Figure 4.4: Baseline hazards

(4.4a. Single Risk -- Self-Employment)

(4.4b. Family Business)

(4.4c. Real Founder)

However, this picture changes dramatically if we differentiate between real founders and family self-employment. Figure 4.4b and Figure 4.4c provide the baseline hazard for taking over a family business and de novo self-employment respectively. The baseline hazard of de novo self-employment is the mirror image of family self-employment. Consistent with our theory, we find that if individuals take over a family business, they do so in the beginning of their career. The longer they postpone that decision, the less likely they are to take over the family business. In contrast, the probability that an individual starts a business from scratch is initially low. Only after enough skills and capabilities are learned through education and working experience, then the individual may be able to become self-employed. Clearly, the results indicate that taking over a family business or starting from scratch are two different processes and that aggregating these two processes might off set and bias duration dependence.

4.6 Conclusions & perspectives

In this paper, we investigated the dynamics of self-employment. We have built an economic theory of the self-employment process, where an individual's human, financial, and social capital determine the waiting time into self-employment. Our analysis reveals clear differences between people who took over an existing family business and people who founded a new firm.

Taking over a family business cannot be compared by starting up a business from scratch. Continuation of a family business is like entering self-employment with a valuable set of assets already existent, while real founders enter self-employment with the burden of financing the sunkness of the start up expenses. Furthermore, the uncertainty of the expected profitability is much lower for a firm that has learned about its true productivity, than for the de novo firm whose expected profitability is more a reflection of the subjective beliefs of the business owner. This is also reflected in the finding that the probability of entering self-employment is age dependent and that the shape of the duration dependence is self-employment type dependent. Because learning may take place in the family business and less sunk costs have to be financed, the hazard of taking over a family business is highest in the earliest stages of one's professional career. In contrast, starting up a business from scratch takes more time. First the would-be self-employed has to accumulate enough skills, capabilities, and capital to overcome set up

costs and to become self-employed. The real founder's probability of entering self-employment is initially low and increases only gradually over his life.

We tested hypotheses about the influence of human and social capital on the waiting time to become self-employed using a data set that is unique in terms of composition and the duration time covered. We have found that the relationship between human capital and self-employment is complex. On the one hand, investment in education produces necessary skills and increase self-employment. While on the other hand, investment in education produces better outside options and reduces self-employment. We found that vocational schooling is most suited for self-employment, while individuals pursuing specialized education are more likely to work as a wage worker.

Next to human capital, we introduced social capital as an important predictor of self-employment. We have found that social capital can reduce the waiting time into self-employment considerably. Self-employed parents transfer entrepreneurial skills and can transfer the family business to their offspring. High educated parents that transfer wealth are especially relevant for real founders who face more binding liquidity constraints. However, social capital is not always beneficial, especially for those who take over a family business. We found that children demand time and reduce the probability of taking over a family business. Furthermore, coming from a large family reduces the chances of taking over the business. Having siblings increases competition for the succession of the family business.

Results of previous (empirical) research that can not differentiate between family self-employment and de novo self-employment must be reconsidered in the light of our study (e.g. Evans and Leighton 1989; Dunn and Holtz-Eakin 2000). First and foremost, the time dimension must be brought back in theorizing and econometric modeling of self-employment selection. We have shown that aggregating these two types of self-employment might off set the dynamic nature of self-employment. Aggregating taking over a family business and starting from scratch resulted in a constant hazard for a considerable period of our researched time span. Moreover, disentangling the different routes into self-employment alters the causal nature of some of the determinants of self-employment.

More significantly, the rejection of the static nature of self-employment opens new paths of inquiry. Studying the waiting time into self-employment bears close resemblance with studying the real options nature of investments. The real options nature of investments dictates that in the presence of irreversible costs, the uncertainty of self-employment success can make individuals reluctant to become self-employed. This may be especially true for real founders whose success changes are more uncertain and who face higher partially irreversible set up costs.

Chapter 5

The Limburg dataset

5 The Limburg dataset

5.1 *The survey*

5.1.1 History

The research on starting entrepreneurs in South-Limburg is embedded in the research institute Business Investment Research Center (BIRC) at the University of Maastricht²⁵. Investment is broadly perceived within BIRC. Within three subprograms the following investment are investigated:

- the dynamics of physical investments of industrial firms
- investments and disinvestments in human capital of large firms
- investments in the form of market entry

The research on starting entrepreneurs is part of the last subprogram. So the decision to become an entrepreneur is perceived as a decision to enter the market of entrepreneurs. Through this perspective, we combine research on the labor market, i.e. the market for entrepreneurs, with research on market entry and exit. After all, every creation of a new company is also the entry of a new player on a market.

Why did we choose to study market entry with the specific case of entrepreneurship? Although more than 90% of the Dutch businesses are considered small businesses, the majority of economic research focuses on middle and large scale businesses. So with this study we hope to enrich the knowledge about small businesses.

The social networks of our respondents are an important theme in this research. There is growing interest from practice and theory in this theme. Nearly no conversation on entrepreneurship with experts from the field and entrepreneurs themselves ends up at a certain point with the importance of the personal networks for entrepreneurship. Entrepreneurs frequently utilize their personal relations. Until now, the mechanisms behind the relationship between networking and successful entrepreneurship are hardly researched. This research is the first step to bridge this gap.

5.1.2 Aim

5.1.2.1 Research population

Starting entrepreneurs form the core of our main questions. Answering the question who becomes self-employed and who not requires that persons who are not self-employed are part of the research population. After all, without such a control group of non-self-

²⁵ This institute was founded in January 1997 by Professor Dr. Gerard A. Pfann and aims to provide new insights in investment decisions. The financial backbone of the institute is a PIONIER subsidy provided by NWO, the Dutch organization for scientific research, to Gerard A. Pfann in 1996.

employed it is not possible to show what differentiates self-employed from non-self-employed. The choice for self-employment is a conscious choice: people think carefully and thoughtfully before they start up their own business. Among the wage workers, there are certainly people who have thought about starting up a business; however there are also people who never have had the idea to start up a business themselves. For research purposes, it is problematic to interrogate individuals about decisions that have not come to their mind. In the evaluation of hypothetical situations people cannot base their answers on their own experience so that their answer often does not reflect their own opinion but what, according to them, others – in our case potential self-employed – think. So the ideal population for researching the determinants of self-employment consists of all individuals that have ever considered becoming self-employed.

5.1.2.2 Time span

In the research, potential self-employed entrepreneurs are investigated who considered becoming self-employed or who started up a business in 1998 or 1999. We choose this time period for two reasons. On the one hand, we were trying to prevent a large lap of time between the moment of interview and the establishment of the business, because people recollect events in the past badly. On the other hand, we had to safeguard that the recently started enterprises had at least had a chance to establish themselves on the market, in order to be able to subsequently evaluate which starting businesses could be considered successful and which until now less successful.

5.1.2.3 Choice for South Limburg

Within the Business Investment Research Center, research on a national scale had already been performed, like for example research towards self-employment on the Brabant dataset and research towards the dynamics of investment decisions of Dutch industrial firms. To stress the positive interplay between the University of Maastricht and the local region we decided to concentrate the research towards self-employed entrepreneurs in South Limburg.

From a more scientific point of view, a regional focus offers also the advantage that some macro-economic factors, which differ on a national scale, are for the research population constant. A regional focus provides us the opportunity to research the phenomenon self-employment at micro level free from macro influences.

5.1.3 Development and execution of the research

5.1.3.1 Execution of research

Because the execution of an empirical research would exceed the capacity of the Business Investment Research Center, it was decided to subcontract the fieldwork to a professional

Figure 5.1: Execution process

survey organization. After an intensive search and selection process, the Business Investment Research Center decided to outsource the fieldwork of the research to R&M in Maastricht and ProInfo in Heerlen.

The development of the questionnaire covers the first part of the execution process (see Figure 5.1 for a schematic description of the complete execution process). In the initial stage, we have had consultations with experts from the field. The aim of these conversations was to obtain a better picture about what is going on in the field and what are important practical issues of starting entrepreneurs. Next to the recently started entrepreneurs, we had conversations with individuals who worked with starting entrepreneurs on a daily basis.

The employees of the Starterscentrum South Limburg acted as an important sounding board for us in order to not lose sight of the practical relevance. In the consultations with bank directors and accountants for example, we mainly informed ourselves about the financial and administrative aspects of starting a new enterprise. These conversations formed the basis of a first version of the questionnaire. Then, the first versions were discussed with several colleagues, who had ample experience with the execution of similar empirical research, and the remarks and comments were incorporated in a second version.

5.1.3.2 Execution process

When the questionnaire passed the test of our colleagues, it was tested in practice. For this we put the questionnaire to the field experts, who were consulted before, and to some recently started entrepreneurs, with the request for critical comments and we invited them in a so-called focus group discussion. On two days, we went through the questionnaire and requested them for a general impression. Then, we discussed about the clarity and clearness of each question and the accompanying answer possibilities. The many suggestions from these focus group discussions were then incorporated in a third version of the questionnaire, the concept questionnaire.

From part two onwards, the selected research organizations R&M and ProInfo were closely involved. The concept questionnaire was once more put through thoroughly. Until then, the concept questionnaire only existed in paper form and now this paper version was translated into an electronic version, so that the interviewers could make use of a laptop at the interviews. In technical jargon, such an electronic questionnaire is called a CAPI-questionnaire, where CAPI stands for “Computer Assisted Personal Interview”. This first version of the CAPI-questionnaire was tested in the interview laboratory of R&M. Once more, three starting entrepreneurs were asked to answer an experienced interviewer in the interview laboratory. After a thorough analysis of the course of the three trial interviews and the answer behavior of the three test subjects, a few more improvements were incorporated in the questionnaire. For example, we discovered that the interview lasted longer than the preliminary expected 60 minutes and that therefore some questions had to be deleted. In the meantime, R&M and ProInfo had already selected the interviewers for the research. The chosen interviewers, who mainly had experience with interviewing ‘professionals’ and companies, completed subsequently an introduction and training specifically aimed at this research on starting entrepreneurs.

At November 27 2000, the first interviews were conducted by R&M and ProInfo, which resulted in 135 interviews at the end of 2000. After Christmas the fieldwork was suspended for four weeks, on the one hand because respondents are hard to reach around the turn of the year, but mainly because we would analyze more specifically the first 135 interviews in order to correct possible problems in this relatively early stage of the research. Fortunately, only marginal changes were needed. For example, we discovered that more than 20% of the respondents were hesitant in providing financial data about their business. Therefore, we decided to offer the respondents the choice not to provide the exact sales but merely an indication. Between February 6 and May 23 2001, the remaining 1088 interviews were conducted (see also Figure 5.2). Only the fieldwork of this research took 6 months. This is clearly longer than normally in a commercial research period. Nevertheless, we choose consciously for a similar long research period, because this improves the response rate. By means of this long research period it is possible to approach respondents at a later time, if they are temporarily very busy.

Figure 5.2: Interview dates

5.1.4 Conduct

5.1.4.1 Response and non-response

Empirical research lives from the cooperation of the respondents. For making inferences from the collected data, it is of utmost importance that a lot of individuals are willing to cooperate in the research. However, a part of the approached individuals were for one or the other reason not willing to collaborate, which creates non-response.

5.1.4.2 Non-response in the Limburg survey

The population consists of two subpopulations, namely businesses which subscribed at the Chamber of Commerce in 1998 and 1999 and individuals in that period that had contact with the Starterscentrum because they intended to start up a business. The total size of these two subpopulations is 7.453 addresses. In 62 cases, addresses were recorded twice, so that the valid original population consists of 7.391 addresses. Of these 7.391 addresses, 5.119 were registered as starter (69%) and 2.272 (31%) as non-starter. A problem with this population is that it is not completely compatible with the by us desired population. As was indicated before, the research focuses on individuals who in 1998 or 1999 became self-employed by starting a business or who at least seriously considered it. During the telephone screening, we discovered that 1.350 (18.2%) did not comply with this constraint. Most of these companies did not belong to our target population because they were founded long before 1998. Furthermore this group consists of business who were not established by a natural person (for example a subsidiary company raised by DSM) or empty administration partnerships who were established solely for tax reasons and do not develop any economic activities. We end up with a final research population of 6.041 addresses.

During the telephone screening, it became apparent that some addresses did not correspond any more. By consulting several other information sources (telephone book, WebPages on the Internet) it was tried to retrieve current data. But also after this update 1.196 (19.8% of 6.041) addresses were untraceable. These 1.196 addresses were from the subpopulation starters and non-starters. We suspect that the main reasons for this fall out are: a) already quitted businesses b) people or businesses, which moved outside the community.

Furthermore, we could not reach 386 addresses because nobody answered the phone, although the telephone number had a connection. It should be noted, that the employees of the telephone screening tried on average 10 times at different times to reach anyone at that number. Finally, 4.459 addresses were reached for the screening.

The aim of the screening was to acquaint the respondents with the research and to ask for their cooperation for a, at their suitable time, personal interview. 1.323 respondents (21.9% of the original population and 29.7% of the reached addresses) preferred not to cooperate with the screening. 120 respondents were not able to cooperate with the research, because they had not mastered the Dutch language or passed away. Finally, we screened 3.016 individuals. 1.577 respondents promised during the screening to cooperate with the subsequent personal interview. These respondents received further information about the research and a confirmation of the appointment for a personal interview. After their initial promise, 354 of these 1.577 respondents renounced to cooperate with the research. Finally, we personally interviewed 1.223 respondents.

What do these numbers mean in terms of non-response? Based on the address database of the Chamber of Commerce and the Starterscentrum, we can assume that 6.041 individuals in 1998 and 1999 in South Limburg seriously thought about becoming self-employed or have become self-employed. These 6.041 addresses form our gross sample. We could not reach 1.856 addresses, because the known address was not valid (1.196), the respondents could not be reached after several trials (540), the respondents were seriously ill or passed away (30), or because the interview was not possible due to

language problems (90). So our net sample is 4.185 addresses, i.e. with 4.185 individuals successful telephone contact was established.

At the time we established successful contact, there exist a possibility that the respondent refuses cooperation. 1.087 respondents (26%) refused to collaborate to the preliminary telephone inquiry, 1.439 (34%) provided limited information via the telephone but preferred not to cooperate with the personal interview and 436 (11%) promised initially cooperation, however renounced at a later moment. 1.223 (29%) spoke to our interviewers in a personal interview. In terms of non-response, this means that of the established contacts, 29% were willing to cooperate to this research and 71% refused it explicitly at some time of the research. Of the 1.223 successful interviews, 841 (69%) were registered as starter and 382 (31%) as non-starter. So the ratio starter and non-starter among the respondents is exactly the same as in the address database. However, during the telephone screening and the interviews, we encountered that 159 registered non-starters were by then starters and 13 non-starters originally were registered as starter. This displacement between the information of the registration and our information from the contact of the respondents is shown in Table 5.1. Finally we interviewed 987 (80.7%) starters and 236 (19.3%) non-starters. These last numbers are an indication of the earlier expressed suspicion that non-response among non-starters is higher.

Table 5.1: Displacement starters and non-starters

		Starter according to interview				
		No	Yes	Total		
Registered Starter	No	223	159	382		
		(58)	(42)	(100)		
		[94]		[16]	[31]	
	Yes	13	828	841		
		(2)	(98)	(100)		
			[6]	[84]	[69]	
Total		236	987	1223		
		(19)	(81)	(100)		
			[100]	[100]	[100]	

5.2 Survey follow up: Matching Limburg dataset with Chamber of Commerce dataset in 2004

One of our aims was to determine the success of the decision to become self-employed. However, as we were interested in individuals who started their business in 1998 or 1999 and that our interviews were conducted from November 2000 till May 2001, the success of the self-employment decision was hard to measure with the existing data. Some of our

targeted firms were only operating for less than a year. Therefore, we decided to conduct a follow up research with our target population.

For this follow up research, we decided to contact the Chamber of Commerce and asked for their cooperation in the research. The staff at the Chamber of Commerce allowed full access to their databases. Therefore, we were able not only to obtain the status of the firm in May 2004 of our 987 registered starters, but also the status of the complete population of registered starters in 1998 and 1999. After a thorough clean up of the database, the new updated database consisted of 5.822 registered starters. Note that this database is somewhat larger than our initial database of starters (5.119). This discrepancy is possible due to late registrations, which the Chamber of Commerce allows for. The Chamber of Commerce database allows for several reasons for deregistration. Table 5.2 shows the reasons for deregistration and their frequencies. Finally, we matched our database with the newly created database to obtain the status of the firm of the interviewed self-employed.

Table 5.2: Deregistration reasons

	Code	Population (N=5822)	%	First Time (N=619)	%	Self- Employed Before (N=184)	%
Deregistration due to rectification	21	10	0.45				
Deregistration by virtue of the Chamber of Commerce	24	260	11.58	7	8.43	3	6.82
Deregistration due to change in legal form or economic activity	25	3	0.13				
Deregistration due to discontinuance (general)	27	1651	73.51	65	78.31	28	63.64
Deregistration due to discontinuance by bankruptcy	28	105	4.67	6	7.23	7	15.91
Delayed deregistration	29	128	5.70	2	2.41	2	4.55
Unknown	31	7	0.31				
Deregistration due to economic transfer	32	74	3.29	3	3.61	4	9.09
Unknown	91	8	0.36				
Total		2246		58		44	
Discontinuance rate by May 2004		38.58%		13.78%		23.91%	

5.3 Some (descriptive) results from the data

In this part of the chapter, the (potential) starter is the focus. Here we would like to answer what differentiates individuals that have started a business from those who have considered becoming self-employed, but finally denounced.

5.3.1 The self-employment decision

5.3.2 Starting entrepreneur versus non-starting entrepreneur

5.3.2.1 Demographics

In this section we describe the differences in objective individual characteristics between starters and non-starters. 82% of the interviewed men have started a business and 80% of the women. So there exist no substantial difference between starters and non-starters with respect to gender. Research among the Dutch labor force shows that woman are less inclined to become self-employed. This figure also emerges from our data. 910 (74.4%) of our respondents is man and merely 313 (25.6%) is woman. These results indicate a two step process. The first step describes if somebody seriously considers self-employment at all and it turns out that man consider this option more often. However among the group of individuals who have seriously considered starting up their own business, an equal portion of man and woman truly starts up a business.

At the time of the interview, the average age of the starter and the non-starter is both 40 years. Figure 5.3 indicates that age is normally distributed for both the starters and the non-starters. Comparing the youngest age category with the oldest age category, we see that young individuals are more likely to abandon the idea to become self-employed.

Table 5.3: Cross-tabulation marital status and self-employment status

Marital Status	Starter	Non-Starter
Married or unmarried living together	78	76
Single or divorced	22	24

Finally, we have data about the marital status of the respondents when they started up a business. The percentages in Table 5.3 clearly indicate that the marital status of starters and non-starters do not differ. About three quart of the starters and non-starters is married or lives together.

Figure 5.3: Age distribution

5.3.2.2 Social background

Previous research among the labor force indicated unanimously that individuals with self-employed parents choose more often to become self-employed themselves. In this research we choose to interview people who at least seriously considered self-employment. In our sample, more than one third (35%) of the respondents are from a family with self-employed parents. In national cross sectional samples this percentage never exceeds 10%. This reveals that individuals with self-employed parents also sooner consider self-employment at all. In our sample, 83% of the respondents with self-employed parents choose for self-employment and 80% of the respondents without self-employed parents choose for self-employment. This difference is not significant (at the 5% level) and therefore the choice for self-employment is not influenced by the fact that the parents are self-employed. So given that somebody seriously considered self-employment, the social background in terms of self-employed parents does not influence the self-employment decision.

Another characteristic of the social background is the educational background of the parents. The average educational level of the parents of starters and non-starters is lower vocational schooling. Furthermore, it becomes apparent from this sample that in the last decennia education expansion has taken place (see Table 5.5 for educational level starters and non-starters). Also other research shows that in the Netherlands, but also in other countries, the educational level of the offspring (in our sample intermediate vocational schooling) is higher than the educational level of the parents (in our sample lower vocational schooling). Table 5.4 provides the distribution of the highest education of the parents for the starters and non-starters. In general, the differences between the two groups are not remarkable. The visual differences in the educational categories VWO and WO are based on small numbers, and not on a systematic difference.

Table 5.4: Distribution parental education

Educational Level	Total	Starter	Non-Starter
Basis	14.9	15.2	13.6
LBO	19.5	19.4	19.9
MAVO/HAVO	16.5	16.3	17.4
VWO	3.1	3.6	1.3
MBO	18.2	18.1	18.2
HBO	14.8	14.9	14.4
WO	5.2	4.8	6.8

5.3.2.3 Human capital

The term human capital stands for the knowledge, skills, and experience of an individual. During the course of life, everybody builds human capital by following education and through learning from experience. Everybody can use its human capital either in a regular job or to become self-employed. Further onwards in this chapter, we present the distribution of several specific elements of the human capital for starters and non-starters. Table 5.5 presents the distribution of schooling over starters and non-starters. Both for the starter as for the non-starter the median is at intermediate vocational schooling. There are only minor differences between the educational level of the starter and the non-starter.

Table 5.5: Distribution respondents over educational levels

Educational Level	Total	Starter	Non-Starter
Basis	4.4	4.0	6.2
LBO	12.8	12.9	12.8
MAVO/HAVO	14.9	15.3	13.3
VWO	3.0	3.4	1.3
MBO	27.4	27.0	29.2
HBO	27.5	26.5	31.4
WO	10.0	11.0	5.8

More than half of the respondents had vocational schooling or higher. This might indicate that schooling types, that besides general skills also teach more practical and direct applicable skills, promote the possibility for individuals to think about self-employment. Respondents with a finished HBO education are more likely not to choose

for self-employment than individuals with another education. We cannot trace this back to the better perspective of higher educated on the labor market. After all, we see that of respondents with a university degree (WO) a larger part of the potential self-employed actually do become self-employed.

Although some of the entrepreneurs (among our respondents 40%) are completely on their own, the establishment of a company often means that the self-employed is also employer. Therefore, the starting entrepreneur employs individuals who he has to supervise. We asked in our research whether the potential entrepreneur had leadership experience in his previous jobs. More than half (53%) of the respondents had leadership experience and less than half did not have had leadership experience (47%) in a previous job. It seems that leadership experience does not affect the probability of starting up a business. 55% of the non-starters and 53% of the starters had leadership experience; a non-significant difference. However, leadership experience has an effect on the type of business started. In Table 5.6 we cross tabbed leadership experience with the size of the company. From the table it is shown that starting entrepreneurs who had leadership experience in their previous job also have a higher probability of starting up a company with employees. The higher the number of employees in a firm the higher is the chance that the starter had leadership experience. Comparative research between the Netherlands and the US shows that in the Netherlands relatively few individuals choose for self-employment compared to the US. Besides several other factors, this is often explained by a less favorable entrepreneurial climate. This might become apparent in the fact that entrepreneurs do not get a second chance when their first attempt failed. In our research, we find that 20% of the respondents have been self-employed before and have gained useful experience for a renewed self-employment decision. From our data it becomes clear that people, who have been self-employed before, have a higher probability to start up a business than people without self-employment experience. Almost 90% of the respondents with self-employment experience established a business, compared to 79% of the respondents without self-employment experience.

Table 5.6: Cross tabulation leadership experience and number of employees

Leadership Experience	None	1	2	3-5	6-9	>9
No	34.7	36.0	12.4	8.8	4.3	3.8
Yes	26.2	34.0	11.1	12.4	7.1	9.2

5.3.2.4 Social capital

The social networks and the social environment of the potential self-employed played a central role in this research. In this section, we report some results on the importance of the social networks of our respondents. Both starters and non-starters indicated that they were mainly satisfied or very satisfied by the support they received from their direct social environment (see Figure 5.4). This indicates that in most cases the direct social environment is willing to support the starting entrepreneur.

Figure 5.4: Satisfaction support social network

5.3.2.4.1 Social network size

Figure 5.5 shows the network size for starters and non-starters. For starters the average network size is 4.83 and for non-starters is 4.53. At first sight these distributions of both groups look similar. This is partly due to the fact that the naming of the network contacts was heavily guided by the questionnaire. At a closer inspection, it is shown that starters have on average significantly larger networks than non-starters. We must be careful to conclude that individuals who are becoming self-employed have larger social networks. Individuals who are planning to become self-employed may have more contact with prospective suppliers and customers.

Figure 5.5: Social network size

5.3.2.4.2 Social network composition

Besides the size of the network, we have also investigated the composition of the social networks. We investigated the composition of the network along the following criteria:

- the proportion self-employed and non-self-employed
- proportion family members and non-family members
- proportion weak and strong ties

Figure 5.6 shows the information about the composition of the network with respect to the proportion of self-employed in the network. Other self-employed can be an important resource in the own network. Contact with other self-employed is useful for a starting entrepreneur for the following three reasons. First, other self-employed have at their disposal important information that is relevant for starting up a business. For example, other self-employed can provide advice to the starting entrepreneur concerning problems with defaulters or the treasury. Second, other self-employed have contact with third parties, who can help the starting entrepreneur. Finally, other self-employed have often similar norms and values, go through similar problems and challenges, and therefore form a natural ally.

