

Self-employment duration in urban and rural locations

Hannu Tervo and Mika Haapanen^{*}

School of Business and Economics, University of Jyväskylä

15 June 2005 (first draft)

Abstract

Previous research has shown that local environment is important for self-employment. Self-employment dynamics varies between areas characterized by different labour market conditions, entrepreneurial tradition and other structural factors. This paper analyses the self-employment spells in Finland with a large registered-based data set from the period of 1987-2002. The main aim is to investigate the role of region-specific factors as compared with individual-specific and other factors on the duration of self-employment spells. We find that the rates of exit out of self-employment and the length of self-employment spells depend upon personal characteristics and cyclical trends in the economy. The self-employment duration is modelled using discrete time survival analysis.

Keywords: Self-employment, duration, urban and rural labour markets

^{*} Contact information: School of Business and Economics, P.O. Box 35, FIN-40014 University of Jyväskylä, Finland. Emails: htervo@econ.jyu.fi and mphaapan@econ.jyu.fi. The paper is a part of a research project (number 200856) financed by the Academy of Finland.

1. Introduction

Individuals in small, dispersed labour markets may be pushed into self-employment if they see no other realistic employment options in the region (Carrasco and Ejrnaes 2003; Ritsilä and Tervo 2002). Typically, rural as compared to urban labour markets are characterized by weaker employment conditions. For example in Finland, rural labour markets hold worse rates of employment and self-sufficiency in jobs than urban labour markets. Self-employment is also on a high level in Finnish rural areas: the non-agricultural self-employment rate has been nearly twice as high in rural than urban locations (Tervo 2004). Besides that labour markets with weak paid-employment opportunities push workers to self-employment, self-employment dynamics also tends to be more common in these areas. As to Finland, individuals with alternation in their working careers are more likely to reside in rural locations than might be expected (Tervo 2004). Alternation between self-employment, paid-employment and non-employment is less common in urban areas which offer varied employment opportunities.

An interesting, but yet largely unstudied question concerns self-employment duration in different labour markets. While the decision to enter self-employment has now been widely studied, fewer studies have dealt with the duration of self-employment spells (see, however, Taylor 1999; Carrasco 1999; Johansson 2000; Falter 2001). It is fairly well known that many individuals who set up in self-employment do not remain self-employed for long. This paper focuses on self-employment duration in Finland. It is concerned with the individual running the firm, allowing many factors to influence the probability of survival.

The main aim is to analyse the role of region-specific factors as compared with individual-specific and other factors on the duration of self-employment spells. Earlier studies show that rates of exit out of self-employment and lengths of self-employment spells depend upon personal characteristics and cyclical trends in the economy. But what is the role of region-specific factors? Do self-employment spells differ in areas characterized by worse labour market situation (rural locations) compared with areas characterized by better situation (urban locations)? Our main finding is that the survival functions differ. The spells typically are longer in rural than urban areas, but this is due to lower hazard rates in rural areas in the first years of the self-employment spells. Obviously weak labour market situation forces individuals to continue as self-employed, even if the firm might not be very profitable. The

push effect works. Instead, the hazard rates are lower in urban areas if self-employment has lasted long enough (at least five years) and the firm has survived the difficult first years. This is the pull effect of increasing market. As long self-employment spells are rarer than short spells, the median spells become longer in rural areas as compared with urban areas.

The data used is based on a Longitudinal Census File and the Longitudinal Employment Statistics File constructed by Statistics Finland. Since 1987, these files are updated annually. These two register-based data sets, together with some other registers, provide panel data on each resident of Finland, from which a 7 percent random sample has been taken for this study. The study period is 1987-2002. The longitudinal data allow us to observe transitions into self-employment and the length of spells remaining in business.

The rest of the paper is organized as follows. First, theoretical standpoints and predictions about the effects of important variables on self-employment duration are described. Second, the data and methods are presented. Third, life table estimates and other descriptive results are presented. Fourth, self-employment duration is modelled using discrete time survival analysis. The paper ends with concluding remarks.

2. Theoretical framework

The theoretical background of this study rests on human capital theory and utility maximising paradigm (Knight 1921; Becker 1975; Evans and Leighton 1989; Tervo and Niittykangas 1994; Taylor 1999), and Jovanovic's (1982) dynamic selection model. Entrepreneurs have imperfect information about their innate abilities, which they can learn about only by trying entrepreneurship. Individuals less suited to self-employment give up quite soon. The longer an individual has been self-employed the more likely (s)he is to continue. The dynamic selection model emphasizes the importance of dynamics, with entrepreneurs learning about their abilities to compete and survive in the market post-entry (Taylor 1999; Parker 2004).

The utility maximizing paradigm predicts that individuals choose the occupation that offers the greatest expected utility. A rational self-employed compares the utility of entrepreneurship with the utility (s)he can obtain in wage-work or in other labour market state in each period, and makes the decision on the continuation according to this comparison. The choice of labour market state is continually revised as individuals' situations change. If the present value of the net expected benefits to continue as a self-employed is positive, the self-

employed does not leave the market, but if it is negative, (s)he quits. The continuation of an individual in self-employment is not necessarily equivalent to survival of a business because an individual can remain self-employed while opening and closing successive businesses (Parker 2004).

The utility maximizing approach should not be understood too literally. In a world of uncertainty and imperfect information, individuals will weigh up the pros and cons of the continuation of self-employment only intuitively. As many factors play a part in the determination of anticipated returns, these returns will be a subject to a constant process of adjustment. Individuals' perceptions of their returns may also suddenly alter in response to change in certain important factors.

The items comprising expected benefits and costs are wide in range, including both financial items and non-financial items. The main focus of this paper is in the analysis of the role of region-specific factors on surviving in self-employment. In addition to circumstances in the local labour market, however, it is assumed that the duration of self-employment, and thus survival, depends upon many other factors such as assets, personal and household characteristics, earlier activity and experience and industry. In general, many of these factors may be assumed to affect duration similarly as they affect the entry to self-employment.¹ The factors and hypothesis related to these factors are derived below, while the next section includes a more specific description of the data including the formation and the means of the variables used (see Table 1).

