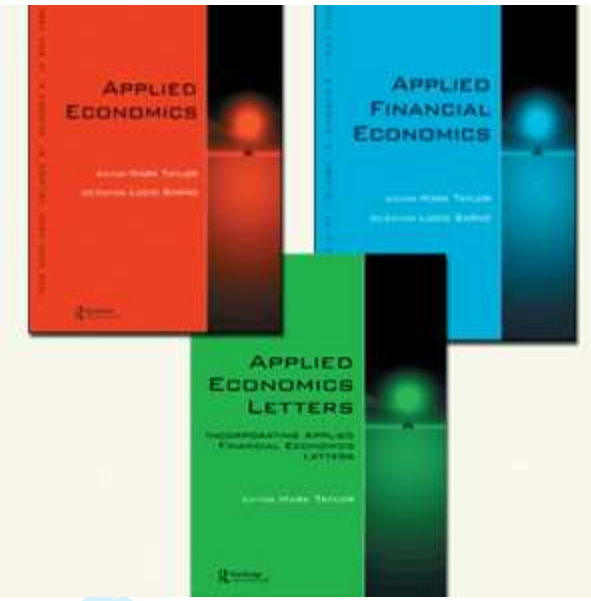
Self-employment duration in urban and rural locations

Haapanen, Mika; Tervo, Hannu

Postprint / Postprint
Zeitschriftenartikel / journal article



Self-employment duration in urban and rural locations

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0018
Journal Selection:	Applied Economics
Date Submitted by the Author:	11-Jan-2007
Complete List of Authors:	Haapanen, Mika; University of Jyväskylä, School of Business and Economics Tervo, Hannu; University of Jyväskylä, School of Business and Economics
JEL Code:	J23 - Employment Determination; Job Creation; Labor Demand; Self-Employment < J2 - Time Allocation, Work Behavior, and Employment Determination/Creation < J - Labor and Demographic Economics, C41 - Duration Analysis < C4 - Econometric and Statistical Methods: Special Topics < C - Mathematical and Quantitative Methods, R23 - Regional Migration|Regional Labor Markets|Population < R2 - Household Analysis < R - Urban, Rural, and Regional Economics
Keywords:	self-employment, duration, urban and rural labour markets

Self-employment duration in urban and rural locations

Mika Haapanen and Hannu Tervo
School of Business and Economics, University of Jyväskylä
P.O. Box 35
FIN-40014 University of Jyväskylä, Finland

Abstract. Previous research has shown that the local environment is important for self-employment. The dynamics of self-employment varies between areas characterized by different labour market conditions, entrepreneurial traditions and other structural factors. This paper analyzes self-employment spells in Finland with a large register-based data set from the period 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. First, the descriptive analysis shows that the exit rates from self-employment and the length of self-employment spells depend upon location (urban vs. rural area) and the cyclical trends in the economy. Second, self-employment duration is modelled using discrete time survival analysis. It is found that rural areas have significantly lower exit rates in the first years of self-employment than urban areas.

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

JEL Classification numbers: J23, C41, R23

Corresponding author:
Professor Hannu Tervo
School of Business and Economics, University of Jyväskylä
P.O. Box 35
FIN-40014 University of Jyväskylä, Finland
e-mail: htervo@econ.jyu.fi

Self-employment duration in urban and rural locations

I. Introduction

Individuals in small, dispersed labour markets may be pushed into self-employment if they see no other realistic employment options in the region (Ritsilä and Tervo, 2002; Moore and Mueller, 2002; Carrasco and Ejrnaes, 2003, Tervo, 2006). 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. Weak paid-employment opportunities push workers into self-employment in Finnish rural labour markets (Tervo, 2004). The non-agricultural self-employment rate has been nearly twice as high in rural than urban locations. Individuals in Finnish rural locations have also more alternation between self-employment, paid-employment and non-employment in their working careers than individuals in urban locations, where employment opportunities are more varied.

While the decision to enter self-employment has been widely studied, fewer studies have dealt with the duration of self-employment spells (see, however, Taylor, 1999; Carrasco, 1999; Johansson, 2000; Falter, 2001; Cueto and Mato, 2006). An interesting, but yet largely unstudied question concerns self-employment duration in different labour markets. 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 played by region-specific factors?

The purpose of this paper is to analyze the role of region-specific factors on the duration of self-employment spells in Finland. As rural areas are characterized by weaker conditions of employment than urban areas, it is hypothesized that exit rates from self-employment are lower in rural than urban areas. The data used are based on a Longitudinal

Census File and the Longitudinal Employment Statistics File constructed by Statistics Finland. Since 1987, these files have been annually updated. 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 longitudinal data allow us to observe transitions into self-employment and how long self-employed individuals remain in business in 1987–2002. The individual level panel data were pooled into a sample of self-employment spells that began during the period 1988–2001. The sample consists of 19,439 spells from 16,867 different individuals. The maximum observed duration of a possibly right-censored self-employment spell is fourteen years. This is a longer period than in most previous studies.