Figure 5.6: Distribution of proportion of self-employed over starters and non-starters

Furthermore, our research shows that people with relatively many other self-employed in their network have a higher probability of becoming self-employed. The median of the share of self-employed in the networks is 0.5 for starters and 0.33 for non-starters and differs significantly. Four out of five respondents, who in the end have started a company, build a network in which at least a quarter of the members are self-employed. More than a quarter of the non-starters have no self-employed contact.

In second instance, we have investigated the extent to which family members are counted as part of someone's personal network. More than a quarter of the respondents has no family member in their network and in more than half of the cases the family

members consists of halve the network. Figure 5.7 shows that the differences between starters and non-starters with respect to the proportion of family members are much less pronounced than the proportion of self-employed. On average 32% of the network contacts of the non-starters are family members and of the starters this is 29%. Although the difference is small, it is significant. Furthermore, it is evident that people with a network with a high proportion of family members are often member of the non-starter group.

Figure 5.7: Distribution of proportion of family members over starters and non-starters

The theory of social networks differentiates between so-called weak and strong ties. Strong ties are in a different way useful than weak ties. Strong ties, like the partner and some very good friends, can provide emotional support and offer help in hard times. Furthermore, the strong ties form someone's frame of reference. The strength of weak ties is that they provide different information and also provide access to other unknown individuals. Because the circle of friends of strong ties looks like one's own circle of friends, strong ties offer much less really new information than weak ties. One can obtain this information also directly. Weak ties are often embedded in different circle of friends and therefore they provide access to new unknown information. An example of a weak tie is the old fellow student, who one encounters at a reunion.

A possible indicator of the strength of a tie is the frequency of contact. For Figure 5.8, we assumed that a strong tie is someone with whom a respondent contacted weekly or more often, while all contacts with a lower frequency are assumed weak ties. Both for starters and non-starters, two third of the network contacts are strong ties. So given these results, there seems no difference between starters and non-starters concerning the proportion of strong ties in the network.

Figure 5.8: Distribution of proportion of strong ties over starters and non-starters

South Limburg has many associations and participation is a possibility to build a social network. Furthermore, it is often suggested that the Limburgian economy is tightly interwoven with associational life. In order to test this suggestion, we asked on a five point Likert scale to what degree starters and non-starters participate in club life. The participation rate for starters is 2.34 and for non-starters at 2.76. Non-starters are significantly more active in the Limburgian associational life. Figure 5.9 provides an explanation. Almost 50% of the respondents are member of a club and are active in the club. But the active non-starters participate more strongly than starters. It should be noted though, that this difference bears on starting entrepreneurs. Exactly the starting up of a company is time consuming so that other activities sometimes suffer.

Figure 5.9: Participation in associational life

5.3.2.5 Starting up: Motives

Table 5.7 classifies the motives for starting up a business. The cells indicate the percentage of the respondents that think a certain motive is important. Becoming self-employed is for many a big challenge and so more than two third of the respondents indicated that the new challenge was an important motive. Almost half of the respondents indicated that they always wanted to become self-employed and that this wish was an important motive for the transfer to self-employment. It strikes that these reasons are more important for starters than non-starters. About one third of the respondents bring forward that they either saw a gap in the market or thought that they could offer a better product or service than was currently available on the market. The availability of certain resources was also important for some of the respondents. 30% of the respondents indicated that the availability of clients or orders promoted the start up and 18% indicated that the sufficient availability of capital was an important motive. Another important motive is that self-employment better suits the combination of working and taking care of the family. People that think of becoming self-employed out of these motives, also do become self-employed.

Table 5.7: Start-up motives

Start Up Motive	Total	Starter	Non-Starter
Looked for new challenge	69.9	70.6	66.9
Self-employment was always goal in life	47.5	49.8	37.7
Could provide better work than market	36.5	36.9	34.7
Saw a hole in the market	32.6	36.4	33.4
Self-employment yield more money	29.7	28.8	33.5
Was secured of clients and orders	28.9	29.9	24.6
Combination of work and family	23.6	25.3	16.5
Had available financial means	17.3	18.1	9.7
Few opportunities at former employer	13.2	13.6	11.9
Acquaintance/ colleague wanted to start together	10.3	10.4	9.7
Threat of losing former job	7.4	7.2	8.1
Self-employment is family tradition	7.0	7.2	5.9

5.3.2.6 Satisfaction

In the meantime, time lapsed between the decision to become self-employed and the moment of interview. Therefore, we asked them how satisfied they look back at the decision to start or not to start a company. On a five-point Likert-scale the average satisfaction of the non-starters is with 3.15 significantly lower than those of starters with 4.30. Figure 5.10 shows the satisfaction distribution of starters and non-starters. People who have actually set up a business are much more satisfied than people, who have

abandoned this idea. Four out of five starters is satisfied or even very satisfied with their decision to become self-employed.

Figure 5.10: Distribution of satisfaction

Beside the subjective observation of the respondents, we have also investigated to what extent the financial situation of the starters and non-starters have improved. In Figure 5.11, we compare the income ahead of the self-employment decision with the income at the time of the interview. This means that we compare the starter's previous income from paid labor with their income as self-employed. It should be noted that comparing the incomes of self-employed with non-self-employed is extremely difficult. On the one hand, the income of employees is often augmented by secondary payments, like a thirteenth month salary, holiday allowance, and obliging costs for child care. Furthermore, some employers pay for pension build-up and other social insurance. On the other hand, self-employed have more margin to reduce their income tax by private expenses. Furthermore, it is known that self-employed work more hours per week than the average employee. Although the yearly income of the self-employed may be higher, then it is still possible that its hourly wage is lower. In this research we focus on the personal yearly income and omit possible differences in working hours.

Non-starters improved yearly on average Dfl. 5.000 and starters on average Dfl. 5.850, but this difference is not significant. The differences between starters and non-starters are not so much in the average income growth but more in the distribution. The income change of starters varies more strongly. For a quarter of it diminishes with 25%, but in contrast, a third experiences an income growth of more than 25%. Almost half of the non-starters experience no income change. For a fifth, it grows with 5 to 25% and another fifth of the non-starters a growth of more than 25%. Only a seventh of the non-starters experience an income reduction of more than 5%. From this result it becomes clear that self-employment is more risky. After all, half of the starting entrepreneurs experience large income differences both in the positive as in the negative direction.

Figure 5.11: Income distribution

5.3.3 Entry and exit timing

Besides the interest in who becomes self-employed and who survives as an entrepreneur, we are interested in the dynamics of these decisions. Therefore, this section investigates when an individual becomes an entrepreneur in his lifecycle for the first time and how long his or her first company survives. Since we only have the exit information of firms that started in 1998 or 1999, we will do this exercise only for the 803 firms that subscribed in 1998 and 1999. For this we split our sample of starters in first time starters and non-first time starters. 52.4% of the starters become entrepreneur for their first time in their life. In the following section, we provide non-parametric estimates of the time it takes before individuals decide to become self-employed²⁶.

5.3.3.1 Non-parametric entry hazard

It is interesting to observe the enormous variety of ages at which individuals decide to become self-employed. It becomes clear from the data that this choice is not made at the beginning of the career. The average age of the individuals that became self-employed for the first time within the time window 1998 and 1999 is 39. For individuals with self-employment experience the average age is 44. The youngest age at which an individual starts its first business is 21 years and an entrepreneur can be as young as 25 when he starts either a second business or tries self-employment for the second time.

Figure 5.12 shows the non-parametric Kaplan-Meier entry hazard for the first time self-employed and the experienced self-employed. The entry hazard shows retro-

²⁶ For the validity of our approach, we assume that the economy is in a steady state.

spective when the group of entrepreneurs has the highest probability to become self-employed over their lifetime. For first time self-employed, the hazard of becoming an entrepreneur is highest at age 50.

Figure 5.12: Non-parametric entry hazard (in months)

5.3.3.2 Non-parametric exit hazard

When an individual starts up a business, it is possible that the firm performs below expectations and the individual decides to exit self-employment or goes bankrupt. At May 2004, 38.6% of the in 1998 and 1999 created businesses do not exist anymore. For the first time starters in our sample this percentage is 13.8%. There exists a large and significant difference between the probability of exit of the firms in the sample and the firms in the population. There are several reasons for this. As was mentioned before, some may have been forced to exit before we could contact them. As business owners have often their business at home, business exit sometimes implies that the family has to move either because they sell the adjoining store or because the new financial situation forces them to live in a smaller home. Also, when the business is performing below expectations, people may be less motivated and enthusiast to tell about their business. Well-performing starters are often proud and enthusiast to tell about their business and are therefore more likely to be willing to participate.

It is well-known that probability of exit of a firm shows an inverted U-shape over time. Initially, the firm faces a low probability of exit. There are abundant financial resources and the entrepreneur knows that it takes time to learn how to manage a company. After a few years, the entrepreneur may find out that the business performs below expectations and that exit is the only option. During this time the probability of exit increases. Afterwards the probability of exit declines because only the able entrepreneurs remain in business. Figure 5.13 shows the exit hazard for the whole population of entre-

preneurs till May 2004. One clearly observes the inverted U-shape relation over time. The probability of exit peaks at 500 days.

Figure 5.13: Non-parametric exit hazard (in days)

5.4 Conclusion

In this chapter, we introduced the Limburg dataset. The Limburg dataset consists of detailed information on labor market history, social background, and social networks of 1223 individuals that seriously considered becoming self-employed around 1998 and 1999. In 2004, this dataset was matched with a dataset of the Chamber of Commerce that provide information about the status of the self-employment decision. In this chapter, we have provided some initial summary statistics of both datasets concerning the self-employment decision.

Chapter 6

Non-response bias and business idea discontinuance rates

6 Non-response bias and business discontinuance rates²⁷

6.1 Introduction

The purpose of this article is to determine whether using only survey respondents produces biased results and determine the extent of this bias. In particular, we look at the bias in estimating the probability that an individual interested in becoming a business owner, abandons the idea of becoming self-employed. Unbiased estimation of the discontinuance probability is only possible when discontinuing the idea to become self-employed is equal among respondents and non-respondents. There are good reasons to believe that an individual that abandons the idea of becoming self-employed would be less likely to respond to the survey than an individual who pursued the idea of becoming self-employed. The death, illness or negative emotions and reduced enthusiasm associated with the decision to abandon the business idea serve as examples.

Previous research has shown a causal link between non-response behavior and the variable under study (dependent variable); for business turnover rates (Holmes and Schmitz 1996) and for unemployment duration (Van der Berg et al. 2005). The bias from ignoring non-response can be estimated by combining survey/ administrative data (Van der Berg et al. 2005; Wang et al. 1992), using the number of call-backs in telephone surveys (Potthoff et al. 1993), or use a structural model of the joint process of non-response and the process under study, like business turnover rates (Holmes & Schmitz 1996).

In this paper we investigate the sources of non-response. More specific, we investigate non-response bias directly by combining personal interview/survey data with telephone screening data, and indirectly by a structural model of the joint process of non-response and business idea continuance in the flavor of Holmes and Schmitz (1996). We deviate from the Holmes and Schmitz model in that we consider only single-owner firms or one owner of a multiple-owner firm. Holmes and Schmitz assume furthermore that non-response is independent from business turnover. We estimate a choice model that relaxes this restrictive assumption.

The empirical ordering (sample analogs) shows that individuals that abandoned the idea of becoming self-employed are less willing to cooperate at the different phases of the research. For example, if an individual decides to take part in the telephone screening the discontinuance probability is 15%. If the individual decides to offer more information during this telephone screening, the discontinuance probability drops below 12%. Finally, if the individual from the telephone screening decides to take part in the final survey, the discontinuance rate drops to 7%. The results from the structural model confirm that the discontinuance rate from the personal interview is biased downward, but the results of the respondents to the telephone screening are remarkably close to the non-respondents of the research. Furthermore, the result from a multinomial logit model indicates that non-

²⁷ This chapter is based on the paper “Non-response bias and business continuance rates” by Kok and Pfann (2006).

response is not independent with business idea continuance. Individuals that decide not to become an entrepreneur are less likely to participate in the survey.

The paper is organized as follows. Section 6.2 provides the data. The Limburg survey is introduced and a taxonomy of non-response is presented. Section 6.3 provides the research problem and gives a first look at the data. Direct evidence of discontinuance bias among respondents is given. Section 6.4 provides a simple model of the joint process of non-response and discontinuance. Section 6.5 investigates a sequential decision model. Section 6.6 concludes.

6.2 The data

6.2.1 The Limburg survey

The survey “Self-employment in the South of Limburg” was a sample drawn from the population of potential business starters in South Limburg in 1998 and 1999. The research population is extracted from two databases: (1) The registration database of the local chamber of commerce contains all new firms, who registered a business in South Limburg in 1998 or 1999. (2) The address database of the “Starterscentrum South Limburg” contains addresses of persons, who had serious plans to become self-employed²⁸. The total size of this population is 7453. 62 addresses were double. During the telephone screening we checked whether the self-employment decision took place between 1998 and 1999. 1350 addresses did not meet this criterion. Our final research population encompasses 6041 individuals.

The research agency attempted to contact the whole population. First contact was established by telephone. During the telephone screening, individuals were informed about the research and asked for their cooperation in a personal interview at a date of their preference. 4472 addresses were reached for the telephone screening. And 3016 individuals completed the telephone screening. Each individual that completed the telephone screening was asked “What happened to your idea to become self-employed?”. 2084 out of the 3016 respondents also provided information about some demographics like, age, education, gender, and parental self-employment. 1577 individuals initially gave permission to cooperate with a personal interview. However, 354 individuals decided afterwards not to partake in the personal interview, leaving 1223 completed surveys.

6.2.2 A taxonomy of non-response

Table 6.1 provides a list of the 25 non-response categories as coded by the survey agency for the telephone screening. Figure 6.1 provides a taxonomy of screening sample non-response. We grouped these 25 non-response categories into four types. Two types of non-response are based on the fact that no direct contact was possible by the survey

²⁸ The Starterscentrum is a foundation sponsored by the chamber of commerce and other local authorities to encourage successful self-employment by providing training and advice.

agency, and the other two types are based on possible direct contact. We distinguish between non-response due to the fact that the individual has moved residence (labelled as MOVED) or due to the inability of the survey agency to contact the individual (NO-CONTACT). Non-response due to refusal of the individual to cooperate with the telephone screening (REFUSAL) and non-response due to other reasons, like illness or language problems (OTHER).

Table 6.1: Non-response categories

Variable	Code	# individuals
Appointment	1	20
Engaged	2	1
No reply	3	18
Quota full	4	1
Refusal	5	281
Terminated	6	30
Uncertain appointment	21	62
2 x no reply	22	37
3 x no reply	23	17
4 x no reply	24	18
5 x no reply	25	122
Answering machine	26	72
Voicemail	27	46
Fax/ modem	28	43
No connection	41	875
Refusal: no interest	42	620
Wrong address	43	321
Long-term ill/ death	44	23
Too old/ dement	45	7
Language problems	46	90
Gone during research period	47	154
Request out of sample	49	109
Too many approaches	50	47
Secret number	51	11
Completed	13	3016
Total number of individuals		6041

Figure 6.1: Flow diagram of non-response taxonomy

Our parameter of interest is the probability that an individual abandons the idea to become self-employed, shortly, discontinuance rate. We constructed this variable from the question: “What happened to your idea to become self-employed?”. Table 6.2 provides the distribution of answers to these questions for the telephone screening sample. The discontinuance rate is defined as the number of individuals that abandoned the idea to become self-employed relative to the total number of individuals that answered this question. We define the continuance rate as $1 - \text{discontinuance rate}$. For the telephone screening sample the overall discontinuance rate is 15.09 %.

Table 6.2: Distribution of business idea continuance

What happened to your idea to become self-employed?		
	Frequency	Percent
I am self-employed	2186	72.48
I am not self-employed, but will be in the coming six months	74	2.45
I am still in the orientation phase	301	9.98
I have abandoned the idea to become self-employed	455	15.09
Total	3016	100

On the basis of the responses to the business idea continuance question, we can classify the respondents in one of the four cell listed in Table 6.3. Cell 1 consists of individuals that could not be directly contacted (non-response category NO-CONTACT and MOVED). 1569 individuals of the 6041 individuals are in this category. Cell 2 consists of individuals that could be contacted, but whom did not partake in the telephone screening because they refused or were not able to fulfill the telephone screening (non-response category REFUSAL and OTHER). Cell 3 consists of individuals who respond to the telephone screening and reported that they continued their business idea. 2561 individuals are in this category. Finally, cell 4 consists of individuals who did also respond to the telephone screening; however they reported that they abandoned the idea to become self-employed.

Table 6.3: Business idea and response status: Cell counts in the Limburg telephone survey

Cell no.	Description of response	Number of individuals in cell
1	Non-response due to no contact	1569
2	Non-response due to refusal	1456
3	Respondent continued business idea	2561
4	Respondent discontinued business idea	455

Our research is concerned with the stability of the discontinuance rate at several stages of the research. In specific, we compare the discontinuance rate for the individuals that completed the personal interview (N=1223), completed the telephone screening (N=3016), and whom did not respond (N=3025). The next section presents the research problem and provides some direct evidence that the discontinuance rate differs among the samples. After that, the discontinuance rate from the non-respondents is estimated by a structural model for the joint probability of non-response and business idea continuance.

6.3 The research problem and some initial statistics

In this section, we present the research problem and notation along the work of Holmes and Schmitz (1996). We also provide evidence that individuals who have abandoned the idea to become self-employed are less likely to give permission for a personal interview than individuals that have not abandoned the idea, conditional on partaking in the telephone survey. Thereby, we provide direct evidence of a causal link between discontinuing the idea to become self-employed and survey non-response.

We follow the notation of Holmes and Schmitz (1996). Define the indicator variable x_i such that $x_i = D$ if individual i discontinues the idea to become self-employed and $x_i = S$ if his or her idea has survived the research time. Let the variable y_i be an indicator variable such that $y_i = R$ if individual i responds and completes the telephone screening and $y_i = N$ if i does not respond to the telephone screening. And let $\Pr(x, y)$ be the joint probability distribution of continuance and non-response.

Our ultimate interest is the probability that an owner discontinues his or her idea to become self-employed; that is $\Pr(x_i = D)$. How is the probability $\Pr(x_i = D)$ estimated from the sample? This unconditional discontinuance rate can be written as the weighted average of the discontinuance rate of the respondents and the non-respondents (Holmes & Schmitz 1996: 233), as

$$(6.1) \quad \Pr(x_i = D) = \Pr(y_i = R)\Pr(x_i = D | y_i = R) + \Pr(y_i = N)\Pr(x_i = D | y_i = N)$$

We have sample analogs of the probability of response and non-response, as well as the probability of discontinuance conditioned on response. We have no sample analog for the term $\Pr(x_i = D | y_i = N)$. Holmes and Schmitz argue that researchers often use $\Pr(x_i = D | y_i = R)$ as an estimate of $\Pr(x_i = D)$. If $\Pr(x_i = D | y_i = R) < \Pr(x_i = D | y_i = N)$, this estimate may be biased. If $\Pr(y_i = N | x_i = D) < \Pr(y_i = N | x_i = S)$, then $\Pr(x_i = D | y_i = R)$ will underestimate $\Pr(x_i = D)$ (Holmes and Schmitz 1996: 234).

In this section we provide some preliminary evidence that $\Pr(y_i = N | x_i = S) < \Pr(y_i = N | x_i = D)$. For this preliminary evidence, we use information on non-response in the population of respondents to the telephone screening. We assess the quality of the estimate $\Pr(x_i = D | y_i = R)$ for $\Pr(x_i = D)$? We address that question in Section 6.4 when we develop some models of the joint process of business idea continuance and survey response.

We introduce some extra notation to validate our claim. Let the variable v_i be an indicator variable such that $v_i = R$ if individual i gives permission to partake in the personal interview and $v_i = N$ if i is not willing to partake in the personal interview. Also let the variable w_i be an indicator variable such that $w_i = R$ if individual i is willing to provide extra information about their personal demographics in the telephone screening and $w_i = N$ if individual i is not willing to provide more information during the telephone screening. Note, that we have sample analogs of these probabilities.

What would be the bias, if we did not go to the trouble of asking the discontinuance status during the telephone screening, when respondents decided not to partake in the personal interview? So, how well is $\Pr(x_i = D | v_i = R)$ a substitute for $\Pr(x_i = D | y_i = R)$ and finally for $\Pr(x_i = D)$?

6.3.1 Sample analogs discontinuance rate

We do some analysis to validate our claim that discontinuance is related to non-response at subsequent stages of cooperation during our research. Table 6.4 shows substantial and significant differences in the discontinuance rate and explanatory variables in the different samples. The individuals that give permission to partake in the personal interview have a substantial lower discontinuance rate than the individuals that do not partake in the personal interview. Table 6.4 provides the sample analogs of the probability of discontinuance $\Pr(x_i = D)$ conditional on the different samples in our research and provides a lower and upper bound of the population discontinuance rate, $\Pr(x_i = D)$. Table 6.4 makes clear that using the discontinuance rate of the personal interview $\Pr(x_i = D | v_i = R)$ as an analog of the population discontinuance rate $\Pr(x_i = D)$ will always result in a downward bias in this sample.

In summary, we have reported the following facts from the Limburg dataset. First, an individual is more likely to non-respond in the phases of the survey if he discontinued his idea to become self-employed. Second, the survey discontinuance rate is always biased downward. These two facts together suggest a connection between an individual's discontinuance event and the probability that the individual fails to respond to the survey. The next sections make a first attempt to uncover this connection by explicitly modelling the process and estimating the parameters of this process.

6.4 A simple model of survey non-response and discontinuing the business idea

This section introduces our simplest model of the joint process of continuing the idea to become self-employed and survey response along the work of Holmes and Schmitz (1996).

Table 6.4: Discontinuance rate at different phases of survey

Sample Name	Survey – Personal Interview	Telephone Interview + Demographics	Telephone Interview	Population
Sample Size	1223	2084	3016	6041
No of Discontinuance	85	251	455	
$\Pr(x_i = D \mid v_i = R)$	0.0695			
$\Pr(x_i = D \mid w_i = R)$		0.12		
$\Pr(x_i = D \mid y_i = R)$			0.15	
$\Pr(x_i = D)$				?
Lower bound (455/6041)				0.0753
Upper bound ((3.025+455)/6041))				0.5760
Discontinuance probability conditioned on				
Not responding	$\Pr(x = D \mid v = N \cap y = R)$	.207		
Responding	$\Pr(x = D \mid v = R \cap y = R)$	.070		
Overall	$\Pr(x = D \mid y = R)$	.151		
Bias from ignoring non-response		.081		

6.4.1 Determination of business idea status

Each individual in the population of prospective business owners dies with probability α . If an individual dies, the individual discontinues his idea to become self-employed with probability 1. We refer to an individual who does not die as a healthy individual.

We model a healthy owner’s decision to continue self-employment or not by assuming that it is based on an expected entrepreneurial ability index z . The entrepreneurial ability index z takes on two levels, $z \in \{1,2\}$, with higher z meaning that it is more desirable to continue the idea of becoming self-employed. Specifically, we assume that, if an individual draws $z = 1$, then becoming an entrepreneur is unprofitable for the individual. At the other extreme, if an individual draws $z = 2$, then the individual thinks his

business idea is profitable. Let w_z be the probability that an individual draws an expected profitability index z .

There is now sufficient structure in place to calculate the probabilities of continuing the idea to become self-employed. Let us calculate these probabilities for a given health status of an individual. If an individual dies, then the individual discontinues his business idea with probability 1. Next, suppose the individual is healthy. The healthy individual continues his business idea if he or she draws $z = 2$. Hence, the healthy individual continues his business idea with probability w_2 and discontinues with probability w_1 .

Let ϕ_x be the probability that the individual experiences event x , conditioned on being healthy, where $x = C$ denotes the event that the partner continues, and $x = D$ if the individual discontinues. We have then that

$$(6.2a) \quad \phi_C = w_2$$

$$(6.2b) \quad \phi_D = w_1$$

6.4.2 Determination of response status

Regarding survey response, individuals who die do not respond to the survey. We assume that a healthy individual does not respond to the survey with probability θ . We follow Holmes and Schmitz and assume non-response independent of the business idea continuance or discontinuance.

6.4.3 Joint determination of continuance and response status

We have now completely described the joint process determining business idea continuance and survey response. We can now calculate the probability of each of the events in the cells of Table 6.3. First, we only model any response versus non-response. The probabilities of each of the different events are:

$$(6.3a) \quad f_1 = \alpha\theta w_1 + \alpha(1-\theta)w_1 + \alpha\theta w_2 + \alpha(1-\theta)w_2 + (1-\alpha)\theta w_2 + (1-\alpha)\theta w_1$$

$$(6.3b) \quad f_2 = (1-\alpha)\theta w_2 + (1-\alpha)\theta w_1$$

$$(6.3c) \quad f_3 = (1-\alpha)(1-\theta)w_2$$

$$(6.3d) \quad f_4 = (1-\alpha)(1-\theta)w_1$$

Consider the event $y_i = R$, $x_i = C$, which corresponds to cell 3 in Table 6.3. This occurs if the individual is healthy, the individual continues the idea to become self-

employed, and the individual responds to the survey. The probability of this event is given in equation (6.3c). The event $y_i = R, x_i = D$ corresponds to cell 4 in Table 6.3 and equation (6.3d) and follows the same reasoning except that the individual does not continue its idea to become self-employed.