I Region-specific factors. Two sets of variables measure the situation in the local labour market. Our key variable defines whether the firm is located in rural or urban area. As rural areas are characterized by weaker employment conditions than urban areas, it is hypothesised that self-employment spells are there longer, especially during the first years of self-employment. Better survival rates in rural areas are assumed to follow more from the push than the pull effect of the market. In addition, rural areas have typically strong traditions in entrepreneurship, which may also have a positive effect on survival rates. Thus, the predicted effect is the same as it is in the entries into self-employment.

¹ Actually, many studies which analyse both self-employment entry and duration use the same variables in both models (e.g. Falter 2001).

A second set of region-specific variables describes unemployment situation in the local labour market. On the one hand, it is assumed that a high unemployment in the region forces self-employed to continue, since other employment options are scarce. On the other hand, the regional unemployment rate, and especially its change, is also an indicator of self-employment opportunities. Decreasing unemployment creates opportunities. Therefore, a rise in the unemployment can be positively related with exits from self-employment.

II Personal and household characteristics. Personal and household characteristics include gender, age, mother tongue, education, marital status and children. In general, these characteristics may help to describe individuals' entrepreneurial ability, the probability of receiving alternative job offers and tastes and preferences.

Earlier studies show that there is a positive relationship between business survival and an entrepreneur's human capital, although formal educational qualifications have shown mixed effects on survival rates (Taylor 1999; Kangasharju and Pekkala 2002; Parker 2004). This is because education increases both an individual's human capital and earning capacity in wage-work. Age also affects survival rates. The rates have been found to be higher for middle aged than for younger or older self-employed (Holtz-Eakin, Joulfaïn and Rosen 1994; Parker 2004).

Many studies have shown that self-employment status is positively associated with marital and parental status. Family support may be assumed to help for self-employed to survive longer. Gender predicts entries into self-employment - females are a minority of self-employed. It can also be assumed that the spells of female self-employed are shorter than the spells of male self-employed. Language may also predict self-employment duration. The great majority of the Finns speak Finnish as their mother tongue, but in certain regions, mainly along the southern and western coastline of the country, Swedish is the first language of the population. The institutional support for the Swedish-speaking minority has traditionally been strong (Liebkind, Broo and Finnäs 1995). The Swedish-speaking community with shared language, culture and social capital may also improve entrepreneurial success.

III Assets. It is assumed, other things equal, that a self-employed of some wealth survives longer than a self-employed without means. Two variables describe assets: income level and house ownership.

IV Prior activity and experience. Experience variables describe lifetime history and activity prior to the self-employment spell. Previous labour market status may have an effect on the

likelihood of remaining business. It can be assumed that a transition from wage-work to self-employment lays a better basis for survival than transitions from non-employment (being unemployed or a student). Individuals entering self-employment from wage-work may be expected to have higher human capital and motivation as well as higher quality information about business opportunities. For example, Carrasco (1999) showed that Spanish self-employed males who were unemployed prior to entering self-employment had hazard rates three times greater than those who previously worked in paid-employment.

Prior self-employment experience indicates the accumulation of business skills. It is an important explanatory variable in accounting for entry into self-employment (Evans and Leighton 1989; Parker 2004), and it can also be assumed to be an important factor for survival. Individuals who come from entrepreneur families can also be supposed to survive more likely in business because of the experience gained in the running of business in their early life as well as the familiarity with the market (Niittykangas and Tervo 2005). Parents' educational level may also predict survival in the business, although the size, and even the sign, of this effect is hard to predefine.

V Industrial sector. Industry at the start of the spell may predict survival, since business conditions vary between the sectors. It may also capture the changing industrial structure in the economy.

3. Data and methods

The data set

The data are based on various registers kept by Statistics Finland. Since 1970, Statistics Finland has compiled a population census every 5 years, and by 1990, the census had become entirely register based. By matching the unique personal identifiers across the censuses, Statistics Finland has constructed a Longitudinal Census File with panel data on the entire population of Finland at 5-year intervals from 1970 to 1995. In addition, since 1987, it has maintained the Longitudinal Employment Statistics file, which is updated annually. Since the same personal identifier is adopted in both the census and the longitudinal employment statistics, the two data sets can be merged, providing panel data on each resident of Finland for 1970, 1975, 1980, 1985 and then annually from 1987 to 2002. By using the personal

identifier, data from various other registers can still be merged with this panel data. In addition, data on spouses and parents can be merged under every individual.

We have in use a 7 percent random sample of those individuals who resided permanently in Finland in 2001. The data set includes thousands of variables from the Longitudinal Census File, Longitudinal Employment Statistics and other registers from the period 1970-2002. The individual level panel data was transformed and pooled into a sample of self-employment spells that began during the period 1988-2001. Hence, our analysis is based on the inflow samples from fourteen cohorts. Due to interval censoring, we only know individual's occupational status at the end of each year. Therefore, the beginning of the self-employment spell is defined in this study as follows:

An individual's self-employment spell begins in year j , if (s)he is not self-employed during the last week of year $j-1$ but (s)he is self-employed during the last week of year j , where $j = 1988-2001$.

The last year of observation is 2002. As consequence, the maximum of the observed duration of possibly right-censored self-employment spell is fourteen years. This is a longer period than in most other previous studies. Since an observed self-employment spell may also start after 1988, some of the spells are right-censored at shorter duration.

The sample of self-employment spells still needs to some additional restrictions:

1. Age-restriction. An individual's self-employment spell is included in the sample only if (s)he is aged between 18 and 55 years old at the beginning of the self-employment spell.
2. Industry-restriction. As usual, we excluded from our sample the self-employment spells in the primary sector. This is because the concept of self-employment is vaguer in agriculture than in other industries, and farm businesses have very different characteristics to non-farm businesses (Blanchflower 2000; Parker 2004). Accordingly, the analysis is concerned non-agricultural self-employment.
3. Region-restriction. Finally, those who live in the Åland Islands are not included in the sample. Åland is a small isolated region with only 26 000 inhabitants. It has a political autonomy and it differs from other Finnish regions in numerous ways. Due to its special character Åland is left out, as in many other studies.