Our analysis affirms the importance of region-specific factors in the determinants of self-employment duration in Finland. First, the descriptive analysis shows that the exit rates from self-employment and the length of self-employment spells depend upon location (urban vs. rural area) and the cyclical trends in the economy. Second, self-employment duration is modelled using discrete time survival analysis. A flexible, semiparametric specification of the baseline hazard is adopted, as in Meyer (1990). The analysis is further generalized by allowing the baseline hazard to differ between the two locations. Our main finding is that, after controlling for other factors, the estimated survival functions differ. Self-employment spells typically are longer in rural than urban areas. This is due to the lower hazard rates in rural areas in the first years of self-employment. It seems that a weak labour market situation forces individuals to continue in self-employment, even though its profitability is questionable. The push effect operates.

The rest of the paper is organized as follows. Section II discusses the theoretical framework for the analysis of self-employment duration. The data and methods are described in Section III. Then life-table estimates and other descriptive results are presented (Section

IV). Self-employment duration is modelled using discrete time survival analysis and results are discussed in Section V. Finally, concluding remarks bring the paper to an end.

II. Theoretical framework

The theoretical background of this study rests on the human capital theory and utility maximising paradigm (Knight, 1921; Becker, 1975; Evans and Leighton, 1989; Tervo and Niittykangas, 1994; Taylor, 1999; Uusitalo, 2001), 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 person compares the utility of entrepreneurship with that of paid employment and other labour market states in each period, and decides accordingly whether or not to continue in her/his current status. The choice of labour market status is continually subject to revision as individuals' situations change. If the present value of the expected net benefits of continuing as self-employed is positive, the individual remains in self-employment, but if it is negative, (s)he quits. Continuation of self-employment is not necessarily equivalent to the survival of a business since an individual can remain self-employed while starting up and closing down 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 continuing in self-employment only intuitively. As many factors play a part in the

determination of anticipated returns, these returns will be 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 range widely, and include both financial and non-financial items. The main focus of this paper is the role of region-specific factors on surviving in self-employment. In addition to factors relating to 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 type of industry. In general, many of these factors can be assumed to affect duration in the same way as they affect entry into self-employment.[1] The factors and hypothesis related to these factors are derived below, while the following section includes a more specific description of the data, including the formation and means of the variables used (see Table 1 below).

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 a rural or an urban area. As rural areas are characterized by weaker conditions of employment than urban areas, it is hypothesized that exit rates from self-employment will be lower, 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 typically have a strong tradition of entrepreneurship, which may also have a positive effect on survival rates. Thus, the predicted effect on survival is the same as it is for entry into self-employment.

A second set of region-specific variables describes the unemployment situation in the local labour market. On the one hand, it is assumed that high unemployment in the region forces self-employed individuals to continue as they are, since other employment options are

scarce. On the other hand, the regional unemployment rate, and especially change in it, is also an indicator of opportunities for self-employment. A decrease in unemployment creates opportunities, and vice versa. Therefore, a rise in unemployment is expected to be positively related to 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 her/his earning capacity in paid employment. Age also affects survival rates, which have been found to be higher for middle-aged than for younger or older self-employed individuals (Holtz-Eakin, Joulfain and Rosen, 1994; Parker, 2004).

Many studies have shown a positive association between self-employment status and marital and parental status. Family support may be assumed to help self-employed individuals to survive longer. Gender also predicts entries into self-employment – females are in their minority among the self-employed. It can also be assumed that spells of female self-employment are shorter than those of male self-employment. Language is another possible predictor of self-employment duration. The overwhelming majority (app. 92%) of the Finns speak Finnish as their mother tongue, but in certain regions, mainly along the southern and western coastlines, Swedish is the first language of a significant part of the population. Institutional support for the Swedish-speaking minority has traditionally been strong

[1] Many studies which analyse both self-employment entry and duration in fact use the same variables

(Liebkind, Broo and Finnäs, 1995). The Swedish-speaking community with its shared language, culture and social capital may also improve entrepreneurial success.

III Assets. Entrepreneurial activity is restricted by liquidity constraints (Evans and Jovanovic, 1989). Lack of financial resources may not only prevent entry into self-employment but also successful running of business (see e.g. Holtz-Eakin, Joulfain and Rosen, 1994). Therefore it is expected, other things being equal, that a wealthier self-employed individual will survive longer than her/his less wealthy counterpart. In the current study three variables describe assets: personal income, income of spouse, and house ownership.