Cell 1 and 2 in Table 6.3 are the non-response categories and both correspond to the events $Y_i = N$. These two cells differ in one important respect. For the event in cell 2, we know that the individuals were healthy, because they refused to partake in the telephone screening. This event occurs if the individual is healthy, but refuses to provide any information concerning his business idea status, so he non-responds.

For the individuals in cell 1, we do not know why these individuals non-responded. Either, they died and were not able to respond, or they did not want to respond. In the last case, this would amount to the same situation as the event in cell 1 of Table 6.3. However, this probability is augmented by the individuals that have actually died from either the entrepreneurs that would continue or discontinue their idea, and from either the ones that would or would not respond to the telephone screening.

The model summarized in equations (6.3a) – (6.3d) has three parameters - α, θ , and, w_2 . (Note that w_1 is determined from $w_1 = 1 - w_2$.) The model can be written as a general multinomial response model with m alternatives, indexed $j = 1, \dots, m$. Then the response probability is:

$$(6.4) \quad P^j = f^j(\alpha, \theta, w_2)$$

Let y_j be an indicator function with $y_j = 1$ if response j is observed and $y_j = 0$ otherwise. We have an independent random sample with observations (y_t) for $t = 1, \dots, T$. The log-likelihood for observation t can then be written as:

$$(6.5) \quad l(y_t, \alpha, \theta, w_2) = \sum_{j=1}^m y_{jt} \ln f^j(\alpha, \theta, w_2)$$

Assuming the disturbances to be independent identical type I extreme value, we estimate this model as a multinomial logit with Biogeme 1.2 (Bierlaire 2003)²⁹. Table 6.5 presents the parameter estimates of the model.

The probability of death is .019. This is a reasonable estimate within a two-year time period because the yearly death rate in the Netherlands is 8.68 per 1000 inhabitants. The probability of not-responding to the telephone survey for a healthy individual is .385. A considerable difference if one regards the sample non-response probability of .500. The probability that an individual perceives his entrepreneurial ability too low to continue his business idea is .137. Once a healthy individual has visited either the Chamber of Commerce or the Starterscentrum there is a probability of .863 that he will pursue his idea, at least for our research period.

²⁹ Empirically, we estimate the model with six parameters and three non-linear constraints.

Table 6.5: Simple model parameter estimates

Parameter	Formal Notation	
Probability of Death	α	.019 (.009)*
Not responding	θ	.385 (.010)**
Low ability	w_2	.137 (.041)**
Discontinuance probability conditioned on		
Not responding	$\Pr(x = D \mid y = N)$	.1533
Responding	$\Pr(x = D \mid y = R)$	.1509
Overall	$\Pr(x = D)$	.1521
Bias from ignoring non-response		.0012

+ significant at 10%; * significant at 5%; ** significant at 1%

With the parameters of our model, we can calculate the probability that an owner discontinues his idea to become self-employed conditioned on not responding to the telephone survey:

$$(6.6) \quad \Pr(x_i = D \mid y_i = N) = \alpha + (1 - \alpha)w_1$$

since an individual discontinues his idea either by dying with probability one, or if he stays healthy he discontinues his idea with probability w_1 . This result is reported in Table 6.5. The probability of discontinuance conditional on not-responding is slightly larger this probability for the respondents. The probability of discontinuance for non-respondents is .1533 and for the respondents it is .1509. There is only a marginal downward bias from using the discontinuance rate for respondents of the telephone survey to estimate the discontinuance rate for the entire population. Because non-respondents account for 50% of the population and because their discontinuance rate is .0024 higher than respondents, the overall discontinuance rate is .0012 ($=.50 \times .0024$) higher than the respondents discontinuance rate. This bias, .0012, from using the telephone screening sample to calculate the discontinuance rate is much lower than the bias, .081, from using the personal interview sample.

6.5 A sequential decision model

Figure 6.2 presents a sequential decision making model for the telephone sample with demographics ($N=2054$). We use this model to investigate the drivers of the process that leads to non-response. We differentiate between personal characteristics and a causal link between non-response and business (dis)continuance.

Figure 6.2 can be seen as a sequential decision flow. At the square node of the left, the individual decides to continue his business idea or to discontinue his business idea. At one point in time, the individual is contacted by the interview agency. Then conditional on the individual's choice to continue or discontinue his business idea, the individual decides to participate in the personal interview. As a result, we end with four events: the individual decides to continue his business idea and decides to participate in the personal interview ($x_i = C; w_i = R$), the individual decides to continue his business idea and decides not to participate in the personal interview ($x_i = C; w_i = N$), the individual decides to discontinue his business idea and decides to participate in the personal interview ($x_i = D; w_i = R$), and the individual decides to discontinue his business idea and decides not to participate in the personal interview ($x_i = D; w_i = N$). At the right extremity of Figure 6.2, we present the empirical counts of the number of individuals that choose one of these four events.

Figure 6.2: Sequential decision model

The model in Figure 6.2 can be written as a general multinomial response model with m alternatives, indexed $j=1, \dots, m$. Then the response probability is:

$$(6.7) \quad P^j = f^j(\beta_k | X_k)$$

where X is a vector of individual characteristics with associated parameter vector β .

Let y_j be an indicator function with $y_j = 1$ if response j is observed and $y_j = 0$ otherwise. We have an independent random sample with observations (y_t) for $t = 1, \dots, T$. The log-likelihood for observation t can then be written as:

$$(7) \quad l(y_t, \beta_k | X_k) = \sum_{j=1}^m y_{jt} \ln f^j(\beta_k | X_k)$$

Assuming the disturbances to be independent identical type I extreme value, we estimate this model as a multinomial logit. Table 6.6 presents the parameter estimates of this model.

The results in Table 6.6 show that there are differences among individuals in their response behavior. However, conclusions are difficult to draw from parameter estimates only in discrete choice models. In order to provide more evidence we report in Table 6.7 the partial effects that show the change in the probability of occurrence of an alternative to a change in an independent variable evaluated at the means of the independent variables.

The results from Table 6.7 shows that women have a higher non-response probability than men. However, this effect is only significant for respondents that decided to continue their business idea. The results show that being a woman reduces the probability of responding conditional on continuing business idea with .076 and increases the probability of non-responding with .063. This evidence suggests that female entrepreneurs are less willing to cooperate with a personal interview when they decide to continue their business idea.

Furthermore, the results show that older individuals are more likely to respond to a personal interview, irrespective of the decision to continue or discontinue their business idea. However, these partial effects are not significant.

The social background of the entrepreneur is also important to understand non-response. Individuals with self-employed parents are more likely to refuse to partake in the personal interview when they decide to continue their business idea. Coming from a self-employed background increases the probability of non-response with 0.55. However, this picture changes for individuals from self-employed backgrounds who decide not to continue their business idea. Although, these individuals with self-employed parents are both less likely to discontinue their business idea (as is reflected by the negative signs of the partial effects), the larger partial effect of self-employed parents on the probability of non-responds indicates that individuals from a self-employed background that do not continue their business idea are less likely to refuse to partake in the personal interview than individuals without self-employed parents. These results indicate that individuals with self-employed parents are only willing to partake in the interview if they decide not to continue the business idea.

Table 6.6: Estimation results sequential decision model

	$x_i = D; w_i = R$	$x_i = D; w_i = N$	$x_i = C; w_i = N$
woman	0.008 (0.259)	0.367 (0.183)*	0.330 (0.109)**
age	0.015 (0.012)	-0.011 (0.010)	-0.006 (0.006)
Education			
Primary school	-0.474 (0.626)	-1.896 (1.024)+	-0.787 (0.337)*
LBO	-0.600 (0.411)	-0.134 (0.294)	-0.110 (0.172)
MAVO, HAVO	-0.694 (0.412)+	-0.741 (0.342)*	-0.594 (0.183)**
VWO	-0.523 (0.752)	0.817 (0.372)*	0.558 (0.252)*
MBO/MEAO	0.048 (0.274)	0.264 (0.219)	0.183 (0.132)
University	-0.918 (0.459)*	-0.420 (0.324)	-0.147 (0.171)
Parents self-employed	-0.440 (0.252)+	-0.438 (0.190)*	0.184 (0.103)+
Constant	-2.789 (0.552)**	-1.471 (0.425)**	-0.492 (0.249)*
Observations	2054	2054	2054

Standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%
Reference category = $x_i = C; w_i = R$

Table 6.7: Partial effects

	$x_i = D; w_i = R$	$x_i = D; w_i = N$	$x_i = C; w_i = R$	$x_i = C; w_i = N$
Woman	-.0050 (.0093)	.0178 (.0131)	-.0757 (.0247)**	.0629 (.0234)**
Age	.0007 (.0005)	-.0007 (.0006)	.0013 (.0012)	-.0013 (.0012)
Self-employed parents	-.0170 (.0085)*	-.0310 (.0112)**	-.0066 (.0231)	.0546 (.0217)*
Education				
Primary school	-.0068 (.0201)	-.0611 (.0166)**	.1910 (.0559)**	-.1232 (.0522)*
LBO	-.0177 (.0106) +	-.0051 (.0185)	.0372 (.0372)	-.0144 (.0351)
MAVO, HAVO	-.0156 (.0112)	-.0309 (.0159) +	.0525 (.0373)**	-.0984 (.0322)*
VWO	-.0225 (.0142)	.0525 (.0373)	-.1347 (.0577)*	.1047 (.0568)
MBO/MEAO	-.0013 (.0102)	.0139 (.0154)	-.0453 (.0295)	.0326 (.0278)
University	-.0246 (.0097)*	-.0207 (.0172)	.0597 (.0371)	-.0144 (.0217)

Standard errors in parentheses

+ significant at 10%; * significant at 5%; ** significant at 1%

The partial effects of our education dummies provide a mixed view. Individuals with a low education are least likely to refuse to partake in the personal interview. This effect is both significant for individuals that did continue and did not continue their business idea. Compared with individuals with a vocational schooling, primary education reduces the probability of refusal when deciding to discontinue the business idea with .061 and .123 when deciding to continue the business idea. The lower educated are most likely to partake in the personal interview when they decide to continue their business idea. This might be attributable to the fact that lower educated perceive more value in cooperating in a personal interview. On the other hand, we can argue that individuals with a low education have a lower opportunity cost of time.

Finally, we investigate the structural relationship between non-response and the probability of continuing or discontinuing business ideas. In order to investigate this important relationship, we consult the alternative specific constants in Table 6.6. The alternative specific constant provides information concerning the rest preference for an alternative in the choice set. The results show all very significant alternative specific constants reflecting the differential propensity to choose a specific alternative. The result

show an interesting ordering in the preference of alternatives. Overall, the ordering shows that more individuals decide to continue than to discontinue their business idea. More interesting information can be extracted from the results when comparing response and non-response propensity between business idea continuers and business idea discontinuers. Individuals that decide to discontinue their business idea are more likely to refuse to partake in the personal interview. In contrast, individuals that decide to continue their business idea are more likely to partake in the personal interview. This evidence suggests that there is a structural relationship between non-response and the variable of interest in this study, namely continuing business ideas. Individuals that do not continue their business idea have a higher chance of non-response.

First, individuals are more likely to cooperate in a personal interview when they continued their business idea ($\beta = -.492$). In contrast, individuals that decide to discontinue their business idea are less likely to participate in a personal interview. These results together indicate that there is a structural relationship between continuing a business idea and non-response. Individuals that decide not to continue their business idea are more likely to refuse to cooperate in a personal interview. This structural relationship between non-response and the probability of continuing business ideas indicates that surveys samples are biased towards individuals that continued their business idea.

6.6 Conclusion

In this paper we analyzed the sources of non-response. Non-response is important because not accounting for non-response may produce biased estimates concerning the variable of interest, in our study the propensity of individuals to continue or discontinue their business idea. We have discovered that non-response in our sample is existent and important. We have found several sources of non-response.

The first source of non-response is the possibility of attrition, due to death for example, from the database at the time of sampling. Our structural model estimates that 1.9 percent of our population leaves our population because of attrition in a two year time period. The analysis shows that probability of discontinuance among the non-respondents is only slightly lower than the probability of discontinuance among the respondents. Furthermore, the results from the structural model show that a healthy individual has 38.5 percent chance of non-response. This is clearly different from the empirical probability of non-response of 50.0 percent.

The second source of non-response is a structural relationship between non-response and business idea continuance. We tested for this possibility on a subset of our sample, namely the individuals that did respond to the telephone screening. During the telephone screening these respondents provided demographic information and information concerning the status of their self-employment decision. Furthermore, we asked whether they would be willing to partake in a personal interview. We modeled the results of these questions as a sequential decision model. The results from this model showed a clear relationship between non-response and business idea continuance. Individuals that continued their business idea were more likely to partake in the personal interview, while individuals that abandoned their business idea less likely to partake in the personal interview.

The third source of non-response is personal characteristics. Woman and individuals from self-employed parents are less likely to respond to request to partake in the personal interview, while low educated individuals are more likely to participate in the personal interview. Differences in response propensity and business idea continuance can indicate that individuals trade off the costs and benefits from participating in a personal interview. And lower educated were still more likely to participate in the personal interview even when they did not continued their business idea. The opportunity cost of time of the less educated is lower. Individuals that did not continue their business idea are least likely to participate in the study. The benefits for participating are lowest for the non-continuers. However, the negative relation between business discontinuance and non-response can also be attributable to negative emotions associated with the abandonment of the business idea. The fact that non-respondents are different can be overcome by weighting techniques; however, the fact that non-response is casually related with the variable of interest, namely business continuance, can only be reduced by increased efforts of researchers in tracing the non-respondents.

Chapter 7

Reluctance to exit self-employment when a fraction of the set up costs is sunk

7 Reluctance to exit self-employment when a fraction of the set up cost is sunk³⁰

7.1 Introduction

Most researchers acknowledge that starting up a business is costly. Capital must be raised to cover investments in buildings, machines, and office equipment. When a business performs below expectations, the owner may decide to quit the business and sell the remaining assets. However, many discarded assets are scrapped rather than sold (Asplund 2000). Therefore, a business owner may accept a temporary loss and decide to remain self-employed and wait if business conditions turn favorable again to recoup partly sunk investment costs. To investigate the determinants and consequences of sunk costs for the creation and dissolution of businesses, this paper presents a model of self-employment entry and exit under uncertainty when a fraction of the set up costs is sunk.

The literature offers several frameworks to study entry and exit behavior of firms, depending on the unit of analysis. At the industry level, entry and exit is often studied in an integrated equilibrium framework, where firm entry and exit is determined simultaneously depending on the industry distribution of entrepreneurial abilities (Jovanovic 1982), active reinvestments and learning (Pakes and Ericson 1998), the changes in industry prices, demand, or some other characteristics of the process for firm shocks (Hopenhayn 1992). At the individual level, entry and exit decisions are seldom considered simultaneously because entry approaches differ from exit approaches. Entry is mainly modeled as a choice in an occupational choice framework, while exit is often perceived as a function of natural selection (Friedman 1953; Williamson 1991). One exception is Gimeno, Folta, Cooper, and Woo (1997) who argue that the survival of a firm is not exclusively a function of economic performance, but may also depends on individual performance threshold levels of the entrepreneur. They claim that entrepreneurs may disentangle their economic viable firms if income from self-employment is below the alternative wage paid in the labor market. This exit choice is also found in our data where only 5 percent of the exits are forced exits in terms of bankruptcy.

Models of entrepreneurial choice compare between self-employment and an alternative means of income such as paid labor (Borjars and Bronars 1989; Evans and Leighton 1989). These selection models ignore that starting up a business involves sunk costs. Asplund (2000) shows that 50-80 percent of the investments in metalworking machinery used in Swedish manufacturing industries are sunk.

In recent models of investments under uncertainty, the presence of sunk costs introduces an option value of waiting for new information and delaying irreversible decisions (Dixit and Pindyck 1994). These models combine the question of whether to invest (the “q” value) with when to invest (the option value). The option value of setting up a self-employed business is largely ignored in the theoretical and empirical self-employment models.

³⁰ This chapter is based on the paper “Reluctance to exit self-employment when a fraction of the set up cost is sunk” by Kok and Pfann (2006).

We believe that the timing of the self-employment decisions is important to understand who becomes self-employed and what the founding characteristics of the firm are. The decision to defer self-employment might increase the opportunity costs of becoming self-employed, e.g. the wage earned in the labour market. Models of entrepreneurial choice argue that individuals enter self-employment only if the value obtained from self-employment exceeds wage earnings. Rather than using the net present value rule the existence of sunk costs dictates that an option value model is more appropriate to analyze the self-employment decision (see Pindyck, 1991).

Our paper aims to extend the self-employment model by adding an option value to self-employment decisions when a fraction of the set up costs is sunk. In the resulting dynamic entrepreneurial choice model, the individual decides on the timing of entry and exit. It is instructive to integrate entry and exit decisions into one model, because the timing of the entry decision may reveal important information why starting up a business is delayed and how this information might influence exit behaviour. More specifically, we investigate the effects of uncertainty of future self-employment earnings streams, sunk costs, and previous wage levels on entry and exit thresholds.

Dixit and Rob (1994) argue that exogenous shocks, like technological progress, preference change, and policy shifts affect individual productivity and earnings over time. The uncertainty of future earnings for entrepreneurs is idiosyncratic, because an external shock can change expectations on self-employment earnings for business owners with different skills, strategy or technology. We follow Dixit and Rob (1994) and Parker (1996; 1997) in modelling the stochastic evolution of future self-employment earnings as a stochastic process.

We apply Dixit's (1989) general theory of entry and exit decisions under uncertainty to produce sharp predictions on how timing and founding characteristics may make entrepreneurs reluctant to exit self-employment. We pursue the simple idea that while comparing future expected income streams from different labour market choices (such as employment or self-employment) people do not only weight the expected value of future income streams, but also consider the option values associated with the uncertainty of income streams and sunk costs.

The theoretical model predicts that higher start up costs reduce the transition probability out of self-employment. In the empirical application, we measure heterogeneity in start up costs by investments in equipment, buildings, machineries, ICT etc. A second prediction derived from the model is that persons with steeper expected entrepreneurial income growth are more reluctant to exit self-employment. Productivity is measured as a function of personal human capital³¹. Further, the model predicts that idiosyncratic uncertainty about idiosyncratic self-employment earnings is inversely related to the decision to leave self-employment. In the empirical application, we utilize the knowledge of sunk costs and the timing of the entry decision to proxy idiosyncratic uncertainty. Since idiosyncratic uncertainty of self-employment income streams should delay the entry decision, the theoretical model predicts that conditional on (observed and unobserved) business owner heterogeneity, the ordering of entry times might reflect the underlying distribution of idiosyncratic uncertainty.

³¹ We follow Lazear (2005) that entrepreneurs should be Jacks-of-all-trade, which implies that entrepreneurs must be able to perform multiple roles.

Furthermore, the model predicts that the wage earned prior to entry increases the initial commitment in partially sunk capital investments. In the empirical application, we estimate the elasticity of previous earnings on capital investments at the time of entry. The structural relationship between prior wage levels and capital investment levels might explain the hitherto theoretically unexplained empirical finding that high prior self-employment entry wage levels reduce the probability of becoming self-employed (Amit et al. 1995; Evans and Jovanovic 1989; Evans and Leighton 1989)³².

The predictions generated by our theoretical model are tested utilizing a sample of de novo entrepreneurs who entered self-employment between 1998 and 1999 in South Limburg, a part of a province of the Netherlands. The results from the econometric analysis confirm the theoretical predictions. Sunk costs and idiosyncratic uncertainty about self-employment earnings reduce the probability of exit. Jack-of-all-trade entrepreneurs have steeper self-employment earnings growth and are therefore more reluctant to exit self-employment. The estimated elasticity of previous income levels to initial capital investments is .29. The positive sign of this elasticity implies an indirect structural effect of previous income levels on self-employment entry and exit via an increase in initial partially sunk capital investment.

The structure of the article is as follows. Section 7.2 presents the entrepreneurial choice model and explains why entrepreneurs enter the market on a different scale. Section 7.3 applies the real option model to the entrepreneurs exit decision problem. Section 7.4 presents the econometric model and presents the data. In section 7.5 the estimation results are given. Section 7.6 concludes.

7.2 Why do entrepreneurs enter on a different scale?

One of the predictions of Jovanovic (1982) learning model is that entrepreneurs enter the market at the same scale. The fact that entrepreneurs enter at the same scale is the result of the assumption that entrepreneurs have unknown entrepreneurial abilities before they enter the market. Since no entrepreneur has a priori more information about his success changes than another entrepreneur, Jovanovic model assumes that rational entrepreneurs should enter self-employment with a firm of equal size. His model does not account for the empirical regularity that firms enter the market with different scales. Figure 7.1 presents the frequency distribution of the size³³ (in log-form) of the firm at the time of entry of a sample of recently started first time entrepreneurs in the Netherlands in 1998 and 1999. The observed distribution of capital investments in a newly created business among entrepreneurs raises the following question: why do entrepreneurs enter at a different scale and what account for these differences?

³² It is noteworthy that Amit et al. (1995) suggest that future work should address the causality between income levels and the probability of entering self-employment. Our inclusion of partially sunk capital investments may provide the missing link between high income levels and low probability of entry.

³³ We proxy the size of the firm by capital investments.

Figure 7.1: Distribution of initial capital investment in firm (log-form)

Frank (1988) develops an intertemporal model of exit that account for different entry scales. He argues that individuals enter on a different scale because individuals have different expectations of success. Individuals that have higher expectations of success will enter on a large scale. As a consequence, these individuals with high expectations will be more reluctant to exit from the market, because it takes a longer series of adverse signals from the market before these entrepreneurs with high expectations about their entrepreneurial ability will adjust their subjective beliefs to their true entrepreneurial abilities and exit from the market. Frank's model has intuitive appeal and has also the advantage over Jovanovic model that Frank assumes that a part of the entry costs are sunk. And before investing these sunk costs, Frank asserts, the firm must expect that these sunken expenditures are justified. A weakness in Frank's model is the source of these expectations of entrepreneurial ability before entry. He argues that the source of these expectations is optimism. More optimistic individuals start larger firms. However, Frank assumes that after entry, all entrepreneurs, optimistic or not, face the same unknown distribution of entrepreneurial abilities and receive an unknown draw from the population distribution. Frank therefore seems to assume irrational behaviour of entrepreneurs because in his model there is no reason for optimism.

A more tractable solution to this problem can be achieved by allowing for individual entry thresholds within an entrepreneurial choice model³⁴. The mechanism of the entrepreneurial choice model is straightforward. Self-employment refers to an income generating mechanism that is different from employment in the labour market. Individuals choose the occupations with the highest expected payoff. Below, we will go through a simplified version of Dunn and Holtz-Eakin (2000) occupational choice model, and discuss the consequences of assuming that a fraction of the set up costs are sunk (Frank, 1988; Asplund, 2000).

³⁴ See e.g. Dunn and Holtz-Eakin (2000) for a recent application of the occupational choice framework to the self-employment decision. We will use their terminology.

The income of an individual working in the labour market consists of his wage and the returns on his assets. Let his wage be a function of its earning ability in a wage job w_i , assets A_i , and the net rate of return r . Then the individual's income as a wage earner is $w_i + rA_i$.

The income of the self-employed is more complex. The main difference between a wage earner and a self-employed is that the self-employed need to invest capital in order to earn income (Evans and Jovanovic 1989). For a self-employed entrepreneur, the income is a function of his gross entrepreneurial earnings and the returns on remaining assets after capital is invested in the firm. Let the individual's gross earnings as an entrepreneur be a function of his entrepreneurial ability θ_i , the amount of capital invested k_i , and some random component ε , summarizing the good fortune of the entrepreneur. Then, the individual's income as an entrepreneur is $\theta_i f(k_i) \varepsilon + r(A_i - k_i)$.

Until now, we have discussed the differences in the income generating mechanism. However, both incomes can be uncertain. Generally among researchers, it is assumed that entrepreneurial income is more volatile than the income as a wage earner³⁵ because (1) the uncertain evolution of business conditions³⁶, and (2) the uncertainty of the entrepreneurs ability or business idea³⁷. Most researchers summarize this information in the formulation of expected income streams. So that, the standard entrepreneurial choice models predict that an individual will opt for entrepreneurship if the expected utility is higher in the self-employment option, that is, if,

$$(7.1) \quad E\{U[\theta_i f(k_i) \varepsilon + r(A_i - k_i); Z_i]\} > E\{U[w_i + rA_i; Z_i]\}$$

Where Z_i is a vector of personal characteristics (e.g. marital status, children, education, social background).