The sample restricted in this way includes 15 182 self-employment spells. If self-employment spells in the primary production had also been included, the number of spells in the sample would have been 25 742. Note that we place no restrictions on the number of self-employment spells per person. Our sample includes self-employment spells from 12 661 different individuals, of whom 2 185 (17.3%) have two or more transitions into self-employment during the period 1988-2001.

Definition of self-employment

In this analysis, the concept of self-employment follows from the statistical definitions used by Statistics Finland (see Statistics Finland 2001). The central variable in this regard is “occupational status” which describes the position of the employed in the labour market: wage and salary earners, and entrepreneurs, which unfortunately cannot be divided into employers and self-employed (sole entrepreneurs). The category of entrepreneurs also comprises unpaid family workers. If an individual is not employed, (s)he belongs to a third category, non-employed. The non-employed are either unemployed or outside the labour force. The data on occupational status is based on the person’s national insurance status and wage, salary and/or entrepreneurial income received. For an entrepreneur, it is required that (s)he had a self-employed person’s pension insurance during the last week of the year and that his/her income from entrepreneurship exceeds his/her wage income, if the person is also in an employment relationship (for details, see Statistics Finland 2001).

Because self-employment is fundamentally a statistical concept used in labour market statistics and national accounting, great diversity can be found among those in this category (Johansson 2000). As is well-known, there exists a “grey area” between paid-employment and self-employment. For example, people may be self-employed by definition but share many of the same features of dependent paid employees. Examples of workers in the ‘grey area’ include salespersons on commission, freelancers, home-workers, tele-workers and unpaid family workers (Parker 2004).

Urban and rural areas

Statistics Finland uses a classification which divides Finnish municipalities (448 in 2000) into three categories: urban-type municipalities (68 in 2000), densely populated municipalities (75 in 2000) and rural-type municipalities (305 in 2000); see e.g. Statistics Finland 2001. This

classification is based both on the degree of urbanisation of the municipality and on the population of the largest urban settlement.

In this study, densely populated municipalities and rural-type municipalities are joined together. Accordingly, only two types of area are used which we call “urban areas” and “rural areas”. Urban areas are those municipalities in which at least 90% of the population lives in urban settlements or in which the population of the largest urban settlement is at least 15 000. An urban settlement refers to a cluster of buildings which are less than 200 metres apart from each other and which together house at least 200 people. Accordingly, rural areas are municipalities in which less than 90% of the population lives in urban settlements and in which the population of the largest urban settlement is at most 15 000. In our sample, the number of recorded self-employment spells in the urban areas is 9 363 (61.7%) and the number of recorded self-employment spells in the rural area is 5 819 (39.3%).

Duration model

The main emphasis of this study is to analyse the duration of self-employment spells and the factors influencing it. To do so, we need a measure for the probability of leaving self-employment in the next period, given that an individual has survived in business up to the current period.

If we could observe the exact length of self-employment spell i , then the continuous time hazard of the spell at duration time t could be parameterised using a proportional hazard specification

$$\theta_i(t) = \lambda(t) \cdot \exp[\beta'X_i(t)], \quad (1)$$

where $\lambda(t)$ is the baseline hazard at the time t , $X_i(t)$ is a vector of (time-varying) covariates, and β is the vector of unknown parameters to be estimated with Cox regression (see e.g. Wooldridge 2002, Ch. 20). However, the duration data available to us is interval censored (grouped) - the status of self-employment spell is only observed at the end of each year. Therefore, discrete time representation for the hazard rate is required.

Fortunately, a discrete time model can be specified, which is consistent with a continuous time model and interval-censored survival data (see e.g. Prentice and Gloeckler 1978; Sueyoshi 1995). Suppose T_i is the actual (unobserved) length of the self-employment spell.

Then the discrete (grouped) interval hazard rate, the probability of a spell being completed by time $t + 1$ given that it was still continuing at time t , can be defined as

$$h_i(t) = \text{prob}(T_i < t+1 | T_i \geq t) = 1 - \exp[-\exp(\beta'X_i(t) + \gamma(t))], \quad t = 1, 2, \dots \quad (2)$$

where $\gamma(t) = \ln\left[\int_t^{t+1} \lambda(u)du\right]$ summarises the pattern of duration dependence in the interval hazards (h_i), and the hazard rate is specified with complementary log-log distribution (Type I extreme value). In the following the discrete time duration model is estimated semi-parametrically by allowing the baseline hazard vary freely with duration time t (see e.g. Meyer 1990).² That is, the full set of γ 's are estimated by adding an indicator variable per duration time t to the model.³ The model can be further generalised by allowing the baseline hazard to differ between populations. In this study we are particularly interested whether it varies between rural and urban regions. Therefore, we allow the γ 's to be different for self-employed individuals in rural and urban areas.

If d_i is the observed duration of the i th self-employment spell (completed or censored) and c_i is an indicator variable equal to 1 if the spell is completed and 0 if it is censored, then the contribution of the i th spell to the log-likelihood is given by

$$\log L_i = c_i \left(\sum_{t=1}^{d_i-1} \ln[1 - h_i(t)] + \ln[h_i(d_i)] \right) + (1 - c_i) \left(\sum_{t=1}^{d_i} \ln[1 - h_i(t)] \right). \quad (3)$$

The log-likelihood for the whole sample of N spells is given by sum of all contributions i , which is maximised with respect to β and γ 's to provide maximum likelihood estimates.⁴ The empirical specification for the set of the covariates X is given below (Table 1).⁵

Variables

The five groups of factors (covariates) and the hypotheses related to them were depicted in Section 2. The following table compiles the covariates, their definitions and means (Table 1).

² That is, our discrete time duration model corresponds to the continuous time Cox proportional hazards model, but the estimates of the baseline hazard are derived directly as part of the estimation procedure.

³ If a constant term is included in the X -vector, one of the interval-specific dummy variables needs to be removed from the model specification.

⁴ Guidelines for practical implementation are given, for example, in Jenkins (1995).