IV Prior activity and experience. The experience variables describe lifetime labour market history and activity prior to the self-employment spell. Previous labour market status may have an effect on the likelihood of remaining in business. It can be assumed that the transition from paid employment to self-employment lays a better foundation for survival than the same transition from non-employment (being unemployed or student). Individuals entering self-employment from paid employment 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. In addition, prior experience of migration is likely to increase the exit rates from self-employment. Starting business in a new environment may involve unexpected risks, for example, in finding customers.

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

in both models (e.g. Falter, 2001).

survival. Individuals who come from entrepreneurial families can also be assumed to have a greater chance of surviving in business because of their experience of running a business gained in their early life as well as their familiarity with the market (Niittykangas and Tervo, 2005).

V Industrial sector. The industry of a self-employed person may further predict survival, as business conditions vary between economic sectors (see e.g. Taylor 1999). It may also capture the changing industrial structure in the economy.

III. 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 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 were transformed and pooled into a sample

of self-employment spells that began during the period 1988–2001. Due to interval censoring, we only know an individual's occupational status at the end of each year. Therefore, the beginning of a 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 is self-employed during the last week of year *j*, where *j* = 1988–2001.

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

Some additional restrictions on the sample of self-employment spells remain to be imposed:

1. Age-restriction. An individual's self-employment spell is included in the sample only if (s)he is aged between 18 and 60 years old at the beginning of that spell.
2. Industry-restriction. As is usual, we excluded from our sample 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 from non-farm businesses (Blanchflower, 2000; Parker, 2004). Accordingly, the analysis concerns 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 political autonomy and it differs from the other Finnish regions in numerous ways. Due to its special character Åland is left out, as in many other studies.

Restricted in this way the sample consists of 19,439 self-employment spells. If spells in primary production had also been included, the number would have been 26,668. Note that we place no restrictions on the number of self-employment spells per person. Our sample consists of all the spells of 16,867 different individuals, 2,264 (13.4%) of whom 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 (students etc.). 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 her/his income from entrepreneurship exceeds her/his 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, people working at home, tele-workers and unpaid family workers (Parker, 2004).

Urban and rural areas

Statistics Finland uses a classification which divides Finland's 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 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 combined.[2] Hence, 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 percent of the population live 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 percent of the population live in urban settlements and in which the population of the largest urban settlement is at most 15,000. In our sample, the total number of recorded spells of self-employment classified as urban is 11,631 (59.8%) and the number of recorded spells of self-employment classified as rural is 7,808 (40.2%).

Duration model

The main emphasis of this to 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-

[2] We combined densely populated municipalities with rural rather than with urban municipalities, since the first two areas have greater similarity in their population densities. The population densities of the densely populated and rural municipalities were 25 and 10 inhabitants per km^2 in 2000, whereas the corresponding figure for the urban municipalities was 308 inhabitants per km^2.

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, for example by using a proportional hazard specification

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

where $\lambda(t)$ is a baseline hazard at time *t*, $X_i(t)$ is a vector of (time-varying) covariates, and β is a 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 a self-employment spell is only observed at the end of each year (maximum of 14 observations per spell). Therefore, the hazard rate requires a discrete time representation.

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

$$h_i(t) = prob(T_i < t+1 \mid T_i \geq t) = 1 - \exp[-\exp(\beta' X_i(t) + \gamma(t))], \quad t = 1, 2, \ldots \quad (2)$$

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 by a complementary log-log distribution (Type I extreme value). Below the discrete time duration model is estimated semi-parametrically by

allowing the baseline hazard to vary freely with duration time t (see e.g. Meyer, 1990).[3] That is, the full set of γ's are estimated by adding an indicator variable per duration time t to the model.[4] The model can be further generalized by allowing the baseline hazard to differ between populations. In this study we are particularly interested in whether it varies between rural and urban regions. Therefore, we allow the γ's to be different for self-employed persons living 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 maximized with respect to β and γ's to provide maximum likelihood estimates.[5] The empirical specification for the set of the covariates X is given below (see Table 1).[6]

Variables

The five groups of covariates and the hypotheses related to them were depicted in Section 2. Table 1 compiles the covariates, their definitions and means. First, some covariates such as gender or prior self-employment experience or parental covariates do not change in time. Another group of covariates such as level of income/wealth or prior activity is measured on

[3] 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.

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

[5] Guidelines for practical implementation are given, for example, in Jenkins (1995).

[6] Note that since the analysis is conditioned on a set of covariates *X*, then duration is implicitly a function of the entire time path of the variable *X*(*t*), *t* = 1, 2,…, which may have changed during the interval (Greene, 2003).

year before the self-employment spell in order to analyse if the initial situation has had an impact on the length of the 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 as 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 exit rates from 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.