Then, what determines capital investments, k , in the firm? If we follow Jovanovic (1982) assumption that θ_i is not known prior entry, and can be only learned ex post, this question can be easily resolved. Note that we know the value of the right-hand side of equation (7.1). This is the value of the income of the previous job. Naturally, this value differs among individuals because of standard human capital arguments. Therefore, if θ_i is unknown and equal among individuals, individuals can only raise their entrepreneurial earnings by raising their capital investments k . How much they have to raise their capital

³⁵ For simplicity, we will assume that the income as a wage earner can be known with certainty.

³⁶ This "objective" uncertainty rises because of productivity and price shocks. The parameters of the stochastic evolution are known to all entrepreneurs and possibly idiosyncratic (Dixit and Rob 1994). The specific path through this distribution is unknown and revealed only ex post. This uncertainty is often modeled as a stochastic process (see e.g. Dixit 1989; Dixit and Pindyck 1992 or Parker, 1996; 1997 for the application of this uncertainty to self-employment decisions). The presence of this type of uncertainty provides individuals an option to delay a decision and to wait for more information to arrive as time progresses.

³⁷ This "subjective" uncertainty arises because entrepreneurs learn only about their true entrepreneurial ability after they have entered (Jovanovic 1982). Since this uncertainty can only be reduced after entry, this type of uncertainty does not provide an option to delay the investment, because there is no chance to obtain information about one's true entrepreneurial ability other than experiencing. Individuals only know the population average entrepreneurial ability.

investment k to overcome the entry threshold, the right hand side of equation (7.1), depends on the income earned in the previous job. Since, the left hand side of equation (7.1) is increasing in k , we know that individuals with high previous income levels have to invest more capital k to raise their entrepreneurial income above the entry threshold level to make it profitable to become an entrepreneur and enter self-employment.

Table 7.1 provides an OLS regression of the entry size of the firm to a set of regressors including our parameter of interest: previous income levels³⁸. The results show that controlling for available debt and equity capital, previous income level is a very strong predictor of initial commitment in capital investments. The elasticity of income to set up costs is .294. Thus a one percent increase in the level of income prior to entry implies a .294 percent increase in initial capital investments. Therefore, individuals from high income cohorts have to overcome more set up costs to raise their expected income level from self-employment high enough to induce entry.

The question then is, what are the consequences for the exit behaviour of entrepreneurs of the theoretical prediction (Frank 1988) and the empirical finding (Asplund 2000) that at least a fraction of these initial commitments in set up costs are sunk? We will discuss the consequences of this assumption for the theory of exit in the next section.

7.3 The entrepreneurs exit problem under heterogeneous irreversible set up costs

From standard occupational choice models and equation (7.1) we know that an entrepreneur decides to exit self-employment if his income as an entrepreneur is lower than the income as a wage earner,

$$(7.2) \quad E\{U[\theta_i f(k_i)\varepsilon + r(A_i - k_i); Z_i]\} < E\{U[w_i + rA_i]; Z_i\}$$

Note that equation (7.2) assumes that k can be sold at the same price that capital investments is bought. What are the consequences for this model of the assumption that a fraction of the set up costs is sunken?

Hypothetically, if k includes a sunk cost component then the decision to exit self-employment includes an option to abandon the investment (Dixit, 1989). The entrepreneur knows that by remaining in business it can avoid incurring sunk costs for re-entry should future developments turn favourable. Therefore, he is willing to incur some current loss to preserve this option. The optimal exit strategy of the entrepreneur depends on entrepreneur-specific profit thresholds levels, Π_c .

³⁸ See Appendix 7.7 for descriptive statistics for variables.

Table 7.1: Explaining initial commitment in set up costs

	Log of Set Up Costs k	
	Estimate	Se
Explanatory Variables		
Income (log)	0.294	[0.118]*
Control Variables		
Debt capital (log)	0.127	[0.014]**
Equity capital (log)	0.105	[0.018]**
Education (base = MBO)		
Ed = primary school	-0.034	[0.451]
Ed = LBO	-0.213	[0.226]
Ed = MAVO/ HAVO	-0.208	[0.223]
Ed = VWO	-0.293	[0.400]
Ed = HBO	-0.376	[0.191]*
Ed = University	-0.003	[0.274]
Labour experience (months)	0.002	[0.001]*
Unemployment experience (months)	-0.005	[0.003]
Married at time of entry	-0.293	[0.193]
Self-employed parents	0.171	[0.143]
Self-employed parents-in-law	0.173	[0.167]
Female	-0.317	[0.212]
Only owner	-0.225	[0.175]
Social Network Size	0.012	[0.028]
Age	0.006	[0.010]
Industry Sector (base = business services)		
personal service incl. health and culture	0.339	[0.232]
manufacturing incl. utility	-0.114	[0.301]
hotels & restaurant sector	1.338	[0.577]*
construction sector	0.144	[0.224]
wholesale & retail sector	0.168	[0.206]
Constant	5.383	[1.245]**
Observations	334	
Model Specification Tests		
R-squared adjusted	0.30	
F(23, 310)	7.31**	

Standard errors in brackets

+ significant at 10%; * significant at 5%; ** significant at 1%

Dixit (1989) in general and Pfann in particular³⁹ (2001; 2006) show that these individual specific thresholds depend on the sunk fraction of set up costs qk , entrepreneurs specific profit uncertainty σ_θ , and expected growth of idiosyncratic profits θ . Where q is the fraction of set up costs that is sunk, with $0 < q \leq 1$ ⁴⁰.

Entrepreneurs with low exit profit thresholds will be more reluctant to exit self-employment and will have lower probabilities of exit and thus larger self-employment spells⁴¹. We will investigate these predictions in detail below and will discuss their empirical proxies in the next section⁴².

Prediction 1. $\partial \Pi_e / \partial qk < 0$

Figure 7.1 and Table 7.1 shows that sunk set up costs vary among entrepreneurs. An entrepreneur with high sunk set up costs is most reluctant to exit self-employment. This result compares to those of Dixit (1989) and Gimeno et al. (1997). Idiosyncratic sunk set up costs are path dependent. Income levels from previous jobs cause increased capital investments. Idiosyncratic set up costs are thus important for exit decisions. This is not the result of the sunk cost fallacy, but the consequences of rational decisions under uncertainty when fractions of the set up costs are sunk.

Prediction 2. $\partial \Pi_e / \partial \theta < 0$

The expected growth of idiosyncratic profits, θ , may vary across workers, and the entrepreneurs with the steepest profits growth will be most reluctant to exit self-employment. Note that we relate the expected growth of idiosyncratic profits to idiosyncratic entrepreneurial ability. This is a well-known result from Jovanovic (1982) learning model and Lucas (1978) general equilibrium model and is related to the fact that more able entrepreneurs with steeper growth rates are more likely to survive.

Prediction 3. $\partial \Pi_e / \partial \sigma_\theta < 0$

The idiosyncratic uncertainty of idiosyncratic profits, σ_θ , may vary across workers. The entrepreneur with the highest uncertainty will be the least likely to exit self-employment because the chance of higher profits increases the firm's expected future returns. The most risky entrepreneur is thus also the most reluctant to exit self-employment. This is an important result, because losses can be substantial before the entrepreneur is convinced to exit.

In summary, the theory of entrepreneurial exit with heterogeneous set up costs and uncertainty extends the earlier entrepreneurial choice models that did not account for the sunk nature of set up costs. The idiosyncratic firm abandonment option value results from the irreversible character of the set up costs and the future uncertainty of each entrepreneur's profit. A new insight is that entrepreneurs with high firm abandonment option values are less likely to exit self-employment when business conditions turn unfavour-

³⁹ Dixit's work focuses on industry wide thresholds and macro uncertainty, like price volatility, whereas Pfann's contribution stems from identifying idiosyncratic thresholds from idiosyncratic uncertainty related to e.g. productivity.

⁴⁰ Asplund estimates that for the manufacturing industry, q may be as high as 0.8.

⁴¹ Note that our model gives a rational explanation why underperforming firms remain (temporary) in business. If it is rational for underperforming businesses to remain temporary in business, then the available capital to remain in business in worse times becomes important too. This argument is related to the liability of adolescence.

⁴² We do not formally derive these thresholds but refer the interested reader to Pfann (2006).

able. An entrepreneurial enterprise with a high option value will be more reluctant to exit self-employment. In the empirical application, we try to identify the existence of the firm abandonment option value by observing the timing of the exit decision of entrepreneurs. Entrepreneurs with high option value are hypothesized to remain longer in business.

7.4 The econometric model and the data

Pfann (2006) shows that the non-linearity of the thresholds in qk , θ , and σ_θ makes it cumbersome to analyze every worker's option values Π_{e_i} . He suggests using linearized decision rules instead, and express Π_{e_i} in linear form as follows

$$(3) \quad \Pi_{e_i} = \frac{\partial \Pi}{\partial qk} qk_i + \frac{\partial \Pi}{\partial \theta} \theta_i + \frac{\partial \Pi}{\partial \sigma_\theta} \sigma_{\theta_i} + Z_i \beta + O(w_i^w), \text{ or}$$

$$\quad \quad \quad (-) \quad \quad (-) \quad \quad (-)$$

$$\Pi_{e_i} = \mu_i M + O(w_i^w)$$

with $O(w_i^2)$ being a zero mean error term with standardized unit variance. Equation (3) implies that given that the self-employment spell has lasted until a particular time t , the probability that someone leaves self-employment in a short interval of time is a function of characteristics qk , θ , and σ_θ that may vary through time but are fixed over a short interval. We assumed that the process of thinking about the uncertainty of future income streams is a continuous time diffusion process. But the outcomes of it, as observed in the data are events. The useful parametric function to characterize this aspect of the distribution of observed spells out of self-employment is that of the hazard rate $\lambda(t)$, with the duration t being the time of an entrepreneur being self-employed, and accordingly the econometric analysis will be based on duration models.

The theory's hypotheses are tested on a dataset consisting of 421 individuals. This sample is part of the South Limburg dataset which contains information about 1223 individuals that either registered a business in South Limburg in 1998 or 1999 or registered at the "Starterscentrum South Limburg" in 1998 or 1999⁴³. This database is matched with a database obtained from the Chamber of Commerce to obtain the current status of the business in May 2004. 987 individuals registered as starter, 805 subscribed in 1998 or 1999, the time frame of our study. The total population of subscriptions in 1998 or 1999 is 5822. Our sample consists of 13.8 % of the total population of entrants that did subscribe at the Chamber of Commerce. From the sample of 803, we excluded 184 individuals that had been self-employed before; 61 individuals that had zero set up costs and 8 missing set up costs; 129 individuals that were not pioneering founders but took over a business; and 22 individuals that did not provide information about equity or debt capital. We end up with a net sample of 421 individuals.

⁴³ The Starterscentrum is a foundation sponsored by the chamber of commerce and other local authorities to encourage successful self-employment by providing training and advice.

Figure 7.2a-7.2c sketches the non-parametric Kaplan-Meier exit hazard for the different samples and the population. Figure 7.2a sketches the exit hazards for our net sample of 421 entrepreneurs and Figure 7.2b the exit hazards for all 803 entrepreneurs in the sample. The non-parametric hazards for both samples are almost identical. The overall picture in Figure 7.2a & 7.2b shows a hump-shaped course of the exit hazard, indicating that the chance of leaving self-employment is age dependent (duration dependence). This shape of the exit hazard is consistent with Jovanovic (1982) and Frank (1988) learning approach and has been empirically validated by numerous empirical studies (e.g. Mata and Portugal 1994) and is also consistent within our framework. The probability of exit is initially low, peaks after four years, and then decreases.

Figure 7.2c sketches the non-parametric hazard for the whole population. The exit hazard for the whole population is highest in the first two years, as is consistent with the findings of Mata and Portugal (1994). The difference in Figure 7.2a and 7.2c point toward a non-response bias for the more exit prone entrepreneurs. Our final sample is thus biased towards more (successful) long-living businesses. Predicted mean and median survival times are therefore not representative for the entire population.

Figure 7.2(a): Non-parametric Kaplan-Meier exit hazard (N=421)

Figure 7.2(b): Non-parametric Kaplan-Meier exit hazard (N=803)

Figure 7.2(c): Non-parametric Kaplan-Meier exit hazard (N=5822)

Furthermore, the dataset provides some information on the reasons for exiting in the population and in the sample. Table 7.2 provides the distribution of deregistration reasons in the population and in our net sample as provided by the Chamber of Commerce. The most common reason is the general deregistration (73.51% for the population and 81.03% in the sample). The second most common reason is due to deregistration by virtue of the Chamber of Commerce. If the Chamber suspects that the business is inactive it has the authority to deregister the business. Deregistration due to bankruptcy is not so common (4.67% in population and 5.17% in sample). The low probability of forced exits

substantiate our belief that exit can be modelled as a choice. The non-parametric hazard differences between the population and the sample result in significant different discontinuance rates. The discontinuance rate in the population is 25% points higher than in the sample (39% versus 15% respectively). For our empirical analysis, we assume that all reasons for exiting are due to shifting values between the two different states (self-employment and employment). The only possible problematic cases like bankruptcy and economic transfer account only for 8.5% of the exits in the sample. Excluding these two exit options did not significantly alter our results.

Table 7.2: Deregistration reasons for exit

Reason	Code	# of individuals			
		Population (N=5822)	%	Net (N=421)	Sample %
Deregistration due to rectification	21	10	0.45		
Deregistration by virtue of the Chamber of Commerce	24	260	11.58	5	8.62
Deregistration due to change in legal form or economic activity	25	3	0.13		
Deregistration due to discontinuance (general)	27	1651	73.51	47	81.03
Deregistration due to discontinuance by bankruptcy	28	105	4.67	3	5.17
Delayed deregistration (activity discontinued more than one year)	29	128	5.70	1	1.72
Unknown	31	7	0.31		
Deregistration due to economic transfer	32	74	3.29	2	3.45
Unknown	91	8	0.36		
Total		2246		58	
Discontinuance rate by May 2004		38.58%		13.78%	

In our analysis we estimate the time it takes before exiting self-employment. Thus a respondent comes at risk at the time of entry into self-employment. Figure 7.2a shows that the self-employment hazard rate is hump-shaped. This can be modelled with the flexible parametric log-logistic distribution⁴⁴. The hazard rate based on the log-logistic distribution of duration with covariates yields

$$(7.4) \quad \lambda(t | X) = \frac{\psi^{\frac{1}{\gamma}} t^{\left(\frac{1}{\gamma}-1\right)}}{\gamma \left[1 + (\psi t)^{\frac{1}{\gamma}}\right]}, \quad \text{with } \gamma > 0, \quad \text{and} \quad \psi = \exp(-\mu, M)$$

⁴⁴ The choice for this specification was confirmed by testing several other distributions, like Generalized Gamma, Weibull, Exponential, and Lognormal. Inspection of AIC values left the Log-Logistic model as the optimal choice.

If the shape parameter $\gamma < 1$ then the hazard rate function is hump-shaped. If $\gamma > 1$, the hazard rate is monotonically decreasing with t . We also test for unobserved heterogeneity adding to equation (7.4) some unobserved individual-specific effect α_j

$$(7.5) \quad \lambda(t | \psi, \alpha_j) = \alpha_j \lambda(t | \psi)$$

where α_j follows the Gamma distribution with mean one and variance θ (Cleves et al. 2004: 279)

$$(7.6) \quad g(\alpha_j) = \frac{\alpha_j^{1/\theta-1} \exp(-\alpha_j/\theta)}{\Gamma(1/\theta) \theta^{1/\theta}}$$

7.4.1 Testing the predictions: Sunk set up costs, uncertainty, and productivity

The theoretical model predicts a relationship between on the one hand the reluctance or *waiting time* to leave self-employment, and on the other hand theoretical constructs such as human capital that may raise self-employment productivity, predetermined at the time of entering self-employment through formal education, or developed through labour market experience or learned from (self-employed) parents or social contacts; the (sunk) set-up costs, and the role of uncertainty on the expected evolvement of self-employment income stream. In the section following the theoretical model we provide empirical concepts to actually measure these constructs.

Set up (sunk) costs (k)⁴⁵: The sum of the investments in computers and software, office furniture and equipment, representation goods, education and training, machines, transport equipment and buildings. Several authors have proposed that high capital investment is associated with more capital at risk of becoming sunk and capital investment has been used as a proxy for sunk costs (Campa 1993; Episcopos 1995; Lambson and Jensen 1998; O'Brien et al. 2003) This is further substantiated by the finding that a substantial fraction of investments in machinery is sunk (Asplund, 2000).

Uncertainty (σ_θ): we use the ordering of entry times conditional on observed and unobserved heterogeneity as a proxy of the underlying distribution of idiosyncratic uncertainty. Under steady state assumptions, two seemingly identical (in terms of unobserved and observed characteristics⁴⁶) individuals may only enter at different dates if their idiosyncratic uncertainty differs. Later entry times are associated with higher levels of expected idiosyncratic uncertainty. We acknowledge that utilizing the ordering of entry times as a proxy for idiosyncratic uncertainty is rather unconventional; however, the

⁴⁵ We assume $q > 0$ and equal among individuals.

⁴⁶ We have very rich and detailed information about the entrepreneurs in terms of human capital, labour market history, social background, sex, social status, marital status, children, and social network size and composition to account for entrepreneur heterogeneity. Including these control variables did not change the effect of our uncertainty proxy.

proxy is derived from the theory. Furthermore, other researchers have empirically related uncertainty to the *timing* of the investment decision (Favero et al. 1994; Hurn and Wright 1994)⁴⁷.

Productivity (θ)⁴⁸: Measurements of human capital are considered good proxies for the expected idiosyncratic growth of idiosyncratic entrepreneurial income. We follow Marshall (1890), Penrose (1959), and Lazear (2005) that entrepreneurs should be jack-of-all-trades. The steepest entrepreneurial income streams are expected from individuals that followed vocational schooling. In the Netherlands MBO and HBO offers the broadest curriculum and therefore we assume that this curriculum is most inline with the jack-of-all-trades hypothesis⁴⁹. We expect an inverted-U shaped relationship between the level of schooling and idiosyncratic productivity and income growth.

We include the following relevant control variables. (1) Debt capital: The amount of debt capital at the start of company. (2) Equity capital: The amount of equity capital at start of company. (3) Foreign: A dummy indicating whether the entrepreneur has the Dutch nationality or some other nationality. (4) Age over 50: A dummy indicating whether the entrepreneur was older than 50 at the time he or she entered self-employment. Individuals in the Netherlands can legally retire around 55.

Table 7.3: Descriptive statistics of independent variables

Variable	N	Mean	Std. Dev	Min	Max
Age	421	39.501	8.411	23	62
Set Up Cost k	421	9.729	1.547	4.394	14.635
Ed = primary school	421	.031	.173	0	1
Ed = LBO	421	.164	.371	0	1
Ed = MAVO/ HAVO	421	.162	.368	0	1
Ed = VWO	421	.031	.173	0	1
Ed = MBO	421	.249	.433	0	1
Ed = HBO	421	.276	.447	0	1
Ed = University	421	.088	.283	0	1
Equity Capital	421	37921	93376	0	1000000
Debt Capital	421	29969	95067	0	1000000
Foreign	421	.057	.232	0	1
Age > 50	421	.109	.312	0	1

⁴⁷ Admittedly, these researchers have used direct proxies of uncertainty to predict the timing of the investment decision.

⁴⁸ We also experimented with other sources of human capital, especially with size and composition of the entrepreneur's social network. These constructs turned out to be insignificant in our analysis and are therefore excluded.

⁴⁹ We acknowledge that Lazear's measurement for jack-of-all-trades is more specific but we justify our assumption on the broadness of the curriculum of vocational schooling.

Table 7.3 presents the descriptive statistics of the variables. The mean age, at which entrepreneurs start a new business for the first time in their life, is between 39 and 40 and 10.9% of the starting entrepreneurs are over 50. Most entrepreneurs had vocational schooling (MBO and HBO in the Netherlands). Mean set up costs of the business are Euro 16798. Mean equity capital at entry is Euro 37921 and debt capital Euro 29969. 5.7% of the starting entrepreneurs have a foreign nationality.

7.5 Estimation results

Table 7.4 shows the results of the duration analysis. Model (1) and model (2) refer to the log-logistic duration model without unobserved heterogeneity and with unobserved heterogeneity. The likelihood ratio test indicates the presence of unobserved heterogeneity. For the discussion of the hypotheses we will refer to model (2).

The uncertainty proxy, sunk costs, and the productivity proxy increase the spell length of self-employment in both models. These findings support our thesis that the sunk nature of the self-employment decision creates option values. More specific, entrepreneurs, who expect more uncertain income distributions, and therefore wait one more year before they enter self-employment, have subsequent 1.7 percent ($\exp(.017)=1.017$) longer self-employment spells. This finding implies that the idiosyncratic uncertainty of self-employment income streams increases the option value to abandon the firm and to exit self-employment. Note that this finding asserts that the timing of the entry and exit decisions is related. Individuals that have to overcome sunk set up costs and expect uncertain self-employment income streams wait longer to collect more information before they enter self-employment. At the same time, this uncertainty makes entrepreneurs more reluctant to exit self-employment once they have entered. This result is consistent with Dixit (1989) finding that uncertainty increases both entry and exit thresholds.

Since we use the logarithmic transformation of set-up costs, the associated coefficient of set up costs k can be interpreted as a constant elasticity, and, accordingly, a one percent change in the set-up costs implies, approximately, a 0.142 percent increase in the spell length of self-employment. The time ratio for set up costs k is $\exp(0.142)=1.152$, which implies that waiting time is 1.152 times higher for an individual with 1 percent more set-up costs. The sunk cost component of set up costs creates an incentive for the entrepreneur to wait to exit and see if business conditions turn favorable again. Finally, our productivity proxy shows that vocational schooling (MBO & HBO) increases self-employment spell length. Our finding indicates that individuals with vocational schooling have steeper income growth in entrepreneurship. Steeper income growth increases the option value to abandon their firm and reduces the probability of exit and increases self-employment spell length.

Table 7.4: Estimates of log-logistic duration model of waiting time out of self-employment

Model Parameters	Model(1)		Model(2)	
	Waiting Time		Waiting Time	
	Estimate	Se	Estimate	Se
Uncertainty σ	0.021	[0.009]*	0.017	[0.009]*
Set up costs k	0.108	[0.041]**	0.142	[0.045]**
Productivity θ (MBO = base)				
Ed = primary school	5.260	[569.968]	4.919	[2,692.902]
Ed = LBO	-0.360	[0.200]+	-0.389	[0.209]+
Ed = MAVO/ HAVO	-0.521	[0.192]**	-0.412	[0.187]*
Ed = VWO	5.433	[518.145]	5.127	[2,445.376]
Ed = HBO	-0.264	[0.182]	-0.229	[0.167]
Ed = University	-0.615	[0.217]**	-0.692	[0.211]**
Control Variables				
Equity capital	-1.22 ^{e-06}	[4.96 ^{e-07}]*	-1.05 ^{e-06}	[6.04 ^{e-07}]+
Debt capital	-6.47 ^{e-07}	[6.01 ^{e-07}]	-6.26 ^{e-07}	[6.90 ^{e-07}]
Foreign	-0.730	[0.185]**	-0.777	[0.202]**
Age > 50 at entry	-0.513	[0.236]*	-0.384	[0.237]
Constant	6.921	[0.503]**	6.392	[0.550]**
Unobserved heterogeneity				
$\ln(\theta)$			1.239	[.726]+
Likelihood-Ratio test $\theta = 0$			4.68*	
Shape Parameter:				
γ	0.370	[0.044]	0.271	[.055]
Model specification Test:				
Likelihood-Ratio test				
Chi2(12)	48.66**		47.19**	
Observations	421		421	

Standard errors in brackets

+ significant at 10%; * significant at 5%; ** significant at 1%

The results of the control variables show that the financing of the firm, either through debt or equity capital, may have an effect on the spell of self-employment. An increase in debt capital or equity capital reduces the spell of self-employment; however these results are only marginally significant for equity capital and not significant for debt capital⁵⁰. These results substantiate our finding that it is not the amount of capital available that influences small business longevity, but the sunkness of the set-up costs. The results in Table 7.4 also indicate that foreign entrepreneurs have lower survival chances than native entrepreneurs.

The estimated γ shows that the baseline hazard of exiting self-employment is hump-shaped. This shape is consistent with the passive learning model (Jovanovic, 1982). It takes time for individuals to learn about their true entrepreneurial ability and to learn about the true stochastic process driving the uncertainty in their business environment. Figure 7.3 plots the predicted baseline exit hazard for the sample. During the first two years the exit hazard is very low, after that period it shows a long period of rising exit hazard. Several reasons might explain this exceptionally long period of rising exit hazards. First, the rising hazards might reflect the rather unexpected long economic downturn starting late 2001 in the European Union. Second, the long learning period might be the result of the non-response bias in our sample. Our sample might be disproportionately supplied by high commitment entrepreneurs as is reflected in the relatively low overall exit rate in our sample. And Frank (1988) shows that it takes more time to alter the initial beliefs of highly committed entrepreneurs.