⁵ Note that since the analysis is conditioned on a set of covariates X , then the duration is implicitly a function of the entire time path of the variable $X(t)$, $t = (0, t)$, which may have changed during the interval (Greene 2003).

Table 1. Descriptions for covariates and their means

Covariate	Description	Mean		
		Urban areas	Rural areas	All
<i>I Region-specific factors</i>				
Rural	1 if self-employed in rural areas, 0 if in urban areas	0	1	0.393
Unemlev [†]	Unemployment level (%) in the region	14.189	16.456	15.080
Unempcha [†]	Unemlev (t) / Unemlev ($t-1$)	1.022	1.005	1.015
<i>II Personal and household characteristics</i>				
Female	1 if female, 0 otherwise	0.388	0.369	0.380
Swedish	1 if native language is Swedish, 0 otherwise	0.046	0.061	0.052
Age [†]	Age in years	39.861	39.234	39.615
Age2 [†]	Age squared divided by 100	16.722	16.240	16.533
Marri [†]	1 if married or cohabiting, 0 otherwise	0.704	0.717	0.709
Childr [†]	1 if has children in the family, 0 otherwise	0.507	0.551	0.524
Interedu ^{†, a}	1 if secondary education (equivalent of 10-12 years of education), 0 otherwise	0.403	0.499	0.441
Highedu ^{†, a}	1 if tertiary education, 0 otherwise	0.351	0.236	0.306
<i>III Assets</i>				
Income [‡]	Annual income subj. to state taxation, 1 000 €	17.046	13.818	15.778
House [‡]	1 if owner-occupier of a house or a flat, 0 otherwise	0.708	0.783	0.737
<i>IV Prior activity and experience</i>				
Wagew [‡]	1 if in paid-employment, 0 otherwise	0.627	0.569	0.604
Unempl [‡]	1 if unemployed, 0 otherwise	0.163	0.201	0.178
Studen [‡]	1 if student or pupil, 0 otherwise	0.062	0.054	0.059
Migrat [‡]	1 if moved to another region, 0 otherwise	0.036	0.050	0.042
Entrexp ^{‡, b}	1 if self-employment experience, 0 otherwise.	0.373	0.527	0.433
Entrfath ^c	1 if father has been self-employed, 0 otherwise	0.228	0.400	0.296
Entrmoth ^c	1 if mother has been self-employed, 0 otherwise	0.229	0.393	0.294
Intedufa ^{a, d}	1 if father has secondary education, 0 otherwise	0.123	0.104	0.116
Intedumo ^{a, d}	1 if mother has secondary education, 0 otherwise	0.152	0.061	0.117
Higedufa ^{a, d}	1 if father has tertiary education, 0 otherwise	0.137	0.110	0.127
Higedumo ^{a, d}	1 if mother has tertiary education, 0 otherwise	0.093	0.042	0.073
<i>V Industrial sector (reference category manufacturing)</i>				
Trade [†]	1 if the industry is wholesale or retail trade, 0 otherwise	0.138	0.132	0.135
Hotel [†]	1 if the industry is hotel or restaurant, 0 otherwise	0.036	0.034	0.035
Building [†]	1 if the industry is construction, 0 otherwise	0.076	0.112	0.090
Transpor [†]	1 if the industry is transport, storage or communications, 0 otherwise	0.083	0.129	0.101
Businser [†]	1 if the industry is business services, 0 otherwise	0.103	0.063	0.087
Otherser [†]	1 if the industry is private services, 0 otherwise	0.089	0.077	0.084
Otherind [†]	1 if self-employed in some other industry, 0 otherwise	0.420	0.360	0.396

Notes: The number of self-employment spells is 15 182 and the number of spell-years (observations) is 57 601. Duration and year dummies are also used: $d_t = 1$, if duration time is t , 0 otherwise; $y_j = 1$, if year is j , 0 otherwise. [†] Time-varying covariate. [‡] Covariate is measured on year before the self-employment spell. ^a Reference category is basic education. ^b Using information from 1970, -75, -80, -85 and 1987-2001. ^c Using information from 1970, -75, -80 and -85. ^d Using information from 1970 and 1980.

First, some covariates such as sex or prior self-employment experience or parental covariates do not change in time. Another group of covariates such as the level of income/wealth or the prior activity is measured on year before the self-employment spell in order to analyse if the initial situation has an impact on the length of the self-employment spell. This assures that a covariate of this type is treated as a cause rather than a consequence of the self-employment spell. Third, a covariate describing the regional location of the company is measured at the outset of the self-employment spell. The reminder of the covariates, such are age, level of education, marital status, family size and regional unemployment, are treated as time-varying covariates. In addition to these covariates, duration dependence in the self-employment is explained with duration time dummies (d_t), one for each year (see notes in Table 1). Year dummies (y_j) are used to capture seasonal changes in the self-employment hazards.

4. Life table estimates

Before discussing the estimation results of the discrete time duration model, it will be useful to take a look at some descriptive statistics from our data on self-employment spells. Table 2 shows the number of started self-employment spells, proportions of self-employed surviving to next year and 2002 for each cohort in both areas (urban and rural). Firstly, yearly changes are notable in the number of started self-employment spells in Finland in 1988-2001. Initial downward trend in the flows into self-employment in 1988-1992 can be explained by the deep recession of the Finnish economy. Once the economy began to recover in 1993 flows into self-employment spells increased notably both in urban and rural areas. Since 1994, there has been a marginally declining trend in the number of started self-employment spells. Nevertheless, it is interesting to note that in recent years more individuals have still been entering self-employment than in the late 1980's or in the early 1990's.

Secondly, the proportions of self-employed surviving to next year are presented. We can see that 74.9% of the individuals who entered self-employment in urban areas in 1988-2001 are still self-employed in the following year. The survival rate is slightly higher in rural areas (75.8%). The survival rates have increased considerably over time. The survival rate is 66% for spells starting in 1988, where as it is more than 85% for spells starting in 2001. Note that these survival rates have increased over time more in rural areas than in urban areas.