IV. 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 in Finland. Table 2 shows the number of self-employment spells started, proportions of self-employed surviving to the end of the next year and 2002 for each cohort in both types of area (urban and rural).

Firstly, the annual changes in the number of self-employment spells started between 1988 and 2001 are noteworthy. The downward trend in the flows into self-employment in 1989–1992 can be explained by the deep recession of the Finnish economy. The downward trend was more dramatic (and began earlier) in urban than rural areas. Flows into self-employment increased considerably in 1993. The 1993 tax reform, the so-called dual income tax system, is one of the key factors responsible for this sudden increase. The reform made self-employment more attractive. The new tax system requires that the income of a self-employed person and the income of the owner of a firm are split into a labour income component and a capital income component (Riihelä *et al.*, 2001). A progressive tax rate on capital income was also replaced by a flat rate of 25 percent (currently 28%), but labour

continued to be taxed on progressive schedule (Saarimaa, 2005). Another explanation for this rise in the flows to self-employment relates to mass unemployment. Since 1984, the Finnish government has provided financial assistance to unemployed job-seekers who set up a new business. In response to the mass unemployment caused by the recession, the number of the start-up grants awarded more than doubled in 1993. Since 1993, there has been a marginally declining trend in the number of started self-employment spells in both types of location.

Secondly, the proportions of self-employed individuals surviving to the end of the next year are presented. On average, around 76 percent of those who entered self-employment in 1988–2001 were still self-employed at the end of the following year. The average survival rate is very similar in both rural and urban areas. The survival rates vary more substantially from year to year. The rates dropped during the deep recession, but have increased markedly in recent years. For example, the survival rate was around 71 percent for spells starting in 1992, whereas it was over 85 percent for spells starting in 2001. Finally, survival rates to the end of 2002 are given for spells originating in 1988–2001. We can see that of the self-employment spells starting in urban (rural) areas in 1988 only 18.2 (17.3) percent survived to the end of 2002. Since 1996, these survival rates have been higher in rural than in urban areas.

Table 3 presents the life-table estimates of self-employment survival and hazard rates for urban and rural areas. The hazard rate is defined as the probability that a self-employed individual who has survived to the beginning of an interval will experience a terminal event in an interval of one year. We can see that the hazard rate is around 24 percent in the first year of self-employment spell in both areas. Over the next two years, the hazard rate drops more in rural than urban areas. Hazard rates continue to decrease as spells get longer, falling to only 3–5 percent during the last interval (so called negative duration dependence). Interestingly, it seems that hazard rates are considerably lower in rural areas compared with urban areas at the

beginning of a spell (i.e. during the first 4 years), but do not differ much during the later intervals (or are even slightly higher in rural areas).

Comparison of the survival functions also indicates a statistical difference between the two areas. The observed level of significance 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 distributions do not differ.[7] A comparison of the mean survival times also shows considerable differences. The estimated mean survival time is 8.49 years for urban areas and 8.66 years for rural areas.[8] As we can see from Table 3, an estimate of the median survival time is four years in both areas.[9]

V. The determinants of self-employment duration

Table 4 displays the results obtained from the discrete time hazards models.[10] A variable with a positive coefficient is associated with an increased hazard rate and decreased survival time, while a variable with a negative coefficient is associated with a decreased hazard rate and increased survival time.

Three model specifications are given that differ in the specification of the 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 self-employment. The second specification is general in the sense that the baseline hazard is allowed to be completely different in rural and urban areas, thus enabling us to study regional differences in duration dependence. That is, do hazard rates in rural and

[7] The same result is obtained with the Kaplan-Meier survival analysis designed for continuous time duration analysis. Wilcoxon and Breslow tests also suggested that we should reject the null hypothesis that the survival functions are the same for the two areas.

[8] In order to account for censoring, survival functions were exponentially extended to zero when computing the estimated mean survival times.

[9] Average hazard rates can also be computed for the two areas. The average hazard rate is 0.151 for urban areas and 0.143 for rural areas.

urban areas differ at each duration time point? Finally, a more parsimonious model specification (3) is presented, which is used to test the hypotheses that

(i) hazard rates are lower in rural than urban areas at the beginning of a self-employment, and

(ii) hazard rates are lower in urban areas than in rural areas in the latter part of the spell.

Before presenting our conclusions regarding these hypotheses, we briefly discuss other results, which turned out to be robust to the specification of duration dependence.

Of the personal characteristics, gender and age behaved as expected. Men have a higher probability of remaining in business than women. Young and old self-employed individuals show a higher propensity to leave self-employment than their middle-aged counterparts. The estimates suggest that exit rates are lowest when a self-employed individual is in her/his forties. The results on the effect of education support the hypothesis that higher education increases duration of self-employment. The language variable did not quite reach statistical significance in all specifications. Hence, contrary to our expectations, we did not find that the Swedish-speaking minority has a significantly higher probability of remaining in business than the Finnish-speaking majority. Marriage increases the duration of self-employment substantially, while having children does not seem to have an impact. These results at least partly strengthen the importance of family support.