Figure 7.3: Baseline exit hazard in months (N=421)

⁵⁰ A logarithmic transformation of these variables reduced the likelihood of the models.

7.6 Conclusion

In this paper we designed a theoretical model to explain the reluctance of entrepreneurs to exit self-employment in the presence of sunk costs and an uncertain evolution of self-employment income. The model includes three important characteristics of that decision. First, initial commitments in partially sunk set up costs are important and irreversible. Second, the stochastic development of future self-employment earnings must be dealt with in the evaluation of value streams from different income streams. The expected growth of idiosyncratic self-employment earnings makes entrepreneurs more reluctant to exit self-employment even when their income is below labour market pay. Third, idiosyncratic uncertainty of idiosyncratic earnings acts as a disciplinary device for the decision making process. The option values of postponing entry and exit decisions can account for the reluctance of individuals to become self-employed and the reluctance of entrepreneurs to leave self-employment. The empirical contents of the model have been tested and confirmed with a parametric duration analysis of small business longevity for a group of 421 individuals over the critical first six years of the life of their first and newly established enterprise.

We have shown that the theoretical model allows for more flexibility regarding the interpretation of empirical results than the traditional static decision framework that is based on net present value comparison of different types of employment choices. Our dynamic framework shows that entry and exit decisions are related and that the timing of the entry decision might reveal important information about small business longevity. Static models assume costless switching. We have shown not only that switching employment options is costly and that sunk switching costs make entrepreneurs reluctant to exit, but we also show that initial commitments in sunk set up cost are endogenous to previous income levels. High income individuals that become self-employed must commit more set up costs to make entry profitable. This initial commitment in sunk set up costs makes entrepreneurs reluctant to exit. Ignoring sunk switching costs, like set up costs, in empirical applications might seriously bias estimates correlated with earnings.

Although we do not formally investigate the option value associated with the entry decision, its implications are obvious and provide some evidence for the hitherto unexplained result in the self-employment literature that high income individuals are less likely to become self-employed. High income individuals have to invest more sunk capital to raise entrepreneurial earnings above entry thresholds. However, since future earnings are uncertain, they have an option value to wait. Therefore, high income individuals will be more reluctant to enter self-employment than low income individuals. Further research should develop on this theme.

7.7 Appendix

Table 7.5: Summary statistics

	Mean	Std. Dev.	Min	Max
Explanatory Variables				
Capital Investment (log)	9.838	1.471	4.394	14.635
Debt capital (log)	3.265	5.014	0	13.816
Equity capital (log)	8.110	3.854	0	13.816
Income (log - yearly)	10.833	.735	6.100	12.612
Ed = primary school	.027	.162	0	1
Ed = LBO	.156	.363	0	1
Ed = MAVO/ HAVO	.150	.357	0	1
Ed = VWO	.033	.179	0	1
Ed = MBO	.257	.438	0	1
Ed = HBO	.287	.453	0	1
Ed = University	.090	.286	0	1
Labour experience (months)	121.635	109.564	0	479
Unemployment experience (months)	4.563	20.091	0	231
Married at time of entry	.808	.394	0	1
Self-employed parents	.359	.481	0	1
Self-employed parents-in-law	.228	.420	0	1
Female	.192	.394	0	1
Only owner	.793	.405	0	1
Business Services	.467	.500	0	1
personal service incl. health and culture	.141	.348	0	1
manufacturing incl. utility	.060	.238	0	1
hotels & restaurant sector	.015	.122	0	1
construction sector	.141	.348	0	1
wholesale & retail sector	.159	.366	0	1
Social Network Size	4.596	2.449	0	10
Age	40.117	8.483	23	62

Table 7.6: Correlation Matrix

	1	2	3	4	5	6	7	8	9	10
1 Capital Investment (log)										
2 Debt capital (log)	.41									
3 Equity capital (log)	.23	-.17								
4 Income (log - yearly)	.18	-.04	.08							
5 Ed = primary school	.04	.06	-.05	-.05						
6 Ed = LBO	-.03	.03	-.09	-.12	-.07					
7 Ed = MAVO/HAVO	-.04	-.03	-.05	-.12	-.07	-.18				
8 Ed = VWO	.02	.06	.04	-.02	-.03	-.08	-.08			
9 Ed = MBO	.06	.02	.03	-.08	-.10	-.25	-.25	-.11		
10 Ed = HBO	-.07	-.07	.08	.19	-.11	-.27	-.27	-.11	-.37	
11 Ed = University	.07	.01	.01	.16	-.05	-.13	-.13	-.06	-.19	-.20
12 Labour experience	.13	-.03	.01	.19	.19	.12	.02	-.03	-.09	-.11
13 Unemployment experience	-.10	-.10	.02	-.10	-.04	.07	.05	-.02	.04	-.07
14 Married at time of entry	-.05	-.01	-.06	.18	.03	.08	-.03	.05	-.01	-.03
15 Self-employed parents	.09	.08	-.02	-.06	-.01	.02	.04	.00	-.04	.03
16 Self-employed parents-in-law	.05	-.04	.10	.03	.09	.00	.07	.06	-.04	-.06
17 Female	-.14	.01	-.02	-.44	.01	.02	.12	-.00	-.04	-.04
18 Only owner	-.15	-.16	-.04	-.04	.04	.02	.01	-.03	.03	-.02
19 Business Services	-.09	-.09	-.08	.23	-.08	-.24	-.09	.03	-.03	.16
20 personal service	-.07	-.08	.01	-.27	-.07	-.01	.14	-.07	-.00	.03
21 manufacturing incl. utility	.03	.07	.07	.01	-.04	.07	-.04	-.05	.02	.01
22 hotels & restaurant sector	.17	.16	.01	.00	-.02	.02	.02	-.02	-.02	.03
23 construction sector	.04	-.01	.04	.00	.04	.28	-.00	-.03	.04	-.20
24 wholesale & retail sector	.08	.13	.04	-.09	.13	.02	.00	.10	.01	-.08
25 Social Network Size	.06	.04	.07	.04	-.04	-.09	-.02	-.03	-.05	.12
26 Age	.05	-.13	.03	.35	-.02	-.04	-.00	.03	-.06	.03

11	12	13	14	15	16	17	18	19	20	21
----	----	----	----	----	----	----	----	----	----	----

.04										
-.06	-.01									
-.06	.20	-.09								
-.06	.03	-.05	-.03							
-.02	.06	.06	.16	.07						
-.05	-.08	-.01	.12	.03	.04					
-.05	.06	-.00	.05	-.02	.01	.04				
.23	.05	.01	-.02	-.09	.02	-.17	.02			
-.13	-.15	.01	-.07	.02	-.01	.39	.12	-.38		
-.04	.01	-.05	.03	.07	-.02	-.06	-.06	-.24	-.10	
-.04	-.10	-.03	-.07	.06	-.07	.00	-.06	-.12	-.05	-.03
-.10	.09	-.03	.09	.06	.05	-.13	.06	-.38	-.16	-.10
-.05	.04	.06	.00	-.02	-.02	.04	-.14	-.41	-.18	-.11
.07	-.09	-.02	-.02	.05	.04	.06	-.15	.00	-.01	.05
.10	.30	.09	.32	-.03	.12	-.10	.11	.15	-.17	-.01

Table 7.6: Correlation Matrix (continued)

	22	23	24	25	26
1 Capital Investment (log)					
2 Debt capital (log)					
3 Equity capital (log)					
4 Income (log - yearly)					
5 Ed = primary school					
6 Ed = LBO					
7 Ed = MAVO/HAVO					
8 Ed = VWO					
9 Ed = MBO					
10 Ed = HBO					
11 Ed = University					
12 Labour experience					
13 Unemployment experience					
14 Married at time of entry					
15 Self-employed parents					
16 Self-employed parents-in-law					
17 Female					
18 Onlyowner					
19 Business Services					
20 personal service					
21 manufacturing incl. utility					
22 hotels & restaurant sector					
23 construction sector	-05				
24 wholesale & retail sector	-05	-18			
25 Social Network Size	.03	-07	.02		
26 Age	-06	-02	-01	-05	

Chapter 8

A discrete choice model for productive social relations

8 A discrete choice model for productive social relations⁵¹

8.1 Introduction

This paper investigates investment behavior in productive social relations of starting entrepreneurs. Social relations may turn into productive social capital. Social relations may add to profitability by economizing on transaction costs, minimizing search costs, enforcing contracts, stimulating trade, and providing timely access to valuable and reliable information. Entrepreneurs who face higher returns on investments in social relations may therefore be more successful entrepreneurs. The accomplishment of social relations is a rational choice. In this paper, we analyze the choice of the number of productive relations in a profit maximizing framework under uncertainty. The main objective of this paper is to show that profitability of productive social relations depends on the productivity of social relations within an industry and idiosyncratic accomplishment costs.

Theory predicts that relation-specific investments influence productivity. Information conveyed in social relations reduces the uncertainty about competitor and supplier prices in imperfect markets (Stigler 1961). Investments which are specific to a certain relationship may create trust and economize on transaction costs and may lead to a concentration of trade between identified partners (Ben-Porath 1980). And social relations provide information that reduces search costs when markets show frictions (Kranton 1996).

Recently, empirical studies started to investigate the connection between social relations and firm performance including social networks into the production function of firms. Barr (2000) predicts that social networks affect firm productivity indirectly through knowledge flows between firms. Fafchamp and Minten (2002) argue that social networks may help overcome or exploit market imperfections and therefore raise profitability. Profitability is increased either directly through price discrimination or indirectly through an increase in the productivity of factor inputs. Both studies find strong evidence for a positive relationship between social networks and firm productivity. Empirically, these studies test for the determinants of social networks. However, their interest is in finding instruments for social capital and not in the formation per se.

Although theoretical and empirical research validates the incentive to invest in productive social relations, they do not explain the differences in the total amount invested by different persons. Figure 8.1 gives the frequency distribution of the number of productive social relations maintained by recently started entrepreneurs in a province of the Netherlands. The observed distribution of productive social relations among entrepreneurs raises the following questions: why do investments in productive social relations differ between persons and what accounts for these differences?

⁵¹ This chapter is based on the paper “A discrete choice model for productive social relations” by Kok and Pfann (2006) and published in IAMB Conference Proceedings 2006 and submitted.

Figure 8.1: Frequency distribution of productive social relations of entrepreneurs

The model presented in this paper explicitly accounts for the accomplishment of productive social relations. Glaeser (2001) and Glaeser, Laibson and Sacerdote (2002) analyze the formation of social capital, as measured by club membership, using a model of optimal individual investment decisions. These studies show that the formation of social capital is based on rational individual investment decisions. The results are consistent with the predictions of a standard model for investment decisions under uncertainty.

We model the demand for productive social relations as an outcome of profit maximizing behavior under uncertainty. Social relations are considered productive inputs in the production process of firms. The productivity of a relation may vary across industries because of differences in the production processes between industries. Relations may also be more productive when markets are not perfectly competitive. Such markets are often characterized by search and transaction costs. The optimization of complex production processes may demand more information than simpler ones, and trust and information may reduce transaction and search costs in heterogeneous inputs and outputs markets more than in their homogenous counterparts. For example, a car manufacturer's trust in his custom-made car parts supplier may reduce contract complexity and transaction costs beyond the level a baker's corn supplier can. Therefore, seemingly identical social relations may differ in productivity to the extent that they are able to overcome market imperfections and optimize the production process of firms within an industry. In our model the costs of accomplishing social relations are heterogeneous. Equilibrium demand is identified in the market for productive relations where marginal returns of accomplishing relations equal marginal costs.

The model's predictions are tested using a dataset of potential entrepreneurs in the South of Limburg, a part of a province in the Netherlands. The econometric analysis shows that young, native, highly educated entrepreneurs in large cities, with self-employed parents, have lower cost of forming productive social relations. The ordering of profits shows decreasing returns to scale for social relations. On average, entrepreneurs obtain full information at 4 to 5 social relations. We conclude that the equilibrium de-

mand for productive social relations can be derived from the demand conditions emerging from industry structure and the supply conditions emanating from entrepreneurs' heterogeneity.

The paper is organized as follows. Section 8.2 sets out the basic theory of optimal investments in productive social relations. We apply a simple investment model to show differences between returns and costs of investments in social relations. The model determines how the extent of productive social relations depends on the market structure and idiosyncratic cost curves. Section 8.3 introduces the econometric demand model. Section 8.4 summarizes the information available in the data and section 8.5 presents the results of the econometric analysis. Section 8.6 concludes.

8.2 Investing In productive social relations

The model views investments in productive social relations as a cost-benefit decision that takes into account the expected returns of the investments and the cost of the investments (Becker 1964). Investing in social capital is like investing in human capital. Individuals increase their productivity by learning from others and cooperating with others. And some may benefit more from their social network than others. Presumably, productivity can be improved only at a cost, for otherwise there would be an unlimited demand for productive social contacts. Included in the cost are the time and effort for accomplishing social relations. These costs may differ for individuals. In our model, differences in incentives and costs drive the process of investing in other people.

Figure 8.2: Within industry equilibrium levels of investments in productive social contacts

Figure 8.2 shows the supply and demand curves for investments in productive social relations. Along the horizontal axis, the amount invested in social relations is measured by its (implicit) accomplishment costs. The curve D shows the marginal benefit (MR in Figure 8.2), for simplicity measured by the rate of return, to an entrepreneur in a particular industry on each additional Euro of investment. D represents the demand curve for productive social relations. The curve S shows the marginal cost of accomplishing a relation⁵² (MC in Figure 8.2). Marginal cost curves differences are reflected in S_L and S_H in Figure 8.2, representing low and high marginal cost respectively. If D exceeds S , the marginal rate of return would exceed the marginal cost, and the entrepreneur's profits would be increased by additional investment, and vice versa. Consequently, equilibrium is achieved in the point where $D = S$, and is given by $\bar{p}$.

8.2.1 Demand curves

Economists stress that the incentive to expand physical and human resources depends on the rate of expected return. Only recently, economic logic expanded to social resources (Burt 1992; Glaeser 2001; Glaeser et al. 2002). In our model, the decision to invest in social resources is a systematic response to the rate of return expected. The demand curve D in Figure 8.2 represents the incentive to invest. However, the incentive to invest is not similar among persons. We ignore "standard" investment incentives like the number of periods and the discount rate, but focus on the influence of the market structure on the incentive to invest in other people⁵³. Investments in social relations would be discouraged if the market would be nearly perfect; thus the incentive to cooperate and to invest time and effort in other people increases as the level of market imperfection increases. In short, the more concentrated the market, the more productive social relations may be (Fafchamps and Minten 2002).

Although market imperfections are a necessary condition for the productivity of social relations, they are not sufficient. The incentive to invest, the rate of expected return, depends on the market structure, the quality of the relationship, and the ability and skills of the investor to benefit from the relation. In perfectly competitive markets, all goods are traded in a market at publicly known prices and agents act as price takers (Mas-Colell et al. 1995). This assumption requires that the characteristics of all goods be observable to all participants in the market. Asymmetric information about heterogeneous in and outputs and their transformation process lead to market failure. Market imperfections lead to transaction costs to reduce uncertainty and moral hazard and they lead to search costs to find information about the production process and optimal prices. More condensed markets provide more opportunities for social relations to become productive. Social relations become productive if they provide information or trust that reduce transaction and search costs in imperfect markets (Fafchamps and Minten 2002). Entrepreneurs benefit more from high quality relationships that provide better and more trusted

⁵² We observe discreteness in social relations, but we assume an underlying continuous process of investments.

⁵³ In Glaeser et al.'s (2002) investment model the incentive to invest depends on the structure of the occupation. More sociable occupations have higher expected returns.

information. Ultimately, the (social) skills and ability of the entrepreneur determine the extent to which the value of the relationship is extracted⁵⁴.

The demand curve in Figure 8.2 depends only on the structure of the market⁵⁵. The incentive to invest may differ across industries to the extent that social relations can capitalize on market failures. The probability that social contacts provide overlapping information increases as a business owner's social network grows (Burt 1992). Because overlapping information causes information redundancy, we expect negatively inclined demand curves.

8.2.2 Supply curves

The supply curves in Figure 8.2 show the accomplishment costs. The marginal cost of accomplishing (trading) partners is of central importance in understanding investments in productive social relations. With equal opportunity of finding equally valuable partners, the market for social relations would be perfect, and marginal cost of finding partners would be the same for entrepreneurs across all industries. In the real world, the market for productive social relations is segmented: spatial distance, status, centrality, club membership, private resources, talents, abilities and family ties create unequal opportunities as well as different marginal cost of accomplishing productive social relations. All factors may affect the probability of accomplishing productive social relations. The cheapest sources for finding productive social relations are a business owner's family, friends and relatives. However, when this source is exhausted, more expensive sources, like clubs and trade organizations must be drawn on. This process is even more complicated, because rational actors will only invest in relations that are Pareto optimal in the sense that the relation must benefit both. Idiosyncratic characteristics create substantial differences in idiosyncratic marginal cost curves. These idiosyncratic curves contribute to differences in the equilibrium price for productive social relations.

8.2.3 Equilibrium

If we assume rationality of the decision-makers and a sufficiently large discount rate on investments in social relations (Burt 2002)⁵⁶, then an entrepreneur's optimal decision is to maximize profits by setting the marginal cost of investments in social relations equal to

⁵⁴ Gleaser et al. (2002: 438) stress this aspect when they define "individual social capital as a person's social characteristics – including his social skills, charisma, and the size of his Rolodex – which enables him to reap market and non-market returns from interactions with others".

⁵⁵ For example, assume an imperfect market with transaction costs because suppliers are uncertain if contracts will be enforced. If a supplier trusts an entrepreneur, he can save on transaction costs and provide the goods at a lower price to the business owner. The maximum amount that can be saved is the total amount of transaction costs, which is determined by the market. If the market would change from imperfect to perfect where contracts are always enforced, then transaction costs would go to zero and the trust provided by the social relation would stop adding to productivity.

⁵⁶ If social relations do not have a large discount rate, then the optimal decision is to select an accumulation pattern that maximizes the net present value of profits. However, Burt (2002) finds extremely high decay rates of relations within a year.

the marginal (rate of) return of these investments. In this simple investment model, differences in either demand or supply conditions lead to differences in social relations equilibrium demand. Entrepreneurs with either higher demand curves or lower supply curves will invest more than their colleagues with unfavorable conditions.

Within an industry, investments and returns may differ only because of idiosyncratic differences. These properties make investments in social relations comparable to the “egalitarian approach” to investment in human capital. This approach assumes that the demand conditions are the same for everyone. All individuals have the same capacity to benefit from investments in social capital. Investments and returns differ only because of the idiosyncratic characteristics of the entrepreneur. These differences change the costs of accomplishing relations and give some the opportunity to invest more than others. It is not unlikely that differences in investment opportunities related to social relations exist. High status individuals have access to more influential individuals and clubs and wealthy children benefit from the social relations of their parents (Becker and Murphy 2000). Eliminating these idiosyncratic differences would eliminate differences in opportunities, and thereby eliminate the important differences in investments and returns. In the following section, we will present our modeling framework for the demand for productive social relations.

8.2.4 Productive social relations

Consider an entrepreneur i in industry j that uses only one type of productive inputs, S_{ij} , which is the knowledge embodied in social relationships. To produce output Y_{ij} , the nominal price per input S_{ij} yields P_{ij}^S . The price for output is P_{ij}^Y . The entrepreneur wants to choose the level of output so as to maximize

$$(8.1) \quad \Pi(S_{ij}) = P_{ij}^Y Y_{ij} - P_{ij}^S S_{ij}$$

and given the actions of all other firms in the market the equilibrium condition is that marginal revenue equals marginal costs. Choosing S_{ij} then is

$$(8.2) \quad P_{ij}^Y \frac{\partial Y_{ij}}{\partial S_{ij}} - P_{ij}^S = 0 \quad \text{or} \quad \frac{\partial Y_{ij}}{\partial S_{ij}} = \frac{P_{ij}^S}{P_{ij}^Y}$$

with necessary condition $\frac{\partial^2 Y_{ij}}{\partial^2 S_{ij}} < 0$ for a maximum.

The indirect profit function represents the equilibrium demand for social relationships, S_{ij}^* . Relationships are only observed in discrete quantities S_{ij} , and these are determined upon evaluation of the breakeven condition

$$(8.3) \quad S_{ij} = n \quad \text{if} \quad \Pi(S_{ij} = n) > 0 \quad \text{and} \quad \Pi(S_{ij} = n + 1) \leq 0$$

The real price $\frac{P_{ij}^S}{P_{ij}^Y}$ is not observable. To accommodate this unobservability, let X_{ij} be a vector of determinants of $\frac{P_{ij}^S}{P_{ij}^Y}$. We assume that industry returns to social capital are equal within industries, but may differ among industries. The equilibrium demand function can then be rewritten as

$$(8.4) \quad S_{ij}^* = \alpha_j + X_{ij}\beta + \varepsilon_{ij}$$

where α_j captures the differences in productivity of social relations among industries j , $\frac{\partial Y_{ij}}{\partial S_j}$, and β captures the effect of the real price proxy X_{ij} on the equilibrium demand.

Let $\pi_{1j} < \pi_{2j} < \dots < \pi_{N_j}$ be unknown profit threshold parameters for industry j , and N_j be the maximum amount of social relationships in industry j . Then it holds that

$$(8.5a) \quad S_{ij} = 0 \quad \text{if} \quad S_{ij}^* \leq \pi_{1j};$$

$$(8.5b) \quad S_{ij} = 1 \quad \text{if} \quad \pi_{1j} < S_{ij}^* \leq \pi_{2j};$$

$$(8.5c) \quad S_{ij} = N_j \quad \text{if} \quad S_{ij}^* > \pi_{N_j}.$$

For our estimation procedure we assume that ε_{ij} has a logistic distribution. The resulting ordered logit model is a member of the GEV-family (Small 1987; Wen & Koppelman 2001). GEV-models are consistent with profit maximization under uncertainty when choices are discrete (McFadden 1984). The response probabilities are

$$(8.6a) \quad \Pr(S_{ij} = 0 | X_{ij}) = \Pr(S_{ij}^* \leq \pi_{1j} | \alpha_j, X_{ij}) = \Pr(\alpha_j + X_{ij}\beta + \varepsilon_{ij} \leq \pi_{1j} | \alpha_j, X_{ij}) = \Lambda(\pi_{1j} - \alpha_j - X_{ij}\beta)$$

$$(8.6b) \quad \Pr(S_{ij} = 1 | \alpha_j, X_{ij}) = \Lambda(\pi_{2j} - \alpha_j - X_{ij}\beta) - \Lambda(\pi_{1j} - \alpha_j - X_{ij}\beta)$$

$$(8.6c) \quad \Pr(S_{ij} = N_j - 1 | \alpha_j, X_{ij}) = \Lambda(\pi_{N_j} - \alpha_j - X_{ij}\beta) - \Lambda(\pi_{N_j-1} - \alpha_j - X_{ij}\beta)$$

$$(8.6d) \quad \Pr(S_{ij} = N_j | \alpha_j, X_{ij}) = 1 - \Lambda(\pi_{N_j} - \alpha_j - X_{ij}\beta)$$

8.3 Data

In order to estimate our model of entrepreneurs' demand for productive social relations, we use a survey among potential entrepreneurs in South Limburg in 1998 and 1999. The sample population is extracted from two databases: (1) The registration database of the local chamber of commerce contains all new firms, who registered a business in South

Limburg in 1998 or 1999. (2) The address database of the “Starterscentrum South Limburg” contains addresses of persons, who explicitly stated they seriously planned to start a company⁵⁷. Together, both databases contain 4339 unique and complete addresses. After an initial telephone screening, 1223 people were willing to participate in the study and were interviewed for on average one hour. Thus, the response rate is 28%. The interviews were conducted from November 2000 till June 2001.

Table 8.1 provides summary information about potential entrepreneurs in the South of Limburg during 1998 and 1999. 1223 potential entrepreneurs are interviewed.

Table 8.1: Population & sampling details

	Sample	Population
Total	1223	
Non-starters	236	
Starters	987	
Start before 1998 or after 1999	184	
Starters 1998/1999	803	3752
Agriculture	12	
Missing SBI code	2	
No information network	14	
Sample size	775	

Based on the interviews, 987 respondents are labelled as starter. From this group, we have deleted 184 entrepreneurs that started before 1998 or after 1999. We also exclude 10 respondents that do not provide any information concerning their social networks and 2 respondents that did not provide information on the sector they operated in⁵⁸. Finally, we exclude 14 observations that operate in the agricultural sector. This results in a net sample of 775 observations.

⁵⁷ The Starterscentrum is a foundation sponsored by the chamber of commerce and other local authorities to encourage successful self-employment by providing training and advice.

⁵⁸ For the zero-network observations, we do not know whether respondents are not willing to give information or whether they make no use of any relation.