Finally, the comparison of survival rates to 2002 shows, for example, that of the spells starting in urban areas in 1988 some 20.4% will survive to 2002. The survival rate for self-

employment spells starting in rural areas is lower (16%). Though, regional differences in the survival rates to 2002 vary with the year of origin.

Table 2. Survival rates in urban and rural areas

Starting in	Urban areas			Rural areas		
	Number of spells	Survive to next year, %	Survive to 2002, %	Number of spells	Survive to next year, %	Survive to 2002, %
1988	446	68.2	20.4	262	64.5	16.0
1989	542	60.7	17.7	266	68.4	21.8
1990	481	67.4	23.9	328	66.2	26.2
1991	396	75.0	29.0	216	74.5	25.0
1992	372	70.2	26.3	221	67.0	30.8
1993	972	75.6	30.1	741	74.8	31.6
1994	827	72.9	34.5	578	73.7	38.4
1995	812	73.9	40.8	520	71.7	38.3
1996	706	77.3	43.8	446	77.6	45.1
1997	800	76.1	46.3	448	79.0	50.7
1998	808	77.7	51.9	467	79.4	54.2
1999	765	79.6	60.3	447	82.8	65.8
2000	755	78.0	69.9	443	81.7	73.1
2001	681	85.2	85.2	436	86.0	86.0
All spells	9363	74.9	27.6	5819	75.8	26.0

Table 3 presents life table estimates of self-employment survival and hazard rates for urban and rural areas. The hazard rate is defined as a probability that a self-employed who has survived to the beginning of an interval will experience a terminal event in the interval of one year. We can see that the hazard rate in urban areas is 25% in the first year of self-employment spell, while it is around 24% in rural areas. These hazard rates decrease notably as spells get longer so that they are only 1-4% in the last interval. Interestingly, it seems that hazard rates are lower in rural areas compared with urban areas in the beginning of the spell, but clearly higher in later intervals. An estimate of the median survival time is, however, longer in rural than in urban areas. Half of the self-employed have exited self-employment after 5 years in rural areas, while in urban areas the estimated median survival time is one year shorter, 4 years. A comparison of surviving functions also indicates a statistical difference between the two areas. The observed significance level for the Log-rank test that

the survival distributions are the same for both areas is less than 0.001, leading us to reject the null hypothesis that the groups do not differ.⁶

Table 3. Life table estimates of self-employment: survival and hazard rates, %

Interval start time (years)	Urban areas			Rural areas		
	Number of observations	Survival rate, %	Hazard rate, %	Number of observations	Survival rate, %	Hazard rate, %
1	9363	74.9	25.1	5819	75.8	24.3
2	6434	61.6	17.8	4033	64.3	15.1
3	4763	53.3	13.4	3100	56.4	12.4
4	3663	47.9	10.3	2422	51.1	9.4
5	2867	43.5	9.1	1942	46.3	9.4
6	2237	40.2	7.7	1533	42.8	7.6
7	1755	37.5	6.6	1215	39.7	7.2
8	1309	34.9	7.1	929	36.7	7.4
9	931	32.8	5.9	638	34.8	5.2
10	583	31.3	4.5	371	32.5	6.7
11	459	29.6	5.7	278	30.2	7.2
12	318	28.5	3.5	204	28.5	5.4
13	192	27.9	2.1	107	27.2	4.7
14	92	27.6	1.1	44	26.0	4.6

Notes: Estimated median survival time is 4 years for urban areas and 5 years for rural areas. Log-rank test for the equality of the survivor functions: $\chi^2 = 14.3$ with d.f. = 1 ($p = 0.0002$).

5. The determinants of self-employment duration

Table 4 displays the results from the discrete time hazards models.⁷ A variable with a positive coefficient is associated with increased hazard rate and decreased survival time, while a variable with a negative coefficient is associated with decreased hazard rate and increased survival time.

Three model specifications are given. They differ in the specification of baseline hazard in terms of the rural variable. The first and simplest specification allows us to test whether hazard rates are universally at different levels in rural and urban areas regardless of the duration of the self-employment spell. The second specification is general in a sense that

⁶ The same conclusion can be reached with Kaplan-Meier survival analysis designed for continuous time duration analysis. All test results (Wilcoxon test and the Breslow test) suggest that we should reject the null hypothesis that the survival functions are the same for the two areas.

⁷ As the data include multiple self-employment spells experienced by one individual, the standard error are adjusted to allow for the same individuals being repeatedly at risk. Similarly, the hazard rates are allowed to differ according to the experience of previous self-employment spells.

baseline hazard is allowed to be completely different in rural and urban areas. Hence, it allows us to study the regional differences in the duration dependence. That is, do hazard rates differ from at each duration time? Finally, a more parsimonious model specification (3) is presented, which is used to test the hypotheses that

- (i) hazard rates are significantly lower in rural areas than in urban areas in the beginning of the self-employment spell, and
- (ii) hazard rates are significantly lower in urban areas than in rural areas in the latter part of the self-employment spell.

Before presenting our conclusions in terms of these hypotheses and the region-specific variables, we briefly discuss other results that turn out to be very robust to the model specifications. Of the personal characteristics, gender and age behave as expected. Men have a higher probability of remaining in business than women. Young and old self-employed show a higher propensity to leave self-employment than middle-aged self-employed. The estimates suggest that the exit rates are lowest when a self-employed is at his/her forties. The results on the effect of education support the hypothesis that higher education increases self-employment duration. Interestingly, the language variable is not significant. Contrary to our expectations, the Swedish-speaking minority does not seem to have a higher probability of remaining in business than the Finnish-speaking majority. Marriage increases self-employment duration, while having children does not seem to have an impact. These results at least partly strengthen the significance of family support.

Initial wealth seems to have an impact on self-employment duration: both house ownership and income level prior to the self-employment spell lower the exit rates and thus increase self-employment duration. Differences in exit rates by prior activity and experience also merge. The estimated effect of previous labour market situation shows that individuals who were in wage-work prior to the self-employment spell are more successful in remaining self-employed than others. Individuals who were students prior to entering self-employment have somewhat greater hazard rates than others. Contrary to some previous results (e.g. Carrasco 1999), prior unemployment experience does not have a statistically significant effect on the hazard rates of leaving self-employment. Migration experience before entering self-employment increases hazard rates as was expected. Industry also plays an important role in self-employment duration. For example, individuals being self-employed in transport, storage or communications, construction, and private services have much slower exit rates than individuals in manufacturing.