Initial wealth seems to have an impact on self-employment duration: both house ownership and the family's income level prior to the self-employment spell lower the exit rates and thus increase length of self-employment. Differences in exit rates by prior activity and experience also merge. The estimated effect of previous labour market status shows that

[10] As our data include multiple self-employment spells experienced by single individuals, the standard errors are adjusted to allow for the same individuals being repeatedly at risk. In addition, the hazard rates are allowed to differ according to the experience of previous self-employment spells.

individuals who were in paid employment prior to the self-employment spell were more successful in remaining self-employed than those who were unemployed or students (cf. Carrasco, 1999). Experience of migration before entering self-employment increases hazard rates, as was expected. Industry also plays an important role in self-employment duration. Individuals self-employed in transport, storage or communications and private services have slower exit rates than individuals in manufacturing. On the contrary, in the hotel and restaurant industry the exit rates from self-employment are greater than in manufacturing.

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 alternate 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, an individual's self-employment duration increases if her/his father has been self-employed. Although the estimated effect of mother's self-employment does not reach significance, these results suggest the importance of early life experience.[11]

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

[11] We also tried out more general specification, which included parents' education as additional variables. Since their estimated coefficients were very small and insignificant, they were dropped from the final specification.

The preceding descriptive analysis based on follow-up life-tables showed that hazard rates are on average lower in rural areas compared with urban areas at the beginning of a spell, but might be even slightly higher at later intervals. The second specification allows us to study whether these findings still hold after other factors have been controlled for. The estimation results are in line with the descriptive analysis. Figure 1 shows the predicted hazard rates for a typical self-employed individual (median observation).[12] We can see that the longer this individual remains self-employed the smaller are the hazard rates in both areas. However, the hazard rates for the second and the third year of the spell are considerably lower in rural areas than in urban areas.[13]

Finally, the third specification lends further support to our findings. The estimated coefficient on the covariate rural$\times\sum_1^4 d_t$ is significantly negative. 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. The estimated coefficient on the covariate rural$\times\sum_5^{14} d_t$ is positive but insignificant.

VI. Conclusion

This paper showed that self-employment duration depends on many factors. Environmental factors and circumstances in the local labour market also have resonance, but this effect is not straightforward. The survival functions between rural and urban areas differ. On average, survival times are longer in rural than urban locations. Rural areas as compared to urban areas

[12] Hazard rates have been predicted using Model 2 estimates in Table 4 and sample median values for covariates. That is, the self-employed individual is a 40-year-old married male with children, whose native language is not Swedish. He is a house-owner and has secondary education. Before the current self-employment spell in wholesale or retail trade he was in paid employment with an annual income of 13,300 €. He did not move or was not previously self-employed and his spouse's annual income was 5,200 €. His parents have not been self-employed. The regional unemployment rate is 15.1%, it has fallen by 4.8% and year is 1997.

[13] The negative estimated coefficients of the covariates rural$\times d_j$ (j = 1, 2, 3, 4) suggest that exit rates are lower in rural areas compared with urban areas during the first four years of self-employment. The reverse seems to hold for most of the remainder of the spell ($t \geq 5$).

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 location does not have a universal effect on exit rates from self-employment, but after separating the lengths of the spells significant effects were found. Rural areas have lower hazard rates in the first years of spells, while urban areas may even have slightly lower rates in 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: aside from the fact that labour markets with weak paid-employment opportunities push workers into self-employment they force individuals to continue in self-employment, even if the firm did not turn out to be very profitable in the first place. If the firms survive the difficult first years, hazard rates will be considerably smaller thereafter.

Acknowledgements

The authors would like to thank session discussants at the 52nd North American Meetings of the Regional Science Association International, Las Vegas (10–12 November 2005), 45th Congress of the European Regional Science Association, Amsterdam (23–27 August 2005), and ONS Analysis of Enterprise Microdata Conference, Cardiff (7–9 September 2005) for their useful comments. The paper forms part of a Research programme on Business Know-how (LIIKE 2, project no. 112116) financed by the Academy of Finland.

References

Becker, G. (1975) *Human Capital*, Columbia University Press, New York.

Blanchflower, D. G. (2000) Self-Employment in OECD Countries, *Labour 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**, 315–41.

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.

Cueto, B. and Mato, J. (2006) An analysis of self-employment subsidies with duration models, *Applied Economics*, **38**, 23-32.