For our empirical analysis, we define entrepreneurs' choices of the type of relations. Productive relations may emerge from different sources, like family, friends, colleagues, acquaintances, or business contacts. The network data were generated through "name-generator" type questions. Interviewees were asked to list the names of persons (1) with whom they discussed important matters during the self-employment decision process; (2) with whom they discussed important matters during the last six months; (3) with whom they were involved in informal social activities (4) whom they considered essential contacts (5) whom they considered difficult contacts (6) and whom they considered an important contact as well, but that was not listed before. A maximum of four answers per question were allowed, leading to a maximum social network of 24. Identical answers at multiple questions were allowed. However, we accounted for identical answers, so we end up with the respondents' net number of social relationships. Our data reveals ten different levels (see Figure 8.1).

The econometric model from equation (8.4) is specified as follows:

The individual profits of an entrepreneur is made up of two components: a sector constant profit from sector specific marginal returns and a variable profit from cost differences that stem from the differences in the real price of accomplishing relations. The industry constant profit is proxied by industry sector dummies. We have classified six sectors with the use of SBI codes: business services (SBI 1000-4500), construction (SBI 4500-5000), trade (wholesale & retail) (SBI 5000-5500), catering (hotels & restaurants) (SBI 5500 – 6000), manufacturing (including utility) (SBI 6000-8500), and personal services (including health and culture) (SBI 8500-9900). Theory predicts higher expected profits for sectors with higher marginal returns.

The vector of characteristics X_{ij} include⁵⁹: self-employed parent(s), age, education, City, and native. Theory predicts that entrepreneurs with higher educated parents have better access to social resources (Lin 1990; Lin 1999). Self-employed parents may provide access to relevant contacts from their own network. Therefore, we expect that self-employed parents reduce the price of obtaining contacts and increase the number of social relations ($\beta_1 > 0$). The effect of age on price and profits is ambiguous ($\beta_2 > 0$ or $\beta_2 < 0$). Young individuals may be more socially active, but older individuals may know more people. Previous research established a positive relation between education and access to social resources (Boxman et al. 1991; Glaeser et al. 2002; Lin 1990; Lin 1999). Entrepreneurs who invest in education face lower real prices for accomplishing contacts and therefore increase the number of social relations ($\beta_3 > 0$). We expect that individuals living in the largest city in the area will have more opportunities to meet people and therefore lower cost ($\beta_4 > 0$). The people in our sample are from Limburg, a province in the South of the Netherlands. This area is known for its associational activity, but also for its closed culture. Cultural differences may pose barriers to enter these associations. Natives will have lower cost of forming relations and therefore being native is expected to increase network size ($\beta_5 > 0$).

⁵⁹ We also used education parents, marital status, self-employed before, gender, a dummy demarcating whether the individual started in 1998 or 1999, and age squared. However, none of these covariates were significant at conventional levels, and are therefore subsequently excluded from the analyses.

Table 8.2(a): Descriptive statistics: Mean and standard deviations by industry

	Total	Business Services	Construction	Trade	Manufacturing	Catering	Personal Services
Variable							
# of social relations	4.82 (2.51)	5.19 (2.53)	4.11 (2.29)	4.74 (2.42)	4.77 (2.63)	4.70 (2.45)	4.58 (2.61)
Parents self-employed	.36 (.48)	.33 (.47)	.40 (.49)	.38 (.49)	.52 (.51)	.30 (.46)	.37 (.48)
Age	38.44 (9.14)	39.82 (9.25)	38.51 (8.78)	38.77 (8.60)	40.27 (10.04)	34.85 (8.43)	35.77 (8.90)
Education	4.57 (1.74)	5.25 (1.56)	3.50 (1.72)	4.16 (1.65)	4.64 (1.74)	3.94 (1.76)	4.48 (1.57)
City	.19 (.39)	.18 (.38)	.09 (.28)	.24 (.43)	.18 (.39)	.25 (.43)	.16 (.37)
Native	.94 (.24)	.96 (.19)	.94 (.24)	.92 (.27)	.95 (.21)	.92 (.28)	.90 (.30)
N	775	301	82	164	44	73	111

Table 8.2(a) shows the descriptive statistics of these variables by industry. Table 8.2(a) provides interesting information about the characteristics of the entrepreneur in each industry. Entrepreneurs in business services have the largest social networks. Entrepreneurs in construction have the smallest social networks. More than half of the starting entrepreneurs in manufacturing have self-employed parents. In catering, less than one third have self-employed parents. Entrepreneurs in business services had the highest level of education, while in construction entrepreneurs had the lowest level of education. The entrepreneur operating in construction also lives more frequently outside a city. The entrepreneur in catering lives most frequently in a city. Finally, almost 10% of the entrepreneurs operating in personal services are foreigner. Only 4% of the entrepreneurs in business services are foreigner.

Table 8.2(b): Descriptive statistics: Mean and standard deviations by number of social relations

Variable	# of social relations									
	1	2	3	4	5	6	7	8	9	10
Parents self-employed	.38 (.49)	.30 (.46)	.29 (.46)	.33 (.47)	.42 (.50)	.39 (.49)	.44 (.50)	.42 (.50)	.42 (.50)	.35 (.48)
Age	41.05 (11.20)	41.23 (9.40)	38.01 (8.56)	38.18 (8.39)	37.43 (9.04)	36.43 (8.20)	37.14 (9.07)	39.42 (8.82)	38.00 (9.15)	38.89 (10.05)
Education	4.11 (1.72)	4.61 (1.74)	4.03 (1.78)	4.29 (1.78)	4.88 (1.65)	4.62 (1.60)	4.80 (1.73)	5.19 (1.64)	5.19 (1.58)	4.93 (1.61)
City	.05 (.23)	.06 (.25)	.15 (.36)	.21 (.41)	.20 (.40)	.22 (.42)	.26 (.44)	.26 (.44)	.35 (.48)	.22 (.42)
Native	.91 (.29)	.92 (.27)	.93 (.26)	.93 (.26)	.90 (.30)	.97 (.16)	.96 (.20)	.97 (.18)	.98 (.14)	1.00 (0.00)
N	56	93	120	126	109	76	70	31	48	46

Table 8.2(b) shows the descriptive statistics for each number of social relations. The statistics show a pattern of increasing social relations at higher levels of education, for entrepreneurs living in a city, and for entrepreneurs that are native. The distribution of age and self-employed parents shows a less clear pattern at successive numbers of social relations.

8.4 Estimation results

Table 8.3 provides the results of our ordered discrete choice model. To calculate expected profit levels, we use the formula:

$$(8.7) \quad \hat{\Pi} = \hat{\alpha}_j + X_{ij} \hat{\beta}$$

Table 8.4 presents the industry specific choice probabilities implied by the ordered logit. To calculate the choice probabilities, we used the formulas in equation (8.6a) – (8.6d). To determine the effect of prices on the choice probabilities, we calculate both the choice

Table 8.3: Ordered choice probabilities for number of social relations (N=775)

Variables	Estimates	Standard Errors
Construction α_2	-0.527	(0.227)*
Trade α_3	-0.235	(0.175)
Manufacturing α_4	-0.322	(0.288)
Catering α_5	-0.315	(0.235)
Personal Services α_6	-0.413	(0.205)*
Parents Self-Employed β_1	0.279	(0.132)*
Age β_2	-0.024	(0.007)**
Education β_3	0.124	(0.039)**
City β_4	0.753	(0.160)* *
Native β_5	0.675	(0.252)**

Cutpoints	Estimates	Predicted profits	N	Mean (s.d)
ψ_1 (from 1 to 2 social relations)	-2.356 (.471)	$\hat{\pi}_1$ (one contact)	56	.013 (.496)
ψ_2 (from 2 to 3 social relations)	-1.206 (.457)	$\hat{\pi}_2$ (two contacts)	93	.122 (.468)
ψ_3 (from 3 to 4 social relations)	-.358 (.453)	$\hat{\pi}_3$ (three contacts)	120	.166 (.525)
ψ_4 (from 4 to 5 social relations)	.361 (.453)	$\hat{\pi}_4$ (four contacts)	126	.281 (.532)
ψ_5 (from 5 to 6 social relations)	.985 (.454)	$\hat{\pi}_5$ (five contacts)	109	.367 (.495)
ψ_6 (from 6 to 7 social relations)	1.484 (.456)	$\hat{\pi}_6$ (six contacts)	76	.441 (.498)
ψ_7 (from 7 to 8 social relations)	2.068 (.461)	$\hat{\pi}_7$ (seven contacts)	70	.536 (.517)
ψ_8 (from 8 to 9 social relations)	2.410 (.464)	$\hat{\pi}_8$ (eight contacts)	31	.478 (.475)
ψ_9 (from 9 to 10 social relations)	3.208 (.477)	$\hat{\pi}_9$ (nine contacts)	48	.545 (.577)
		$\hat{\pi}_{10}$ (ten contacts)	46	.466 (.514)

Notes. Business services is base category.
* significant at 5%; ** significant at 1%
Standard errors in parentheses.

probabilities including prices (indicated by T in Table 8.4), as well as setting prices at zero (indicated by I in Table 8.4). The choice probabilities correspond to the full model and the constant only model respectively. Based on the estimated choice probabilities, we calculate the corresponding expected equilibrium demand for productive social relations with the formula:

$$(8.8) \quad E(S_j) = \sum_{n=1}^{10} \text{Pr}(S_j) * S$$

where S denotes the number of social relations in industry j . To determine whether price differences exist between industries, we also calculate the average real price level for each industry utilizing the inverse effect of the parameter estimates on the real price proxy. Finally, we calculate the average profit level per industry.

A positive parameter in Table 8.3 implies that the probability of observing an increasing number of social relations increases if a covariate X increases. Rewriting equation (1) in terms of real prices:

$$(8.9) \quad \Pi(S_{ij}) = Y_{ij} - \frac{P_{ij}^s}{P^r} S_{ij}$$

Given fixed Π and Y , from equation (8.9) follows that if S increases, the real price $\frac{P^s}{P^r}$ decreases. A positive parameter in Table 8.3 not only implies an increase in the probability of increasing social relations, but also a decrease in the real price of accomplishing social relations.

Self-employed parents increase the probability of more social relations. Living in a city increases the probability of more social relations. Being native increases the probability of more social relations. The positive parameter estimates for education implies that an increase in the level of education increases the probability of more social relations. Note that the positive parameter estimate in Table 8.3 also implies that being native, living in a city, having a high education, and having self-employed parents reduces the cost of accomplishing social relations. Only age has a negative effect. Young entrepreneurs have more social relations than older entrepreneurs. Age increases the cost of accomplishing social relations.

Besides the effect of the real price proxy on social relations, the results in Table 8.3 indicate that there are significant differences in the demand for productive social relations between industries. The negative signs for the α 's imply that the demand for productive social relations is always lower in these industries compared to the business services. However, the negative signs are only significant for the construction and the personal services. Economic logic suggests that higher returns to social relations will induce higher investment rates. Combining this logic with our results would suggest that social relations are most productive in business services and least productive in construction and personal services.

Differences in the amount of social relations among industries may also be the result of a selection effect. There might be some unobserved heterogeneity, like the social ability and skills of the entrepreneur that are not captured by our covariates. Entrepreneurs with high social ability should endogenously select industries with high returns on

social relations. Self-selection does not undermine our argument that social relations are more productive among certain industries; it just provides another mechanism why observed investment levels may differ among industries.

The lower right part of Table 8.3 provides the predicted profits at successive level of social relations. The ordering of the profits follows the expected structure of increasing profit levels at successive number of social relations up to seven relations. From seven social relations onwards, profits remain relatively stable. Figure 8.3 sketches the ordering of predicted profits graphically. The figure shows that predicted profits increase at a decreasing rate. This result of negatively inclined demand curves reflects our proposition that overlapping information between social relations causes information redundancy.

Figure 8.3: Predicted expected profit levels by number of social relations

The results in Table 8.4 provide the choice probabilities disaggregated by industry. Setting prices at zero, the (I) column presents the baseline demand for productive social relations for each industry. The results show an inverse U-shape between the number of social relations and its likelihood, with a peak at three social relations. Correcting for prices, the inverse U-shape remains intact, but peaks at four social relations, except for trade which peaks at three. For all industries, prices increase the probability of more social relations. We use the choice probabilities to calculate the expected equilibrium demand for productive social relations for each industry. The results indicate that in these imperfect markets, entrepreneurs obtain full information at 4 to 5 social relations. More imperfect markets need more information to clear. With 5.2 social relations, business services have highest demand, and with 4.1, construction has lowest. Within the business services sector, products are more heterogeneous than in the construction sector. These heterogeneous products create uncertainty which increases potential transaction costs. Transaction costs may be overcome by the trust in its trading partners, providing a larger incentive to invest, explaining the higher demand for productive social relations in business services. Although in absolute terms a difference of one social relation seems trivial, this number indicates almost 27% more demand for social relations in business

services than in trade. Given our low estimated network size, it is not unlikely that we only captured each entrepreneur’s core social network. If each core network contact provides 5 more indirect contacts, then the total retrievable network would already increase to 26 for business services and 21 for trade, ignoring the possibility that core network contacts may lead to identical indirect contacts.

In Table 8.4, we also tested whether estimated price and profit level differences differed among industries. The ANOVA tests the null hypothesis of no differences in price levels among industries. We significantly reject the null hypothesis of no differences. The results show that real price levels are lowest in business services and highest in construction. The descriptive statistics in Table 8.2 provides some hints about the drivers of these price differences. In business services, entrepreneurs have a higher level of education and higher proportion of them live in cities. Table 8.3 shows that education and city significantly reduces the real price of accomplishing relations.

Table 8.4: Industry choice probabilities, equilibrium demand, prices, and profits

Probabilities (in %)		Business Services		Construc- tion		Trade		Manufactur- ing		Catering		Personal Services		Test
		I	T	I	T	I	T	I	T	I	T	I	T	
1	social relation	8.6	5.0	13.8	10.4	10.7	6.8	11.5	7.1	11.5	7.0	12.5	7.7	
2	social relations	14.4	9.3	19.8	16.4	16.8	11.9	17.7	12.3	17.6	12.2	18.6	13.2	
3	social relations	18.1	13.8	20.6	19.3	19.5	16.2	19.9	16.6	19.8	16.5	20.2	17.2	
4	social relations	17.8	16.4	16.6	17.6	17.6	17.5	17.3	17.6	17.4	17.6	17.1	17.7	
5	social relations	13.9	15.4	11.1	12.9	12.7	14.9	12.3	14.7	12.3	14.7	11.7	14.4	
6	social relations	8.7	11.2	6.3	7.8	7.6	9.9	7.2	9.7	7.2	9.8	6.8	9.3	
7	social relations	7.3	10.4	4.9	6.3	6.1	8.7	5.7	8.4	5.7	8.5	5.3	7.9	
8	social relations	3.0	4.6	1.9	2.5	2.5	3.7	2.3	3.5	2.3	3.6	2.1	3.3	
9	social relations	4.4	7.1	2.7	3.6	3.5	5.5	3.3	5.3	3.3	5.3	3.0	4.8	
10	social relations	3.9	6.7	2.3	3.2	3.1	5.0	2.8	4.8	2.9	4.8	2.6	4.4	
Expected equilib- rium demand		4.4	5.2	3.7	4.1	4.1	4.7	4.0	4.7	4.0	4.7	3.9	4.5	
														Anova
Average level	Price	-.578 (.474)		-.324 (.437)		-.502 (.480)		-.540 (.509)		-.546 (.491)		-.534 (.461)		F(5,769) 3.82**
Average level	Profit	.578 (.474)		-.202 (.437)		.266 (.480)		.218 (5.9)		.231 (.491)		.121 (.461)		F(5,769) 43.06**
N		301		82		164		44		73		111		

Notes. * significant at 5%; ** significant at 1%

We used the profit function of equation (8.7) to test whether profit differences exist between industries. The ANOVA tests the null hypothesis of no differences in profit levels among industries. Also, this null hypothesis is significantly rejected. In line with equilibrium demand for productive social relations, profits levels are lowest in construction and highest in business services.

8.5 Conclusion

In this paper we designed a model to explain the accomplishment of entrepreneurs' productive social networks in concentrated markets. The accomplishment of productive social networks in a business context follows economic logic. Individual incentives drive the formation of productive social networks. Differences in expected returns and costs trigger the equilibrium demand for productive social relations.

A discrete choice formulation is used to draw inferences about the unobserved profitability of entrepreneur's social network. The analysis shows that expected profitability increases at successive levels of social relations. Moreover, social contacts differ in productivity across markets. When markets deviate from their perfect form, frictions and uncertainty emerge and transaction and search costs arise. Social relations may dampen uncertainty and overcome or capitalize on frictions and thereby reduce these costs. Whereas the demand for productive social contacts emerges from the peculiarities of the industry the entrepreneur operates in, the price of establishing social contacts follows from entrepreneur heterogeneity. The price of social contacts is not only a function of individual characteristics, like education, social background, age, and nationality, but also a matter of location. Living in a city is a cheap source of social resources.

Understanding why investment in social capital, like social networks, differs between individuals is important for government policies. We echo Gleaser et al. (2002: 456) that "individuals accumulate social capital when the private incentives for such accumulation are high". However, we add that providing incentives to invest is only one side of the medal. Opportunities for investment in social capital are still structurally segmented, by race, education, and social class. As investments in human capital may be inhibited by financing constraints, investments in social capital may be constrained by social structure.

Chapter 9

Summary & conclusion

9 Summary & conclusion

9.1 Summary

At the end of this dissertation, we take some time to reflect on what we have learned and, also, in what directions future research can be undertaken. In this dissertation, we utilized two datasets to answer several research questions related to entrepreneurship and self-employment. In chapter 3 and 4 we investigated individuals' entry decisions with the aid of the North Brabant dataset. In chapter 6 we studied non-response related to business idea continuance in the newly created Limburg dataset. Chapter 7 investigated entrepreneurs exit decision when a fraction of the set up costs is sunk. Chapter 8 scrutinized the social network accumulation decisions of entrepreneurs. We have studied these entrepreneurial decisions as an individual's choice being sensitive to the entrepreneur social context.

In this dissertation, we started from the premise that the entrepreneur is so strongly embedded in his social context, both horizontally through space by the social and economic structure and vertically through time by his past actions and choices and future aspirations, that entrepreneurial decisions can not be researched from a static atomized perspective, but should be studied as a dynamic phenomenon mutually reinforcing social and economic context.

In chapter 3, our analysis confirms the superiority of the dynamic duration models to model the transition to self-employment to their static counterparts applied to multi-period data. We have shown theoretically that the static model is merely an aggregated version of a duration model and therefore always inefficient. If the process that generates the data is the continuous Poisson process and the data is grouped into intervals, then the natural candidate for an empirical dynamic model is a grouped duration model with a type I extreme value error distribution (cloglog). We have studied the consequences of time aggregation in a study on self-employment selection in the Netherlands between 1953 and 1993. The results showed the clear superiority of the dynamic models in terms of parameter estimate precision and prediction accuracy. Time aggregation can seriously bias parameter estimates. For example, we have shown that the self-employed father increases the probability to become self-employed in the dynamic model with 40 up to 200% points as compared with the static estimator. The dynamic model with duration dependence classifies 95% more individuals in the highest probability decile than the static model. Estimation precision is important for policy makers that aim to predict who becomes an entrepreneur. Furthermore, our dynamic approach explicitly models the timing of the entry decision providing information for policy makers when individuals are most at risk of self-employment.

In chapter 4, we extended the knowledge on the timing of the entry decisions by differentiating two routes into self-employment: taking over control in a family business or founding a new business. We have developed an economic theory of the self-employment process, where an individual's human, financial, and social capital determine the waiting time into self-employment. Our analysis reveals clear differences between people who took over an existing family business and people who founded a

new firm. The results from the analysis show that the hazard of entering self-employment is age dependent and the shape of the duration dependence is self-employment type dependent. Because learning may take place in the family business and less sunk costs have to be financed, the hazard of taking over a family business is highest in the earliest stages of one's professional career. In contrast, starting up a business from scratch takes more time. First the would-be self-employed has to accumulate enough skills, capabilities, and capital to overcome set up costs and to become self-employed. The real founder's probability of entering self-employment is initially low and increases only gradually over his life. Therefore, taking over a family business cannot be compared by starting up a business from scratch. Continuation of a family business is like entering self-employment with a valuable set of assets already existent, while real founders enter self-employment with the burden of financing the sunkness of the start up expenses. Furthermore, the uncertainty of the expected profitability is much lower for a firm that has learned about its true productivity, than for the *de novo* firm whose expected profitability is still more a reflection of the subjective beliefs of the business owner.

Furthermore, chapter 4 investigated the determinants of self-employment selection. Next to an individual's human capital, our results indicate that an individual's social context is an important predictor of self-employment selection. The entrepreneur's social context both promotes and inhibits the self-employment decision. Our analyses reveal that self-employed parents transfer not only entrepreneurial skills, but also the opportunity to take over the family business. Parents with a high education that are probably also wealthier are especially important for real founders who face higher liquidity constraints. A spouse increases self-employment because he or she provides emotional and financial support as well as cheap labor. On the other hand, children sometimes constrain the choice for self-employment. And coming from large entrepreneurial family constraints the possibility to take over the family business.

The results from our analyses in chapter 3 and chapter 4 have important implications for the field of entrepreneurship and self-employment. Results of previous (empirical) research that can not differentiate between family self-employment and *de novo* self-employment must be reconsidered in the light of our study (e.g. Evans & Leighton 1989; Dunn & Holtz-Eakin 2000). First and foremost, the time dimension must be brought back in theorizing and econometric modeling of self-employment selection. Furthermore, the results of chapter 4 have shown that aggregating family self-employment and *de novo* self-employment might offset the dynamic nature of self-employment. Aggregating taking over a family business and starting from scratch resulted in a constant hazard for a considerable period of our researched time period. Moreover, disentangling the different routes into self-employment alters the causal nature of some of the determinants of self-employment selection.

Starting entrepreneurs were the focus in chapter 3 and chapter 4. However, answering the question who becomes self-employed and who not requires also that persons who are not self-employed are part of the research population. After all, without such a control group of non-self-employed it is not possible to show what differentiates self-employed from non-self-employed. The Limburg dataset was created to investigate the choice for entrepreneurship. For making inferences from the collected data, it is of utmost importance that as many possible individuals are willing to cooperate in the research. However, a part of the approached individuals were for one or the other reason

no willing to collaborate. In chapter 6, we investigated the sources of non-response in general and the link between the outcome of the self-employment decision and non-response in particular. The results from our models suggest several sources of non-response. One of the reasons of non-response is attrition, e.g. because of death, from the sample. Furthermore, some individuals are more likely to non-respond than others. However, the most important finding from our analyses is that there is a structural relationship between the probability that an individual decides to abandon his business idea and his subsequent probability of non-response. Individuals that continued their business idea were more likely to partake in the personal interview, while individuals that abandoned their business idea were less likely to partake in the personal interview. As a result, researchers and policy makers that use surveys to investigate the self-employment decision must be aware of this causal bias. Abandoning business ideas in the population is probably more common than the survey results will show, leading to possible underestimation of possible constraints.

In chapter 7 we shifted our attention from the entry decision to the exit decision. In this chapter we designed a theoretical model to explain the reluctance of entrepreneurs to exit self-employment in the presence of sunk costs and an uncertain evolution of self-employment income. The results indicate that the sunk fraction of the set up costs make entrepreneurs reluctant to exit self-employment. The fact that sunk costs lead to reluctance provides evidence for the existence of option values associated with the self-employment decision. A business owner may accept a temporary loss and decide to remain self-employed and wait if business conditions turn favorable again to recoup partly sunk investment costs.

Furthermore, this chapter investigated why entrepreneurs invest different amounts of sunk costs in their firm. Our results show that the level of wage earned in the last job prior entry is an important determinant of the initial commitment in sunk set up costs. This confirms our prediction from an entrepreneurial choice model extended with sunk set up costs. The entrepreneurial choice model applied in this chapter shows that individuals have to raise their capital investments to overcome individual entry thresholds. The entry threshold is the income one can earn as a wage worker. Therefore, the entrepreneurial choice model dictates that individuals with high previous income levels have to invest more capital to raise their entrepreneurial income above the entry threshold level to make it profitable to become an entrepreneur and enter self-employment. We find that the elasticity of previous income level to initial commitment in sunk capital investments is .294. Thus a one percent increase in the level of income prior to entry implies a .294 percent increase in initial capital investments.

Although we do not formally investigate the option value associated with the entry decision, its implications are obvious and provide some evidence for the hitherto unexplained result in the self-employment literature that high income individuals are less likely to become self-employed. High income individuals have to invest more sunk capital to raise entrepreneurial earnings above entry thresholds. However, since future earnings are uncertain, they have an option value to wait. Therefore, high income individuals will be more reluctant to enter self-employment than low income individuals. This results from chapter 7 complement the results from chapter 4 and 5. We believe that the timing of the self-employment decisions is important to understand who becomes self-employed and what the founding characteristics of the firm are. The decision to defer

self-employment might increase the opportunity costs of becoming self-employed, e.g. the wage earned in the labor market.