Table 4. Parameter estimates of the discrete time duration models

Covariate	Model 1		Model 2		Model 3	
	Coeff.	(Std.err.)	Coeff.	(Std.err.)	Coeff.	(Std.err.)
<i>I Region-specific factors</i>						
Rural	-0.040	(0.025)				
Rural $\times \sum_1^4 d_t$					-0.070**	(0.027)
Rural $\times \sum_5^{14} d_t$					0.121*	(0.058)
Unemplev	0.002	(0.003)	0.002	(0.003)	0.002	(0.003)
Unempcha	0.113*	(0.054)	0.108*	(0.054)	0.108*	(0.054)
<i>II Personal and household characteristics</i>						
Female	0.127**	(0.025)	0.128**	(0.025)	0.128**	(0.025)
Swedish	-0.024	(0.053)	-0.021	(0.053)	-0.022	(0.053)
Age	-0.074**	(0.010)	-0.075**	(0.010)	-0.075**	(0.010)
Age2	0.088**	(0.012)	0.089**	(0.012)	0.089**	(0.012)
Marri	-0.298**	(0.036)	-0.300**	(0.036)	-0.300**	(0.036)
Childr	0.061	(0.032)	0.062	(0.032)	0.062	(0.032)
Interedu	-0.080**	(0.028)	-0.081**	(0.028)	-0.081**	(0.028)
Highedu	-0.111**	(0.033)	-0.110**	(0.033)	-0.110**	(0.033)
<i>III Assets</i>						
Income	-0.006**	(0.001)	-0.006**	(0.001)	-0.006**	(0.001)
House	-0.196**	(0.025)	-0.194**	(0.025)	-0.195**	(0.025)
<i>IV Prior activity and experience</i>						
Wagew	-0.326**	(0.032)	-0.326**	(0.032)	-0.326**	(0.032)
Unempl	0.040	(0.036)	0.039	(0.036)	0.039	(0.036)
Studen	0.122**	(0.045)	0.121**	(0.045)	0.122**	(0.045)
Migrat	0.205**	(0.049)	0.206**	(0.049)	0.205**	(0.049)
Entrexpe	0.166**	(0.026)	0.166**	(0.026)	0.166**	(0.026)
Entrfath	-0.156**	(0.035)	-0.156**	(0.035)	-0.156**	(0.035)
Entrmoth	-0.028	(0.034)	-0.027	(0.034)	-0.028	(0.034)
Intedufa	0.012	(0.036)	0.013	(0.036)	0.013	(0.036)
Intedumo	-0.009	(0.038)	-0.010	(0.038)	-0.009	(0.038)
Higedufa	-0.020	(0.034)	-0.019	(0.034)	-0.020	(0.034)
Higedumo	0.080	(0.044)	0.080	(0.044)	0.079	(0.044)
<i>V Industrial sector (reference category manufacturing)</i>						
Trade	0.021	(0.057)	0.021	(0.057)	0.022	(0.057)
Hotel	0.198**	(0.071)	0.198**	(0.071)	0.199**	(0.071)
Building	-0.173**	(0.065)	-0.174**	(0.065)	-0.174**	(0.065)
Transpor	-0.593**	(0.077)	-0.604**	(0.077)	-0.600**	(0.077)
Businser	0.063	(0.067)	0.063	(0.067)	0.064	(0.067)
Otherser	-0.362**	(0.070)	-0.363**	(0.070)	-0.362**	(0.070)
Otherind	0.589**	(0.051)	0.588**	(0.051)	0.589**	(0.051)

Table 4. (Continued)

Covariate	Model 1		Model 2		Model 3	
	Coeff.	(Std.err.)	Coeff.	(Std.err.)	Coeff.	(Std.err.)
<i>VI Duration dummies</i>						
d_1	0.395	(0.202)	0.401*	(0.203)	0.415*	(0.203)
d_2	0.065	(0.206)	0.122	(0.207)	0.086	(0.206)
d_3	-0.113	(0.207)	-0.091	(0.209)	-0.092	(0.208)
d_4	-0.309	(0.209)	-0.290	(0.211)	-0.288	(0.209)
d_5	-0.341	(0.211)	-0.378	(0.214)	-0.396	(0.212)
d_6	-0.463*	(0.213)	-0.482*	(0.218)	-0.518*	(0.214)
d_7	-0.553*	(0.217)	-0.615**	(0.225)	-0.608**	(0.218)
d_8	-0.413	(0.219)	-0.464*	(0.230)	-0.468*	(0.220)
d_9	-0.582*	(0.232)	-0.569*	(0.245)	-0.635**	(0.233)
d_{10}	-0.630*	(0.248)	-0.860**	(0.283)	-0.678**	(0.248)
d_{11}	-0.420	(0.255)	-0.556	(0.288)	-0.467	(0.256)
d_{12}	-0.777**	(0.293)	-1.021**	(0.362)	-0.827**	(0.293)
d_{13}	-1.054**	(0.398)	-1.445**	(0.544)	-1.100**	(0.398)
d_{14}	-1.197	(0.611)	-1.903	(1.019)	-1.241*	(0.611)
Rural $\times d_1$			-0.027	(0.035)		
Rural $\times d_2$			-0.169**	(0.051)		
Rural $\times d_3$			-0.069	(0.065)		
Rural $\times d_4$			-0.059	(0.085)		
Rural $\times d_5$			0.083	(0.098)		
Rural $\times d_6$			0.037	(0.120)		
Rural $\times d_7$			0.140	(0.143)		
Rural $\times d_8$			0.114	(0.159)		
Rural $\times d_9$			-0.042	(0.222)		
Rural $\times d_{10}$			0.539	(0.281)		
Rural $\times d_{11}$			0.345	(0.301)		
Rural $\times d_{12}$			0.562	(0.428)		
Rural $\times d_{13}$			0.871	(0.673)		
Rural $\times d_{14}$			1.423	(1.229)		
<i>VII Year dummies (reference category 1989)</i>						
y_{1990}	0.076	(0.086)	0.078	(0.086)	0.078	(0.086)
y_{1991}	-0.195	(0.121)	-0.189	(0.121)	-0.188	(0.121)
y_{1992}	-0.396**	(0.102)	-0.393**	(0.102)	-0.394**	(0.102)
y_{1993}	-0.253*	(0.100)	-0.253*	(0.100)	-0.251*	(0.100)
y_{1994}	-0.020	(0.091)	-0.022	(0.091)	-0.019	(0.091)
y_{1995}	0.125	(0.089)	0.127	(0.089)	0.127	(0.089)
y_{1996}	0.099	(0.088)	0.101	(0.088)	0.101	(0.088)
y_{1997}	-0.110	(0.086)	-0.110	(0.086)	-0.108	(0.086)
y_{1998}	-0.149	(0.084)	-0.150	(0.084)	-0.149	(0.084)
y_{1999}	-0.189*	(0.083)	-0.190*	(0.083)	-0.190*	(0.083)
y_{2000}	-0.168*	(0.081)	-0.170*	(0.081)	-0.168*	(0.081)
y_{2001}	-0.214**	(0.082)	-0.215**	(0.082)	-0.214**	(0.082)
y_{2002}	-0.773**	(0.088)	-0.774**	(0.088)	-0.773**	(0.088)
Log-likelihood	-21705.05		-21693.73		-21700.24	