Evans, D. S. and Jovanovic, B. (1989) An Estimated Model of Entrepreneurial Choice under Liquidity Constraints, *Journal of Political Economy*, **97**, 808–27.

Evans, D. S. and Leighton, L. S. (1989) Some Empirical Aspects of Self-Employment, *The American Economic Review*, **79**, 519–35.

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., Joulfain, 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**, 129–38.

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, 649–70.

Kangasharju, A. and Pekkala, S. (2002) The Role of Education in Self-Employment Success in Finland, *Growth and Change*, **33**, 216–37.

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**, 757–82.

Moore C. S. and Mueller R. E. (2002) The transition from paid to self-employment in Canada: the importance of push factors, *Applied Economics*, **34**, 791-801.

Niittykangas, H. and Tervo, H. (2005) Spatial Variations in Intergenerational Transmission of Self-Employment, *Regional Studies*, **39**, 319–32.

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.

Riihelä, M., Sullström, R., Suoniemi, I. and Tuomala, M. (2001) Inequality in Finland during 1990s', in *Down from the Heavens, up from the Ashes. The Finnish Economic Crisis of the 1990s in the Light of Economic and Social Research* (eds.) J. Kalela, J. Kiander, U. Kivikuru, H.A. Loikkanen and J. Simpura, VATT Publications 27(6), pp. 385–410.

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.

Saarimaa, T. (2005) Taxation and Debt Financing of Home Acquisition: Evidence from the Finnish 1993 Tax Reform, VATT Discussion papers 366, Helsinki.

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**, 411–31.

Taylor, M. (1999) Survival of the Fittest? An Analysis of Self-Employment Duration in Britain, *Economic Journal*, **109**, C140–55.

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. (2006) Regional unemployment, self-employment and family background, *Applied Economics,* **38**, 1055-62

Tervo, H. and Niittykangas, H. (1994) The Impact of Unemployment on New Firm Formation in Finland, *International Small Business Journal*, **13**, 38–53.

Uusitalo R. (2001) Homo entreprenaurus?, *Applied Economics*, **33**, 1631-38.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT, London.

Table 1. Description of covariates and their means

Covariate	Description	Mean		
		Urban areas	Rural areas	All
I Region-specific factors				
Rural	1 if workplace of self-employed is in rural area, 0 if in urban area	0.000	1.000	0.414
Unemplev†	Unemployment level (%) in the region	14.498	16.592	15.365
Unempcha†	Unemplev (*t*) / Unemplev (*t*–1)	1.070	1.053	1.063
II Personal and household characteristics				
Female	1 if female, 0 otherwise	0.365	0.341	0.355
Swedish	1 if native language is Swedish, 0 otherwise	0.047	0.061	0.053
Age†	Age in years	40.274	39.531	39.966
Age2†	Age squared divided by 100	17.096	16.498	16.848
Marri†	1 if married or cohabiting, 0 otherwise	0.729	0.745	0.735
Childr†	1 if has children in the family, 0 otherwise	0.517	0.574	0.541
Interedu†, a	1 if secondary education (equivalent of 10–12 years of education), 0 otherwise	0.412	0.494	0.446
Highedu†, a	1 if tertiary education, 0 otherwise	0.320	0.212	0.275
III Assets				
Income‡	Annual income subject to state taxation, 10 000 €	1.692	1.381	1.563
Spincome‡	Annual income of spouse, 10 000 €	1.100	0.817	0.983
House‡	1 if owner-occupier of a house or a flat, 0 otherwise	0.727	0.791	0.754
IV Prior activity and experience				
Unempl‡, b	1 if unemployed, 0 otherwise	0.149	0.176	0.160
Otheract‡, b	1 if other activity (incl. student), 0 otherwise	0.193	0.204	0.197
Migrat‡	1 if moved to another region, 0 otherwise	0.036	0.053	0.043
Entrexpe‡, c	1 if self-employment experience, 0 otherwise	0.359	0.503	0.419
Entrfath[d]	1 if father has been self-employed, 0 otherwise	0.222	0.383	0.289
Entrmoth[d]	1 if mother has been self-employed, 0 otherwise	0.225	0.382	0.290
V Industrial sector (reference category manufacturing)				
Trade†	1 if the industry is wholesale or retail trade, 0 otherwise	0.207	0.203	0.205
Hotel†	1 if the industry is hotel or restaurant, 0 otherwise	0.048	0.052	0.050
Building†	1 if the industry is construction, 0 otherwise	0.117	0.161	0.135
Transpor†	1 if the industry is transport, storage or communications, 0 otherwise	0.078	0.111	0.092
Businser†	1 if the industry is business services, 0 otherwise	0.101	0.057	0.083
Otherser†	1 if the industry is private services, 0 otherwise	0.113	0.083	0.100
Otherind†	1 if self-employed in some other industry, 0 otherwise	0.246	0.203	0.228

Notes: The number of self-employment spells is 19,439 and the number of spell-years (observations) is 79,918. 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] Reference category is employed. [c] Using information from 1970, 1975, 1980, 1985 and 1987–2001. [d] Using information from 1970, 1975, 1980 and 1985.