In chapter 8, we investigated one possible factor of successful self-employment, namely an entrepreneur's social network. In this chapter, we modeled the demand for productive social relations as an outcome of profit maximizing behavior under uncertainty. Social relations are considered productive inputs in the production process of firms. The productivity of a relation can vary across industries because of differences in the production processes between industries. Relations can also be more productive when markets are not perfectly competitive. Such markets are often characterized by search and transaction costs. Therefore, seemingly identical social relations may differ in productivity to the extent that they are able to overcome market imperfections and optimize the production process of firms within an industry. In our model the costs of accomplishing social relations were heterogeneous. Equilibrium demand is identified in the market for productive relations where marginal returns of accomplishing relations equal marginal costs.

The model's predictions are tested using the Limburg dataset. The econometric analysis shows that young, native, highly educated entrepreneurs in large cities, with self-employed parents, have lower cost of forming productive social relations. The ordering of profits shows decreasing returns to scale for social relations. On average, entrepreneurs obtain full information at 4 to 5 social relations. We conclude that the equilibrium demand for productive social relations can be derived from the demand conditions emerging from industry structure and the supply conditions emanating from entrepreneurs' heterogeneity.

Understanding why investment in social capital, like social networks, differs between individuals is important for government policies. Individuals accumulate social capital when the incentives for such accumulation are high and the costs for accomplishing social capital are low. However, our study shows that the opportunities for investment in social capital are still structurally segmented, by race, education, and social class. As investments in human capital may be inhibited by financing constraints, investments in social capital may be constrained by social structure.

Reflecting on these results, what have we learned of the influence of the social context on entrepreneurial decisions? Most importantly, we have shown the strong social embeddedness of the entrepreneur. First of all, the social background of an individual is important to understand the chance that he will become an entrepreneur. Individuals with self-employed parents have a higher chance of becoming self-employed and enter self-employment faster than individuals without self-employed parents. This dissertation found several reasons for this finding. We agree with other researchers that self-employed parents can transfer skills and capital to their offspring. Furthermore, we add to this understanding that a significant part of the effect of self-employed parents on self-employment selection rest in the fact that self-employed parents can provide their offspring the opportunity to take over the control in a family business. However, the effect of self-employed parents diminishes over time as the individual increases his entrepreneurial abilities through education and work experience and increases his own wealth through saving. But also after the individual enters self-employment, the self-employed parent can be important. Our study shows that self-employed parents are an important source for the accomplishment of productive social relations. Entrepreneurs from self-

employed parents have larger productive social networks and entrepreneurs with large social networks are more successful entrepreneurs. Highly educated and probably wealthier parents are especially important for real founders who do not take over a business and who have to overcome sunk set up costs and who face more uncertain success chances. However, the social background is not always productive for the prospective entrepreneur. The more siblings an individual has, the fewer are his chances to take over the family business.

But the context of the entrepreneur transcends his social background. The past actions of the entrepreneur bear a considerable influence on his future behavior. His past accumulation decisions influence the timing of the entry decision. Investments in education relax both human capital and liquidity constraints through the learning of entrepreneurial capabilities and the accumulation of capital through higher earnings. However, the accumulation of labor market job specific skills and capabilities can make individuals reluctant to exit their labor market job and enter self-employment. The influence of path dependence becomes especially important if one investigates the effect of labor market experience on the founding characteristics of the firm. Next to the accumulation of labor market job specific human capital, experience and high income from labor market jobs create idiosyncratic entry thresholds. The individual only enters self-employment if his income from self-employment exceeds labor market income. In order to raise entrepreneurial earnings the individual has to invest partially sunk capital in the form of set up costs. Individuals from high labor market income cohorts must therefore overcome more set up costs. Although high income individuals are less liquidity constraints, our study has shown that sunk costs make individuals reluctant to enter or exit self-employment when faced with uncertain future entrepreneurial earnings. If sunk costs do not only make individuals reluctant to exit self-employment but also to enter self-employment, then individuals from high income cohorts will be both less likely to enter as exit self-employment, a process called hysteresis.

For these reasons, we provide ample evidence for the embeddedness of the entrepreneur in its social context, and the influence of being embedded in a social context on the decisions of the entrepreneur.

9.2 *Suggestions for future research*

Thanks to the dynamic nature of our datasets, we could show many interesting aspects of the self-employment decisions not available in static studies. The rejection of the static nature of the self-employment decision opens new paths of inquiry. Studying the waiting time into self-employment bears close resemblance with studying the real options nature of investments. The real options nature of investments dictates that in the presence of irreversible costs, the uncertainty of self-employment success can make individuals reluctant to become self-employed and reluctant to exit self-employment. This may be especially true for real founders whose success chances are more uncertain and who face higher partially irreversible set up costs.

Our research provided empirical evidence of the real options nature of the self-employment decision. Further research is needed to assess the empirical validity of the real options nature of the entry decision. Furthermore, our data allowed us only to use

proxies for the key parameters in the real options model. Future research should try to obtain direct measures or better proxies of sunk costs, idiosyncratic uncertainty, and idiosyncratic profit growth. Especially measuring idiosyncratic uncertainty is the “holy grail for the empirical analysis of dynamic irreversible adjustment models” (Pfann 2006: 166).

From an empirical perspective, more research is needed that investigate the dynamics of the self-employment decision. Our study has shown the time dependent shape of the entry hazard both for real founders and for individuals that take over control in a family business. Other samples from different countries must be consulted to generalize the time dependence of self-employment. Furthermore, our analyses have shown that variables can have different effects over time. Future research on the dynamics of self-employment selection can make use of parametric spline models that can account for these differences.

Furthermore, our analyses of different types of self-employment seemed fruitful. Further research can investigate the dynamics of other types of self-employment like the Schumpeterian innovator to assess its dynamics and determinants. Multiple risk models are the preferred techniques to estimate these models. These multiple risk models can also be extended to the analysis of exit behavior of different types of self-employment. An obvious candidate would be a study investigating the exit dynamics of de novo businesses and taken over family business.

Our study has shown that self-employment is strongly embedded within a social context. Future research should follow this endeavor providing a more complete picture of the entrepreneur and its origins. For example, there is a promising literature studying discrete choices with social interactions (Brock and Durlauf 2001). Unfortunately, their findings have not been incorporated in this study, but studying self-employment choices within their framework can provide interesting information on determinants of self-employment selection. Furthermore, this dissertation provided only a limited view of the social context. We have limited ourselves to the micro social context and have not addressed the macro social context. Studies that incorporate both the micro and macro social context can be more conclusive in assessing the relevant worth of both approaches. More studies are needed that explain why and how social relations make individuals more productive. We have investigated how differences between industries and difference in accomplishment costs influence the use of social networks. However, our study touches the benefits of social relations only superficially by observing differences among industries. More studies are needed that investigate smaller sectors and explicitly model the industry characteristics related to transaction and search costs. Our discrete choice model for productive social relations can be easily extended to allow industry characteristics as determinants to the use of social networks. Moreover, our analysis of social network size can be extended with models that focus not only on network size but also on the type and strength of the relation and the structure of the social network.

With these suggestions for future research we have reached the end of this dissertation, in which we have tried to do justice to the dynamic nature of entry and exit to the market for self-employed entrepreneurs. Furthermore, we have showed the importance of the social context to these processes. We have shown that the entrepreneur is not just an isolated phenomenon outside its social context, but that he is strongly embedded in its social structure that at the same time can promote and hinder his development.

Bibliography

Bibliography

- ADDISON, J. T. and P. PORTUGAL (1987). "On the distributional shape of unemployment duration." Review of Economics and Statistics **69**: 520-526.
- ALBA-RAMIREZ, A. (1994). "Self-employment in the midst of unemployment: The case of Spain and the United States." Applied Economics **26**(354-367).
- ALLISON, P. (1982). Discrete time methods for the analysis of event histories. Sociological Methodology 1982. S. Leinhardt. San Francisco, Jossey-Bass.
- ALLISON, P. (1984). Event History Analysis. Newbury Park CA, Sage.
- ALT, J. E., G. KING and C. S. SIGNORINO (2000). "Aggregation among binary, count, and duration models: Estimating the same quantities from different levels of data." Political Analysis **9**: 21-44.
- AMIT, R., E. MULLER and I. COCKBURN (1995). "Opportunity costs and entrepreneurial activity." Journal of Business Venturing **10**: 95-106.
- ASPLUND, M. (2000). "What fraction of a capital investment is sunk costs?" Journal of Industrial Economics **XLVIII**(3): 287-304.
- AUDRETSCH, D. B. and T. MAHMOOD (1995). "New Firm Survival: New Results Using a Hazard Function." Review of Economics and Statistics **77**(1): 97-103.
- BARR, A. (2000). "Social Capital and Technical Information Flows in the Ghanaian Manufacturing Sector." Oxford Economic Papers **52**(3): 539-59.
- BASU, A. and S. C. PARKER (2001). "Family finance and new business start-ups." Oxford Bulletin of Economics and Statistics **63**: 333-358.
- BATES, T. (1990). "Entrepreneur human capital inputs and small business longevity." Review of Economics and Statistics **72**(4): 551-59.
- BATES, T. (1997). Race, Self-Employment and Upward Mobility. Baltimore, John Hopkins University Press.
- BECKER, G. S. (1964). Human Capital. New York, Columbia University Press.
- BECKER, G. S. (1996). Accounting for Tastes. Cambridge, Ma, Harvard University Press.
- BECKER, G. S. and K. M. MURPHY (2000). Social Economics: Market Behavior in a Social Environment. Cambridge, Ma, Harvard University Press.
- BECKER, G. S. and N. TOMES (1979). "An equilibrium theory of the distribution of income and intergenerational mobility." Journal of Political Economy **87**: 1153-1189.
- BEN-PORATH, Y. (1980). "The F-connection: Families, friends, and firms and the organization of exchange." Population and Development Review **6**(1): 1-30.
- BERGSTRÖM, R. and P. A. EDIN (1992). "Time aggregation and the distributional shape of unemployment duration." Journal of Applied Econometrics **7**: 5-30.
- BIERLAIRE, M. (2003). "Biogeme: A free package for the estimation of discrete choice models." Proceedings of the 3rd Swiss Transportation Research Conference.
- BLANCHFLOWER, D. G. and B. D. MEYER (1994). "A longitudinal analysis of the young self-employed in Australia and the United States." Small Business Economics **6**: 1-19.
- BLANCHFLOWER, D. G. and A. J. OSWALD (1998). "What Makes an Entrepreneur?" Journal of Labor Economics. January **16**(1): 26-60.
- BLAU, P. O. and O. D. DUNCAN (1967). The American Occupational Structure. New York, Wiley.
- BLUMBERG, B. F., J. B. KOK and G. A. PFANN (2006). The dynamics of business succession and de novo self-employment: A multiple risks approach. Maastricht University.
- BOOTH, A. L. and S. E. SATCHELL (1995). "The hazards of doing a PhD: An analysis of completion and withdrawal rates of British PhD students in the 1980's." Journal of the Royal Statistical Society, Series A **158**: 297-318.
- BORJARS, G. J. and S. G. BRONARS (1989). "Consumer discrimination and self-employment." Journal of Political Economy **97**: 581-605.
- BOSMA, N., M. VAN PRAAG, R. THURIK and G. DE WIT (2004). "The value of human and social capital investments for the business performance of startups." Small Business Economics **23**: 227-236.
- BOURDIEU, P. (1986). The forms of capital. Handbook of Theory and Research the Sociology of Education. J. G. Richardson. Westport, Greenwood: 241-258.

- BOXMAN, A. W., P. M. DE GRAAF and H. D. FLAP (1991). "The impact of social and human capital on the income attainment of Dutch managers." Social Networks 13: 51-73.
- BROCK, W. A. and S. N. DURLAUF (2001). "Discrete choice with social interactions." Review of Economic Studies 68(2): 235-60.
- BROCKHAUS, R. H. (1982). The psychology of the entrepreneur. The Encyclopedia of Entrepreneurship. C. Kent, D. Sexton and K. Vesper. Englewood Cliffs, Prentice Hall.
- BRÜDERL, J. and P. PREISENDÖRFER (1998). "Network support and the success of newly founded businesses." Small Business Economics. May 10(3): 213-25.
- BURT, R. S. (1992). Structural Holes: The Social Structure of Competition. Cambridge, MA, Harvard University Press.
- BURT, R. S. (2002). "Bridge Decay." Social Networks 24(4): 333-363.
- CALVO, G. and S. WELLISZ (1980). "Technology, entrepreneurs, and firm size." Quarterly Journal of Economics 95: 663-677.
- CAMPA, J. M. (1993). "Entry by foreign firms in the United-States under exchange-rate volatility." Review of Economics and Statistics 75: 614-622.
- CANTILLON, R. (1755). Essai sur la nature du commerce en général. T. Tsuda. Tokyo, Kinokuniya.
- CARLAND, J. W., F. HOY, W. R. BOULTON and J. A. C. CARLAND (1984). "Differentiating entrepreneurs from small business owners: A conceptualization." The Academy of Management Review 9(2): 354-359.
- CARROLL, G. R. (1983). "Dynamic analysis of discrete dependent variables: A didactic essay." Quality and Quantity 17: 425-460.
- CARROLL, G. R. and E. MOSAKOWSKI (1987). "The career dynamics of self-employment." Administrative Science Quarterly 32: 570-589.
- CHESHER, A. and K. LANCASTER (1983). "The estimation of models of labour market behaviour." Review of Economic Studies 50: 609-624.
- CLEVES, M., W. GOULD and R. GUTIERREZ (2004). An Introduction to Survival Analysis using Stata: Revised Edition. College Station, Stata Press.
- COLEMAN, J. S. (1986). "Social theory, social research, and a theory of action." American Journal of Sociology 91: 1309-1335.
- COLEMAN, J. S. (1988). "Social capital in the creation of human capital." American Journal of Sociology 94: S95-S120.
- COLEMAN, J. S. (1990). Foundations of Social Theory. Cambridge, Ma, Harvard University Press.
- COOPER, A. C. and W. C. DUNKELBERG (1986). "Entrepreneurship and paths to business ownership." Strategic Management Journal 7: 53-68.
- COSH, A. D., A. HUGHES and E. WOOD (1999). Innovation in UK SMEs: Causes and consequences for firm failure and acquisition. Entrepreneurship, Small and Medium Sized Enterprises and the Macro Economy. A. Z.J. and B. Carllson. Cambridge, Cambridge University Press.
- COX, D. (1972). "Regression models and life-tables (with discussion)." Journal of the Royal Statistical Society, Series B 34: 187-220.
- COX, D. and D. OAKES (1984). Analysis of Survival Data. London, Chapman Hall.
- CRAMER, J. S., J. HARTOG, N. JONKER and C. M. VAN PRAAG (2002). "Low risk aversion encourages the choice for entrepreneurship: An empirical test of a truism." Journal of Economic Behavior & Organization 48: 29-36.
- DE GRAAF, N. D. and H. D. FLAP (1988). "'With a little help from my friends': Social resources as an explanation of occupational status and income in West Germany, The Netherlands, and the United States." Social Forces 67(2): 452-472.
- DE WIT, G. and F. A. VAN WINDEN (1989). "An empirical analysis of self-employment in the Netherlands." Small Business Economics 1: 263-272.
- DIXIT, A. (1989). "Entry and Exit Decisions under Uncertainty." The Journal of Political Economy 97(3): 620-638.
- DIXIT, A. and R. S. PINDYCK (1994). Investment under Uncertainty. Princeton, Princeton University Press.
- DIXIT, A. and R. ROB (1994). "Switching costs and sectoral adjustments in general equilibrium with uninsured risk." Journal of Economic Theory 62: 48-69.
- DOLTON, P. J. and G. H. MAKEPEACE (1990). "Self-employment among graduates." Bulletin of Economic Research 42: 35-53.

- DUESSENBERRY, J. (1960). Comment on "An economic analysis of fertility". Demographic and Economic Change in Developed Countries. U.-N. B. C. f. E. Research. Princeton, N.J., Princeton University Press.
- DUNN, T. and D. HOLTZ-EAKIN (2000). "Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links." Journal of Labor Economics 18(2): 282-305.
- ELBERS, C. and G. RIDDER (1982). "True and spurious duration dependence: The identifiability of the proportional hazard model." Review of Economic Studies 49: 403-409.
- EPISCOPOS, A. (1995). "Evidence on the relationship between uncertainty and irreversible investment." Quarterly Review of Economics and Finance 35: 41-52.
- EVANS, D. S. and B. JOVANOVIĆ (1989). "An estimated model of entrepreneurial choice under liquidity constraints." Journal of Political Economy 97(4): 808-27.
- EVANS, D. S. and B. JOVANOVIĆ (1989). "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." The Journal of Political Economy 97(4): 808-827.
- EVANS, D. S. and L. S. LEIGHTON (1989). "Some empirical aspects of entrepreneurship." American Economic Review 79(3): 519-35.
- FAFCHAMPS, M. and B. MINTEN (2002). "Returns to social network capital among traders." Oxford Economic Papers 54(2): 173-206.
- FAVERO, C. A., A. PESARAN, M. HASHEM and S. SHARMA (1994). "A duration model of irreversible investment: Theory and empirical evidence." Journal of Applied Econometrics 9: S95-S112.
- FLAP, H. D. (1990). Conflict, Violence, and Loyalty. Bern, Peter Lang.
- FRANK, M. (1988). "An intertemporal model of industrial exit." Quarterly Journal of Economics 103: 333-344.
- FRANK, R. H. (1992). "Melding sociology and economics: James Coleman's foundations of social theory." Journal of Economic Literature 30: 147-170.
- FRIEDMAN, M. (1953). Essays in Positive Economics. Chicago, University of Chicago Press.
- FUKUYAMA, F. (1995). Trust: The Social Virtues and the Creation of Prosperity. New York, Free Press.
- GEROSKI, P. A. and R. POMROY (1990). "Innovation and the evolution of market structure." Journal of Industrial Economics 91: 841-855.
- GIMENO, J., T. B. FOLTA, A. C. COOPER and C. Y. WOO (1997). "Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms." Administrative Science Quarterly 42: 750-783.
- GLAESER, E. L. (2001). "The formation of social capital." Isuma 2(1): 1-13.
- GLAESER, E. L., D. LAIBSON and B. SACERDOTE (2002). "The economic approach to social capital." The Economic Journal 112: 437-458.
- GRAMBSCH, P. M. and T. M. THERNEAU (1994). "Proportional hazards tests and diagnostics based on weighted residuals." Biometrika 81: 515-526.
- GRANOVETTER, M. (1973). "The strength of weak ties." American Journal of Sociology 78: 1360-1380.
- GRANOVETTER, M. (1974). Getting a Job. Cambridge, MA, Harvard University Press.
- GRANOVETTER, M. (1985). "Economic action, social structure, and embeddedness." American Journal of Sociology 91: 481-510.
- GRONAU, R. and D. S. HAMERMESH (2001). "The demand for variety: A household production perspective." NBER Working Paper 8509.
- GUIO, L., P. SAPIENZA and L. ZINGALES (2000). "The role of social capital in financial development." NBER Working Paper 7563.
- HAN, A. and J. A. HAUSMAN (1990). "Flexible parametric estimation of duration and competing risk models." Journal of Applied Econometrics 5(1): 1-28.
- HANNAN, M. T. and J. FREEMAN (1989). Organizational Ecology. Cambridge, Harvard University Press.
- HARTOG, J. (1989). "Survey non-response in relation to ability and family background: Structure and effects on estimated earnings functions." Applied Economics 21(3): 387-95.
- HASTIE, T. J. and R. J. TIBSHIRANI (1993). "Varying-coefficients models (with discussion)." Journal of the Royal Statistical Society, Series B 55: 757-796.
- HECKMAN, J. J. and B. SINGER (1984). "Econometric duration analysis." Journal of Econometrics 24: 63-132.
- HOLMES, T. J. and J. A. SCHMITZ (1990). "A theory of entrepreneurship and its application to the study of business transfers." Journal of Political Economy 98: 265-294.

- HOLMES, T. J. and J. A. SCHMITZ (1996). "Managerial tenure, business age, and small business turnover." Journal of Labor Economics **15**: 283-291.
- HOLMES, T. J. and J. A. SCHMITZ (1996). "Nonresponse bias and business turnover rates: The case of the characteristics of business owners survey." Journal of Business and Economic Statistics **14**: 231-241.
- HOLTZ-EAKIN, D., D. JOULFAIAN and H. S. ROSEN (1994). "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints." Journal of Political Economy **102**(1): 53-75.
- HOLTZ EAKIN, D., D. JOULFAIAN and H. S. ROSEN (1994). "Entrepreneurial decisions and liquidity constraints." Rand Journal of Economics **25**(2): 334-47.
- HOMANS, G. (1950). The Human Group. New York, Harcourt, Brace.
- HOPENHAYN, H. A. (1992). "Entry, exit, and firm dynamics in long run equilibrium." Econometrica **60**(5): 1127-1150.
- HURN, A. S. and R. E. WRIGHT (1994). "Geology or economics? Testing models of irreversible investment using North Sea Oil Data." Economic Journal **104**: 363-371.
- JENKINS, S. (1995). "Easy ways to estimate discrete time duration models." Oxford Bulletin of Economics and Statistics **57**: 129-138.
- JENKINS, S. (2004). Survival Analysis. University of Essex.
- JOHNSON, W. (1978). "A theory of job shopping." Quarterly Journal of Economics **22**: 261-278.
- JONKER, N. (1995). Vervolgonderzoek Noord Brabantse Zesdeklassers codebook (Continued Research Project Sixth Formers North Brabant Codebook). mimeo. Amsterdam.
- JOVANOVIĆ, B. (1982). "Selection and the evolution of industry." Econometrica **50**(3): 649-670.
- JOVANOVIĆ, B. and S. LACH (1989). "Entry, exit, and diffusion with learning by doing." American Economic Review **79**: 690-699.
- KAUERMANN, G., G. TUTZ and J. BRÜDERL (2005). "The survival of newly founded firms: a case study into varying-coefficient models." Journal of the Royal Statistical Society, Series A **168**: 145-158.
- KENNEDY, P. (2003). A Guide to Econometrics. Oxford, Blackwell Publishing.
- KIHLSTROM, R. E. and J.-J. LAFFONT (1979). "A general equilibrium entrepreneurial theory of firm formation based on risk aversion." Journal of Political Economy **87**: 719-748.
- KNACK, S. and P. KEEFER (1997). "Does social capital have an economic payoff? A cross-country investigation." Quarterly Journal of Economics **112**(4): 1251-88.
- KNIGHT, F. H. (1921). Risk, Uncertainty, and Profit. Chicago, Chicago university press.
- KOK, J. B. and G. A. PFANN (2006). A discrete choice model for productive social relations. Maastricht University.
- KOK, J. B. and G. A. PFANN (2006). Non-response bias and business discontinuance rates. Maastricht University.
- KOK, J. B. and G. A. PFANN (2006). Reluctance to exit self-employment when a fraction of the set up cost is sunk. Maastricht University.
- KRANTON, R. E. (1996). "Reciprocal exchange: A self-sustaining system." American Economic Review **86**(4): 830-51.
- LABAND, D. N. and B. F. LENTZ (1983). "Like father, like son: Toward an economic theory of occupational following." Southern Economic Journal **50**: 474-493.
- LAI, G. W., N. LIN and S. LEUNG (1998). "Network resources, contact resources, and status attainment." Social Networks **20**(2): 159-178.
- LAMBSON, V. E. and F. E. JENSEN (1998). "Sunk costs and firm value variability: Theory and evidence." American Economic Review **88**: 307-313.
- LANCASTER, T. (1990). The Econometric Analysis of Transition Data. Cambridge, Cambridge University Press.
- LAPORTA, R., F. LOPEZ-DE-SALANES, A. SHLEIFER and R. VISHNY (1997). "Trust in large organizations." American Economic Review Papers and Proceedings **87**: 333-338.
- LAUMANN, E. O. (1966). Prestige and Association in an Urban Community. Indianapolis, Bobbs-errill.
- LAZEAR, E. (2004). "Balanced skills and entrepreneurship." American Economic Review **94**: 208-211.
- LAZEAR, E. (2005). "Entrepreneurship." Journal of Labor Economics **23**(4): 649-680.
- LE, A. T. (1999). "Empirical studies of self-employment." Journal of Economic Surveys **13**: 381-416.
- LIGHT, I. and E. BONACICH (1988). Immigrant Entrepreneurs: Koreans in Los Angeles, 1965-1982. Berkely, University of California Press.
- LIN, N. (1990). Social resources and social mobility: a structural theory of status attainment. Social Mobility and Social Structure. R. Breiger, L. Cambridge, Cambridge University Press: 247-271.