Notes: Standard errors are adjusted for clustering at the individual level. Number of observations: 57 601. Variable definitions are given in Table 1. * (**)= Statistically significant at the 0.05 (0.01) level.

What is surprising is the result that prior self-employment experience has a significant negative effect on self-employment duration. For example, Taylor (1999) found previous business experience to be an important determinant of survival in Britain. On the other hand, many individuals have alternation between self-employment and other labour markets states in Finland (Tervo 2004), and for these individuals previous self-employment spells do not inevitably predict a high probability of remaining self-employed. However, if father has been self-employed individual's self-employment duration increases. Although the estimated effect of mother's self-employment does not reach significance, these results suggest the importance of early life experience. Parent's educational level does not, however, affect survival.

Turning now to our main question, the effect of environment on self-employment duration, the results of the first model specification show that the site (urban vs. rural) does not seem to have a universal effect on exit rates. As regards local unemployment, we do not find evidence that a high unemployment in the region forces self-employed individuals to continue. Instead, we find that a rise in the local unemployment rate increases exits from self-employment.

The preceding descriptive analysis based on follow-up life tables showed that hazard rates might be lower in rural areas compared with urban areas in the beginning of the spell, but higher in later intervals. The second specification allows us to study whether this still holds after other factors have been controlled for. The estimation results are in line with the descriptive analysis. Figure 1 shows predicted hazard rates for a typical self-employed (median observation).⁸ We can see that in urban areas the hazard rates continue to drop with years self-employed, whereas in rural areas the hazard rates hardly decrease after self-employment has lasted for five years.⁹

⁸ Predicted hazard rates have been computed using model 2 estimates in Table 4 and sample median values for covariates. That is, regional unemployment rate is 14.9%, unemployment rate has decreased 5% and year is 1998. The self-employed is 40-years-old married male with children, whose native language is not Swedish. He has house and secondary education and his annual income is 13 000 €. Before self-employment spell he was in paid-employment, did not move or were previously self-employed. His father or mother has not been self-employed, and the parents have only basic education. He is not self-employed in manufacturing, trade, hotel, building, transport or services.

⁹ The negative estimated coefficients of covariates $rural \times d_j$ ($j = 1, 2, 3, 4$) imply that exit rates are lower in rural areas compared with urban areas during the first four self-employment years. The reverse is true in the later part of the spell ($t \geq 5$).