Table 2. Survival rates in urban and rural areas

Starting in	Urban areas			Rural areas		
	Number of spells	Survive to end of the next year, %	Survive to end of 2002, %	Number of spells	Survive to end of the next year, %	Survive to end of 2002, %
1988	880	75.9	18.2	637	75.8	17.3
1989	1,107	72.4	18.4	689	75.5	21.8
1990	942	71.7	21.7	745	71.7	23.9
1991	720	76.5	27.6	549	72.7	23.9
1992	715	71.3	26.2	529	71.3	28.7
1993	988	75.8	29.7	771	73.9	30.4
1994	844	72.6	33.8	587	73.6	38.0
1995	825	73.5	40.4	536	71.1	37.3
1996	720	77.6	43.8	458	77.1	44.1
1997	818	76.2	46.1	460	78.7	50.2
1998	820	77.2	51.5	480	78.8	53.1
1999	779	79.3	59.8	461	81.8	64.0
2000	770	77.5	69.5	455	80.9	72.3
2001	703	85.5	85.5	451	85.6	85.6
All spells	11,631	75.7	39.4	7,808	75.8	39.4

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	11,631	75.7	24.3	7,808	75.8	24.2
2	8,203	61.8	18.3	5,534	63.6	16.2
3	6,167	52.6	15.0	4,310	55.1	13.3
4	4,779	46.7	11.3	3,442	48.9	11.2
5	3,817	42.0	10.0	2,800	44.0	10.0
6	3,058	38.3	8.8	2,288	39.7	9.7
7	2,475	35.4	7.5	1,864	36.4	8.4
8	1,956	32.6	7.9	1,508	33.4	8.2
9	1,517	30.7	5.8	1,161	31.4	6.0
10	1,136	29.0	5.7	857	29.6	5.6
11	884	27.1	6.5	657	27.7	6.5
12	628	25.7	5.1	483	26.4	4.8
13	392	24.2	5.9	282	24.9	5.7
14	165	23.5	3.0	116	23.6	5.2

Notes: Estimated mean survival time was 8.49 years for urban areas and 8.66 years for rural areas. Estimated median survival time is 4 years for urban and rural areas. Log-rank test for the equality of the survivor functions: $\chi^2 = 17.03$ with d.f. = 1 ($p < 0.001$).

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 Region-specific factors						
Rural	-0.032	(0.021)				
Rural$\times\sum_1^4 d_t$					-0.056*	(0.023)
Rural$\times\sum_5^{14} d_t$					0.068	(0.044)
Unemplev	0.002	(0.003)	0.002	(0.003)	0.002	(0.003)
Unempcha	0.119**	(0.039)	0.113**	(0.039)	0.115**	(0.039)
II Personal and household characteristics						
Female	0.126**	(0.022)	0.127**	(0.022)	0.127**	(0.022)
Swedish	-0.093*	(0.047)	-0.092	(0.047)	-0.092	(0.047)
Age	-0.102**	(0.007)	-0.102**	(0.007)	-0.102**	(0.007)
Age2	0.123**	(0.009)	0.123**	(0.009)	0.123**	(0.009)
Marri	-0.260**	(0.031)	-0.260**	(0.031)	-0.261**	(0.031)
Childr	0.038	(0.027)	0.038	(0.027)	0.039	(0.027)
Interedu	-0.069**	(0.023)	-0.070**	(0.023)	-0.070**	(0.023)
Highedu	-0.109**	(0.028)	-0.108**	(0.028)	-0.108**	(0.028)
III Assets						
Income	-0.049**	(0.010)	-0.049**	(0.010)	-0.049**	(0.010)
Spincome	-0.024**	(0.009)	-0.024**	(0.009)	-0.023**	(0.009)
House	-0.206**	(0.022)	-0.206**	(0.022)	-0.205**	(0.022)
IV Prior activity and experience						
Unempl	0.287**	(0.027)	0.287**	(0.027)	0.287**	(0.027)
Studen	0.230**	(0.026)	0.231**	(0.026)	0.231**	(0.026)
Migrat	0.174**	(0.042)	0.175**	(0.042)	0.174**	(0.042)
Entrexpe	0.132**	(0.022)	0.132**	(0.022)	0.132**	(0.022)
Entrfath	-0.140**	(0.030)	-0.141**	(0.030)	-0.141**	(0.030)
Entrmoth	-0.054	(0.029)	-0.054	(0.029)	-0.054	(0.029)
V Industrial sector (reference category manufacturing)						
Trade	0.064	(0.038)	0.064	(0.038)	0.064	(0.038)
Hotel	0.232**	(0.049)	0.232**	(0.049)	0.232**	(0.049)
Building	-0.030	(0.043)	-0.030	(0.043)	-0.030	(0.043)
Transpor	-0.765**	(0.060)	-0.766**	(0.060)	-0.767**	(0.060)
Businser	0.020	(0.050)	0.020	(0.050)	0.020	(0.050)
Otherser	-0.290**	(0.049)	-0.290**	(0.049)	-0.290**	(0.049)
Otherind	0.682**	(0.037)	0.682**	(0.037)	0.682**	(0.037)