- LIN, N. (1999). "Social networks and status attainment." Annual Review of Sociology 25: 467-487.
- LINDEBOOM, M. and J. THEEUWES (1991). "Job duration in the Netherlands: The co-existence of high turnover and permanent job attachment." Oxford Bulletin of Economics and Statistics 53: 243-264.
- LINDH, T. and H. OHLSSON (1996). "Self-employment and windfall gains: Evidence from the Swedish Lottery." Economic Journal 106: 1515-1526.
- LOURY, G. C. (1977). A dynamic theory of racial income differences. Women, Minorities, and Employment Discrimination. P. A. Wallace and A. M. LaMond. Lexington, MA, Lexington Books: 153-186.
- LUCAS, R. E. (1978). "On the size distribution of business firms." Bell Journal of Economics 9: 508-523.
- MARSHALL, A. (1890). Principles of Economics. London, Macmillan and co.
- MAS-COLELL, A., M. D. WHINSTON and J. R. GREEN (1995). Microeconomic Theory. New York, Oxford University Press.
- MATA, J. and P. PORTUGAL (1994). "Life duration of new firms." Journal of Industrial Economics XLII(3): 227-245.
- MCCLELLAND, D. C. (1961). The Achieving Society. Princeton, Van Nostrand.
- McFADDEN, D. (1984). Econometric analysis of qualitative response models. Handbook of Econometrics. M. Intriligator and Z. Griliches. Amsterdam, Elsevier. **volume II**: 1396-1456.
- MEYER, B. D. (1990). "Unemployment insurance and unemployment spells." Econometrica 58: 757-782.
- MILL, J. S. (1848). Principles of Political Economy with some of their Applications to Social Philosophy. London, John W. Parker.
- MULLIGAN, C. B. (1999). "Galton versus the human capital approach to inheritance." Journal of Political Economy 107(6): S184-224.
- NARENDRENATHAN, W. and B. M. STEWART (1991). "Simple methods for testing the proportionality of cause-specific hazards in competing risks models." Oxford Bulletin of Economics and Statistics 53: 331-340.
- O'BRIEN, J. P., T. B. FOLTA and D. R. JOHNSON (2003). "A real options perspective of entrepreneurial entry in the face of uncertainty." Managerial and Decision Economics 24: 515-533.
- PAKES, A. and R. ERICSON (1998). "Empirical implications of alternative models of firm dynamics." Journal of Economic Theory 79(1): 1-46.
- PARKER, S. C. (1996). "A time series model of self-employment under uncertainty." Economica 63: 459-475.
- PARKER, S. C. (1997). "The effects of risk on self-employment." Small Business Economics 9: 515-522.
- PARKER, S. C. (2004). The Economics of Self-Employment and Entrepreneurship. Cambridge, Cambridge university press.
- PENNINGS, J. M., K. LEE and A. VAN WITTELOOSTUIJN (1998). "Human capital, social capital, and firm dissolution." Academy of Management Journal 41(4): 425-440.
- PENROSE, E. (1959). The Theory of the Growth of the Firm. Oxford, Oxford University Press.
- PETERSEN, T. (1991). Time-aggregation bias in continuous-time hazard-rate models. Sociological Methodology 1991. P. V. Marsden. Cambridge MA, Blackwell.
- PETERSEN, T. and K. W. KOPUT (1992). "Time-aggregation hazard-rate models with covariates." Sociological Methods and Research 21: 25-51.
- PFANN, G. A. (2001). "Options to quit." Economic Letters 70: 259-265.
- PFANN, G. A. (2006). "Downsizing and heterogeneous firing costs." Review of Economics and Statistics 88(1): 158-170.
- PORTES, A. (1998). "Social capital: Its origins and applications in modern sociology." Annual Review of Sociology 24: 1-24.
- PORTES, A. and P. LANDOLT (1996). "The downside of social capital." American Prospect 26: 18-21.
- POTTHOFF, R. F., K. G. MANTON and M. A. WOODBURY (1993). "Correcting for nonavailability bias in surveys by weighting based on number callbacks." Journal of the American Statistical Association 88: 1197-1207.
- PRENTICE, R. L. and L. A. GLOECKLER (1978). "Regression analysis of grouped survival data with application to breast cancer data." Biometrics 34: 57-67.
- PUTNAM, R. (1993). Making Democracy Work: Civic Traditions in Modern Italy. Princeton, Princeton University Press.
- PUTNAM, R. (1995). "Tuning in, tuning out: The strange disappearance of social capital in America." Political Science & Politics 28(4): 664-683.
- PUTNAM, R. (2000). Bowling Alone. New York, Simon & Schuster.

- QUADRINI, V. (1999). "The importance of entrepreneurship for wealth concentration and mobility." Review of Income and Wealth 45: 1-19.
- REES, H. and A. SHAH (1986). "An empirical analysis of self-employment in the UK." Journal of Applied Economics 1: 95-108.
- RIDDER, G. (1982). Life Cycle Patterns in Labor Market Experience: A Statistical Analysis of Labor Market Histories of Adult Men. Amsterdam, University of Amsterdam.
- ROED, K. and T. ZHANG (2002). "A note on the Weibull distribution and time aggregation bias." Applied Economic Letters 9: 469-472.
- ROSS, S. M. (1993). Introduction to Probability Models: Fifth Edition. San Diego, Academic Press.
- ROYSTON, P. and M. K. B. PARMAR (2002). "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects." Statistics in Medicine 21: 2175-2197.
- SANDERS, J. M. and V. NEE (1996). "Immigrant self-employment: The family social capital and the value of human capital." American Sociological Review 61: 231-249.
- SAY, J. B. (1803). A Treatise on the Political Economy or the Production, Distribution, and Consumption of Wealth. New York, E.M. Kelley Publishers.
- SCHOENFELD, D. (1982). "Partial residuals for the proportional hazards regression model." Biometrika 69: 239-241.
- SCHUMPETER, J. (1934). The Theory of Economic Development. Cambridge Mass., Harvard university press.
- SHUMWAY, T. (2001). "Forecasting bankruptcy more accurately: A simple hazards model." Journal of Business 74: 101-124.
- SINGER, J. D. and J. B. WILLETT (1993). "It's about time: Using discrete-time survival analysis to study duration and timing of events." Journal of Educational Statistics 18(155-195).
- STEIGUM, E. (1978). "A financial theory of investment behavior." Econometrica 51: 637-646.
- STIGLER, G. J. (1961). "The economics of information." Journal of Political Economy 69(3): 213-225.
- SUEYOSHI, G. T. (1995). "A class of binary response models for grouped duration data." Journal of Applied Econometrics 10: 411-431.
- TAYLOR, M. P. (1999). "Survival of the fittest? An analysis of self-employment duration in Britain." Economic Journal 109: C140-C155.
- THOMAS, J. M. (1996). "On the interpretation of covariate estimates in independent competing-risks models." Bulletin of Economic Research 48: 27-39.
- VAN DER BERG, G. J., M. LINDEBOOM and P. J. DOLTON (2005). Survey non-response and unemployment duration. Working paper.
- VAN PRAAG, C. M. (1996). Determinants of Successful Entrepreneurship. Amsterdam, University of Amsterdam.
- VAN PRAAG, C. M. (1999). "Some classic views on entrepreneurship." De Economist 147: 311-335.
- VAN PRAAG, C. M. (2001). "Business survival and succes of young small business owners." Small Business Economics 21: 1-17.
- VAN PRAAG, C. M. (2005). Successful Entrepreneurship: Confronting Economic Theory with Empirical Practice. Elgar, Cheltenham.
- VAN PRAAG, C. M. and J. S. CRAMER (2001). "The roots of entrepreneurship and labor demand: Individual ability and low risk aversion." Economica 68(269): 45-62.
- VAN PRAAG, C. M. and H. VAN OPHEM (1995). "Determinants of willingness and opportunity to start as an entrepreneur." Kyklos 48: 513-540.
- WANG, R., J. SEDRANSK and J. H. JINN (1992). "Secondary data analysis when there are missing observations." Journal of the American Statistical Association 87: 952-961.
- WEBSTER, F. A. (1977). "Entrepreneurs and ventures: An attempt at classification and clarification." Academy of Management Review 2: 54-61.
- WILLIAMSON, O. E. (1991). "Strategizing, economizing, and economic organization." Strategic Management Journal 12: 75-94.
- ZEIRA, J. (1987). "Investment as a process of search." Journal of Political Economy 95: 204-210.

Nederlandstalige samenvatting

Zelfstandig ondernemerschap is een belangrijke motor voor de Nederlandse economie. Zelfstandige ondernemers vullen de gaten in de markt die niet attractief genoeg zijn voor grote multinationals en zij vormen een belangrijke werkgever. Binnen deze grote groep zelfstandige ondernemers bevindt zich ook de zogenaamde “entrepreneur”. Het is voornamelijk deze entrepreneur die verantwoordelijk is voor innovatie en verandering in het economische systeem. Het is daarom niet verwonderlijk dat zelfstandig ondernemers een belangrijke onderzoekspopulatie vormt in wetenschappelijk onderzoek.

Wij mogen echter de zelfstandige ondernemer niet opvatten als een autonoom opererend individu buiten een brede sociale context. De zelfstandige ondernemer beïnvloedt niet alleen de economische en sociale context door het opstarten en opzeggen van bedrijven, het creëren en vernietigen van banen en het brengen van innovaties, hij is op zijn beurt afhankelijk en wordt gevormd door zijn sociale en economische context. De zelfstandige ondernemer komt vaak uit een familie van ondernemers, die hem belangrijke vaardigheden mee kan geven en die hem zelfs de mogelijkheid kan bieden om een familiebedrijf over te nemen. Echter ook een individu's geschiedenis kan een belangrijke rol spelen bij het worden van ondernemer. Zijn investeringen in educatie, zijn sparen op ontvangen loon, zijn huwelijk, en tevens ouderlijke erfenissen kunnen allemaal bijdragen om de droom van zelfstandig ondernemerschap te realiseren. Maar tegelijkertijd kan de accumulatie van arbeidsmarktspecifiek kapitaal, de zorg en verantwoordelijkheid voor vrouw en kinderen hem angstig en onwillig maken om te worden wie hij werkelijk is.

Wij geloven dat de zelfstandige ondernemer zo sterk ingebed is in zijn sociale context, zowel horizontaal door de ruimte van economische en sociale structuren als verticaal over de tijd door zijn gemaakte keuzes en toekomstige aspiraties, dat de beslissingen van ondernemers niet kunnen worden onderzocht vanuit een statisch atomistisch perspectief, maar moet worden bestudeerd als een dynamisch fenomeen die in onderlinge spanning en betrekking staat met een brede sociale context.

De centrale onderzoeksvragen in dit proefschrift proberen recht te doen aan deze visie. Niet alleen stellen wij de geaccepteerde onderzoeksvragen binnen de zelfstandig ondernemerschap literatuur naar wie er zelfstandig ondernemer wordt en wie hier op een later tijdstip weer vanaf ziet, maar wij vullen deze vragen aan met de onderzoeksvragen naar wanneer iemand zelfstandig ondernemer wordt en wanneer hij er mee ophoudt. Verder kijken wij hoe deze toe en uitredingsprocessen tot de markt van zelfstandige ondernemers zich gedragen in een brede sociale context.

In hoofdstuk 3 en 4 wordt de dynamiek van de beslissing om tot de markt van zelfstandige ondernemers toe te treden onderzocht. Naast de vraag wie er zelfstandig ondernemer wordt behandelen wij de vraag wanneer hij voor het eerst in zijn beroepsleven ondernemer wordt. Het verkrijgen van kennis betreffende de timing van zelfstandig ondernemerschap is belangrijk omdat zij inzicht kan geven in de snelheid waarmee een persoon kan toetreden tot de markt van zelfstandige ondernemers zodat hij zijn carrière als zelfstandig ondernemer kan maximaliseren. Ondernemerschap ervaring is namelijk een belangrijke indicator voor toekomstige ondernemerschapbedrijvigheid en toekomstig ondernemerschap succes in termen van overlevingskansen en bedrijfswinsten.

De onderzoeksvraag naar het tijdstip wanneer iemand ondernemer wordt en toetreedt tot de markt van zelfstandige ondernemers kan worden geschat met behulp van duurmodellen. Mede door het ontbreken van longitudinale data is dit op individueel niveau nog niet of nauwelijks gedaan in het wetenschappelijk onderzoek naar zelfstandige ondernemers en daarom wordt in hoofdstuk 3 uitgebreid ingegaan op de toepasbaarheid, bruikbaarheid en het voorspellende vermogen van duurmodellen op de onderzoeksvraag naar wanneer iemand voor het eerst tot de markt van zelfstandige ondernemers toetreedt. Hiervoor gebruiken wij de Noord Brabant dataset. Deze dataset bestaat uit een cohort individuen die drie keer geïnterviewd zijn over de periode 1953 tot 1993. In dit hoofdstuk laten wij de voordelen zien van dynamische modellen t.o.v. statische modellen aan de hand van een replicatie van een studie die deze data benaderde met een statische schatter. Eerst laten wij zien dat de statische schatter een geaggregeerde versie is van de dynamische schatter. De resultaten van de statistische analyse laten zien dat de dynamische schatters superieur zijn in termen van parameter precisie en voorspellend vermogen. Verder laat de data zien dat zelfstandig ondernemerschap leeftijdsafhankelijk is. Op jonge leeftijd is de kans om zelfstandig te worden klein waarna de kans langzaam toeneemt en op latere leeftijd weer afneemt.

In hoofdstuk 4 breiden we de kennis betreffende de timing van de toetredingsbeslissing uit door twee verschillende routes naar zelfstandig ondernemerschap te onderscheiden, namelijk, het overnemen van een familiebedrijf of het oprichten van een nieuw bedrijf. Hoewel vele onderzoekers het belang hiervan benadrukken is dit de eerste studie die zowel een economische theorie ontwikkeld die deze twee verschillende routes verklaart en die deze theorie ook empirisch kan testen. In deze theorie verklaart een individu's humaan, financieel, en sociaal kapitaal de tijd voordat een individu toetreedt tot de markt van zelfstandig ondernemers. Onze analyse brengt duidelijke verschillen aan het licht tussen mensen die een bestaand familie bedrijf overnemen en die een nieuw bedrijf oprichten. De resultaten van de analyse laten zien dat de kans op zelfstandig ondernemerschap leeftijdsafhankelijk is en dat de vorm van de tijdsafhankelijkheid per route naar zelfstandig ondernemerschap verschilt. Omdat het vak geleerd kan worden binnen een bestaand bedrijf en minder verzonken kosten gefinancierd hoeven te worden is de kans om een familiebedrijf over te nemen het grootst in de vroegste stadia van iemands professionele carrière. Daar tegenover vergt het opstarten van een nieuw bedrijf meer tijd. Eerst moet de potentiële zelfstandig ondernemer genoeg vaardigheden, capaciteiten, en kapitaal opbouwen om opstartkosten te overkomen en zelfstandig ondernemer te worden. De kans van een oprichter van een nieuw bedrijf om zelfstandig ondernemer te worden is dus initieel laag en neemt slechts gestaag toe over zijn leven wanneer meer kapitaal verzameld is.

Daarnaast hebben wij in hoofdstuk 4 ook de determinanten van de keuze voor zelfstandig ondernemerschap onderzocht. Onze resultaten laten zien dat naast een individu's humaan kapitaal, zijn sociale context een belangrijke indicator voor ondernemerschap is. De resultaten laten zien dat zijn sociale omgeving zowel de kans op ondernemerschap kan stimuleren als beperken. De statistische analyse van deze studie laat zien dat ouders die zelf ondernemer zijn niet alleen ondernemerschapspecifieke vaardigheden overgeven aan hun kinderen, maar vaak ook de mogelijkheid om het bedrijf over te nemen. Hoog opgeleide ouders, die ook welvarender zijn, spelen echter juist een belangrijke rol voor de oprichters van nieuwe bedrijven die meer liquiditeit nodig hebben om de

verzonken kosten van hun onderneming te financieren. Een echtgenote vergroot de kans op ondernemerschap omdat zij zowel emotionele als financiële ondersteuning kan geven en een goedkope arbeidskracht is. Aan de andere kant verkleinen kinderen de kans op ondernemerschap doordat zij tijd en energie opeisen. Tot slot laten de analyses zien dat de grootte van het gezin waarin men opgroeit de kans op het overnemen van een familiebedrijf significant verkleint.

De resultaten van hoofdstuk 4 hebben belangrijke implicaties voor het wetenschappelijk onderzoek naar zelfstandig ondernemerschap. Resultaten van voorgaande studies die het onderscheid niet kunnen maken tussen het oprichten van een nieuw bedrijf en het overnemen van een familie bedrijf dienen heroverwogen te worden in het licht van onze studie. Verder stellen wij voor dat de tijdsdimensie teruggebracht wordt in zowel theorievorming als mede in de econometrische schattingstechnieken naar de keuze voor zelfstandig ondernemerschap. De resultaten van hoofdstuk 4 laten namelijk zien dat het samenvoegen van zelfstandig ondernemen binnen het familie bedrijf en het opstarten van een nieuw bedrijf de dynamiek van ondernemerschap kan doen opheffen.

In hoofdstuk 5 introduceren wij de zogenaamde Limburg dataset. Deze dataset bestaat uit 1223 personen die serieus over zelfstandig ondernemen hebben nagedacht in 1998 en 1999. Deze dataset is in 2004 aangevuld met data van de Kamer van Koophandel waarin o.a. de gegevens zijn opgenomen over de status van het opgerichte bedrijf tot en met mei 2004. Hoofdstuk 5 beschrijft het verloop van het onderzoek die tot de Limburg dataset heeft geleid en presenteert een aantal beschrijvende statistieken. Veel aandacht in dit onderzoek is gericht op de non-response. Non-response is ook het thema in hoofdstuk 6. Daarnaast gebruiken wij deze data in hoofdstuk 7 en 8 om diverse theoretische voorspellingen empirisch te testen.

In hoofdstuk 6 onderzoeken we de oorzaken van non-response en specifiek de link tussen non-response en de beslissing om zelfstandig ondernemer te worden. Hiervoor bouwen wij een structureel model om de onderliggende populatie karakteristieken te benaderen. Verder bouwen wij een sequentieel beslissingsmodel om de link tussen non-response en de zelfstandigheidbeslissing te onderzoeken. De resultaten van deze modellen suggereren de volgende oorzaken voor non-response. Eén van de oorzaken van non-response is verloop binnen de onderzoekspopulatie door bijvoorbeeld ziekte of sterven. Verder zijn er individuele verschillen die de non-response kans beïnvloeden. Mensen met een lagere opleiding zijn bijvoorbeeld meer geneigd mee te werken aan het onderzoek. Echter de belangrijkste bevinding van deze analyses is dat er een structurele relatie bestaat tussen de kans dat iemand besluit om van zijn idee om een bedrijf te starten afziet en de kans dat hij weigert aan het onderzoek deel te nemen. Individuen die besluiten om hun bedrijfsidee voort te zetten zijn veel vaker geneigd om aan het onderzoek deel te nemen dan zij die van hun bedrijfsidee afzien. Hierdoor dienen onderzoekers en beleidsmakers opmerkzaam te zijn voor een zogenaamde causale bias. Deze zorgt er namelijk voor dat het afzien van een bedrijfsidee in werkelijkheid vaker voorkomt dan de onderzoeksresultaten laten zien. Dit kan weer leiden tot onderschatting van de structurele beperkingen voor het toetreden tot de markt van zelfstandige ondernemers, zoals het onderschatten van liquiditeitsbeperkingen.

In hoofdstuk 7 verplaatsen wij onze aandacht naar de dynamiek van de beslissing om te stoppen met een bedrijf en zo eventueel de markt van zelfstandige ondernemers weer te verlaten. In dit hoofdstuk ontwerpen wij een theoretisch model om de weerstand

van ondernemers om met hun bedrijf te stoppen te meten wanneer er verzonken kosten met het bedrijf gemoeid zijn en de evolutie van toekomstige bedrijfsinkomsten onzeker is. De resultaten laten zien dat de fractie van de opstartkosten die verzonken zijn er voor zorgen dat de ondernemer weerstand heeft om zijn bedrijf op te zeggen. Dit kan er op duiden dat er een optiewaarde aan het opzeggen van een bedrijf kleeft. Een ondernemer kan hierdoor er voor kiezen een tijdelijk verlies te accepteren en besluiten om zelfstandig ondernemer te blijven en wachten of de bedrijfscondities verbeteren om zo de gedeeltelijk verzonken opstartkosten terug te verdienen.

Verder gaat hoofdstuk 7 in op de vraag waarom startende ondernemers verschillen in de grootte van kapitaal investeringen waarmee zij de markt van zelfstandige ondernemers toetreden. Hier wordt wederom gekeken naar de invloed van de brede sociale context op deze beslissing en voornamelijk naar de invloed van eerdere keuzes m.b.t. de arbeidsmarkt op de grootte van de kapitaalinvestering van beginnende ondernemers. De resultaten laten zien dat in het bijzonder de hoogte van het laatst genoten loon een belangrijke determinant is voor de hoogte van de initiële kapitaalinvestering in gedeeltelijk verzonken opstartkosten. Dit resultaat is in lijn met de verwachtingen van een door ons met verzonken kosten uitgebreid keuzemodel voor ondernemerschap. Dit keuzemodel laat zien dat mensen hun initiële kapitaal investeringen moeten verhogen om individueel bepaalde toetredingsdrempels te overkomen. Dit model voorspelt dat individuen met een hoog laatst genoten loon meer gedeeltelijk verzonken kapitaal moeten investeren om hun persoonlijke drempel te overkomen en het voor hen winstgevend wordt om zelfstandig ondernemer te zijn. Specifiek vinden wij dat de elasticiteit van het voorafgaand genoten loon in verhouding tot initiële kapitaalinvesteringen in gedeeltelijke verzonken opstartkosten 0.294 is. Dit wil zeggen dat elke procent vermeerdering van het laatst genoten loon voor de toetredingsbeslissing een 0.294 procent toename in initiële kapitaalinvestering betekent.

In hoofdstuk 8 onderzoeken wij een mogelijke factor die kan bijdragen aan het succes van een onderneming, namelijk de grootte van het sociale netwerk van een ondernemer. In dit hoofdstuk modeleren wij de vraag naar productieve sociale relaties als een uitkomst van winstmaximaliserend gedrag onder onzekerheid. Wij veronderstellen dat sociale relaties productieve input zijn in het productieproces van een bedrijf. De productiviteit van een sociale relatie kan variëren tussen industrieën vanwege verschillende productieprocessen. Relaties kunnen ook productiever zijn wanneer markten niet perfect competitief zijn. Zulke markten worden vaak gekenmerkt door zoek en transactiekosten. Hierdoor kunnen op het oog gelijke sociale relaties verschillen in hun mate van productiviteit al naar gelang dat zij deze markt imperfecties kunnen overkomen en het productieproces kunnen optimaliseren. In ons model zijn de kosten om deze sociale relaties aan te gaan heterogeen. Evenwichtsvraag in de markt van sociale relaties wordt dan ook bereikt wanneer de marginale opbrengsten van een sociale relatie gelijk zijn aan de marginale kosten van het aangaan van deze relatie.

De voorspellingen van dit model zijn getoets op de Limburg dataset. De econometrische analyse laat zien dat jonge, hoog opgeleide startende ondernemers in grote steden met ouders, die ook zelfstandig ondernemer zijn, lagere kosten hebben om sociale relaties aan te gaan. De ordering van de verwachte winsten laten verminderende opbrengsten van sociale relaties zien en de verwachte winsten verschillen per industrie. Gemiddeld bereikt een zelfstandige ondernemer een optimale hoeveelheid informatie door 4 a 5 sociale

relaties aan te gaan. Hieruit kunnen wij concluderen dat de evenwichtsvraag naar productieve sociale relaties afgeleid kan worden uit de vraagcondities die ontstaan uit imperfecties binnen de marktstructuur en de aanbodscondities die ontstaan uit individuele verschillen tussen ondernemers.

Hiermee zijn wij aan het einde gekomen van dit proefschrift waarin wij getracht hebben recht te doen aan het dynamische karakter van toe en uittreding tot de markt van zelfstandige ondernemers. Verder hebben wij het belang laten zien van de sociale omgeving op deze processen. We hebben laten zien dat de ondernemer niet slechts een losstaand fenomeen buiten de sociale context is maar dat hij stevig ingebed zit in de sociale structuur die hem zowel mogelijkheden biedt om zichzelf te ontplooien maar die het hem tevens moeilijk maakt om een bedrijf te starten of om een bedrijf succesvol te maken.

Curriculum Vitae

Jasper Kok was born in Colmschate, the Netherlands, on January 17, 1977. He followed his high school education at the Emelwerda College in Emmeloord, which he completed in 1996. After finishing high school, he studied International Business at the University of Maastricht. He graduated in 2001 with a master thesis on the flexibility of strategic alliances in industry networks. After his thesis he worked for the local government in Amsterdam. In 2002 he started working on his Ph.D. thesis at the University of Maastricht, as a member of the department of Organization and Strategy. He completed the Meteor Ph.D. Research program in 2004. Since September 2006 he is a full time teacher at the University of Maastricht at the department of Organization and Strategy. In September 2006, he also started as a part-time student Psychology at the Academia of Integral Human Sciences in Utrecht, specializing in transpersonal counseling.