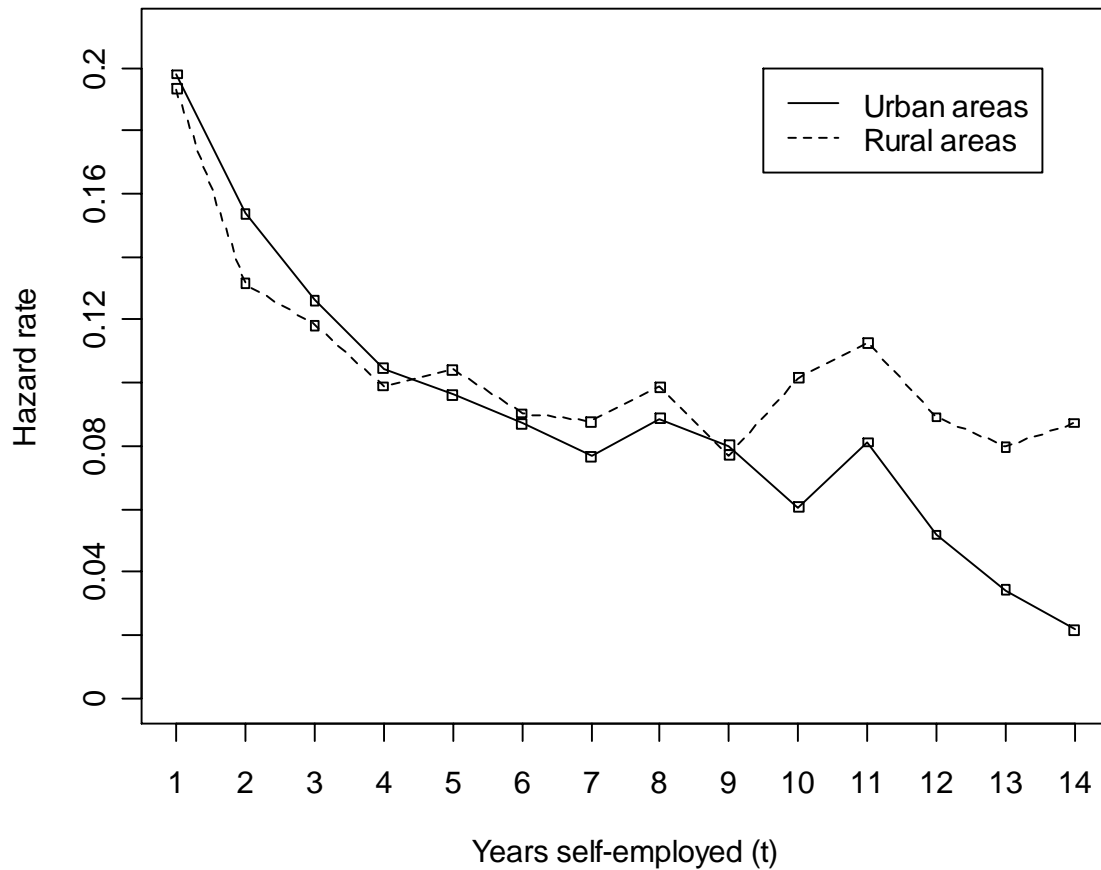


Figure 1. Predicted hazard rates in urban and rural areas (Model 2 in Table 4)

Finally, the third specification gives further support for our findings. The estimated coefficient on the covariate $\text{rural} \times \sum_1^4 d_t$ is significantly negative and the estimated coefficient on the covariate $\text{rural} \times \sum_5^{14} d_t$ is significantly positive. This implies that exit rates are significantly lower in rural areas than in urban areas when a self-employment spell has lasted less than five years. Reverse is true when a self-employment spell has lasted five years.

6. Conclusion

This paper showed that self-employment duration depends on many factors. Environmental factors and circumstances in the local labour market have also resonance, but this effect is not straightforward. The survival functions between rural and urban areas differ. On average, median survival times are longer in rural than urban locations. Rural areas as compared to urban areas in Finland are characterized by weaker employment conditions. Self-employment is also on a higher level in rural areas. Our estimation results showed that, after controlling for other factors, the site does not have a universal effect on self-employment duration, but after separating the lengths of the spells significant effects were found. Rural areas have lower hazard rates in the first years of the spells, while urban areas have lower rates in the spells lasting at least five years. Long-lasting self-employment spells are, however, much rarer than short spells.

These results would suggest that the push effect is strong in rural locations: besides that labour markets with weak paid-employment opportunities push workers into self-employment they force individuals to continue as self-employed, even if the firm might not be very profitable. If the firms survive the difficult first years, hazard rates decrease notably. The decrease is greater in urban locations compared with rural locations, which tells about the greater pull effect of increasing market.

References

- Becker G. (1975) *Human Capital*. Columbia University Press, New York.
- Blanchflower D.G. (2000) Self-Employment in OECD Countries. *Labor Economics* 7:471-505.
- Carrasco R. (1999) Transitions to and from Self-Employment in Spain: An Empirical Analysis. *Oxford Bulletin of Economics and Statistics* 61: 315-341.
- Carrasco R and Ejrnæs M. (2003) Self-Employment in Denmark and Spain: Institution, Economic Conditions and Gender Differences. Centre for Applied Microeconometrics, Institute of Economics, University of Copenhagen.
- Evans D. S. and Leighton L. S. (1989) Some Empirical Aspects of Self-Employment. *The American Economic Review* 79(3): 519-535.
- Falter J-M. (2001) Self-Employment Entry and Duration in Switzerland. Papers from the 13th EALE Annual Conference 13-16 September 2001, Jyväskylä, Finland. CD-Rom.
- Greene W. (2003) *Econometric Analysis*, 5th edition. Prentice Hall, Englewood Cliffs, New Jersey.
- Holtz-Eakin D., Joulfaïn D. and Rosen H. S. (1994) Sticking it Out: Entrepreneurial Survival and Liquidity Constraints. *Journal of Political Economy* 102: 53-75.
- Jenkins S. P. (1995) Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics*, 57(1): 129-138.
- Johansson E. (2000) *Essays on the Determinants of Self-Employment*. Publications of the Swedish School of Economics and Business Administration, Nr 85, Helsingfors.
- Jovanovic B. (1982) Selection and the Evolution of Industry. *Econometrica* 50: 649-670.
- Kangasharju A, and Pekkala S. (2002) The Role of Education in Self-Employment Success in Finland. *Growth and Change* 33 (2): 216-237.
- Knight F. (1921) *Risk, Uncertainty, and Profit*. Houghton Mifflin, New York.
- Liebkind K., Broo R. and Finnäs F. (1995) The Swedish-Speaking Minority in Finland. In *Cultural Minorities in Finland. An Overview Towards Cultural Polity*. Publications 66, Finnish National Commission for Unesco.
- Meyer B. D. (1990) Unemployment Insurance and Unemployment Spells. *Econometrica*, 58(4): 757-782.
- Niittykangas H. and Tervo H. (2005) Spatial Variations in Intergenerational Transmission of Self-Employment. *Regional Studies* 39: 319-332.

- Parker S. C. (2004) *The Economics of Self-Employment and Entrepreneurship*. Cambridge University Press, Cambridge.
- Prentice R. and Gloeckler, L. (1978) Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34: 57-67.
- Ritsilä J. and Tervo H. (2002) Effects of Unemployment on New Firm Formation: Micro-Level Panel Data Evidence from Finland. *Small Business Economics* 14: 31-40.
- Statistics Finland (2001) Population Census 2000 Handbook, Helsinki.
- Sueyoshi, G. T. (1995) A Class of Binary Response Models for Grouped Duration Data. *Journal of Applied Econometrics*, 10(4): 411-431.
- Taylor M. (1999) Survival of the Fittest? An Analysis of Self-Employment Duration in Britain. *The Economic Journal* 109: C140-C155.
- Tervo H. (2004) Self-Employment Dynamics in Rural and Urban Labour Markets. Papers from the 44th ERSA Congress, 27-30 August 2004, Porto, Portugal. CD-rom.
- Tervo H. and Niittykangas H. (1994) The Impact of Unemployment on New Firm Formation in Finland. *International Small Business Journal* 13(1): 38-53.
- Wooldridge J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT, London.