Table 4. (Continued)

Covariate	Model 1		Model 2		Model 3	
	Coeff.	(Std.err.)	Coeff.	(Std.err.)	Coeff.	(Std.err.)
VI Duration dummies						
d_1	0.609**	(0.151)	0.606**	(0.152)	0.629**	(0.152)
d_2	0.345*	(0.154)	0.393*	(0.155)	0.365*	(0.155)
d_3	0.205	(0.156)	0.253	(0.157)	0.226	(0.156)
d_4	0.040	(0.157)	0.037	(0.159)	0.061	(0.158)
d_5	-0.018	(0.159)	-0.029	(0.162)	-0.049	(0.159)
d_6	-0.047	(0.161)	-0.106	(0.165)	-0.079	(0.161)
d_7	-0.153	(0.164)	-0.214	(0.170)	-0.185	(0.164)
d_8	-0.056	(0.166)	-0.089	(0.174)	-0.088	(0.166)
d_9	-0.286	(0.175)	-0.319	(0.187)	-0.318	(0.175)
d_{10}	-0.292	(0.181)	-0.305	(0.197)	-0.324	(0.181)
d_{11}	-0.111	(0.185)	-0.136	(0.204)	-0.143	(0.185)
d_{12}	-0.344	(0.204)	-0.335	(0.233)	-0.376	(0.205)
d_{13}	-0.115	(0.226)	-0.104	(0.261)	-0.147	(0.227)
d_{14}	-0.331	(0.338)	-0.588	(0.472)	-0.363	(0.338)
Rural$\times d_1$			-0.003	(0.031)		
Rural$\times d_2$			-0.135**	(0.044)		
Rural$\times d_3$			-0.131*	(0.054)		
Rural$\times d_4$			-0.002	(0.067)		
Rural$\times d_5$			0.016	(0.079)		
Rural$\times d_6$			0.125	(0.091)		
Rural$\times d_7$			0.126	(0.109)		
Rural$\times d_8$			0.064	(0.121)		
Rural$\times d_9$			0.065	(0.161)		
Rural$\times d_{10}$			0.021	(0.190)		
Rural$\times d_{11}$			0.048	(0.203)		
Rural$\times d_{12}$			-0.034	(0.274)		
Rural$\times d_{13}$			-0.037	(0.325)		
Rural$\times d_{14}$			0.524	(0.604)		
VII Year dummies (reference category 1989)						
y_{1990}	0.114	(0.068)	0.119	(0.068)	0.115	(0.068)
y_{1991}	0.149	(0.091)	0.159	(0.091)	0.155	(0.091)
y_{1992}	0.164*	(0.079)	0.168*	(0.079)	0.165*	(0.079)
y_{1993}	0.195*	(0.080)	0.197*	(0.080)	0.195*	(0.080)
y_{1994}	-0.001	(0.075)	-0.001	(0.075)	-0.002	(0.075)
y_{1995}	0.117	(0.074)	0.119	(0.074)	0.116	(0.074)
y_{1996}	0.084	(0.073)	0.086	(0.073)	0.083	(0.073)
y_{1997}	-0.149*	(0.071)	-0.149*	(0.071)	-0.150*	(0.071)
y_{1998}	-0.193**	(0.070)	-0.193**	(0.070)	-0.195**	(0.070)
y_{1999}	-0.248**	(0.069)	-0.248**	(0.069)	-0.250**	(0.069)
y_{2000}	-0.242**	(0.067)	-0.242**	(0.067)	-0.243**	(0.067)
y_{2001}	-0.271**	(0.068)	-0.270**	(0.068)	-0.272**	(0.068)
y_{2002}	-0.863**	(0.074)	-0.862**	(0.074)	-0.864**	(0.074)
Log-likelihood	-30,276.0		-30,267.0		-30,272.6	

Notes: Standard errors are adjusted for clustering at the individual level. Number of observations: 79,918. Variable definitions are given in Table 1. * (**) = Statistically significant at the 0.05 (0.01) level.

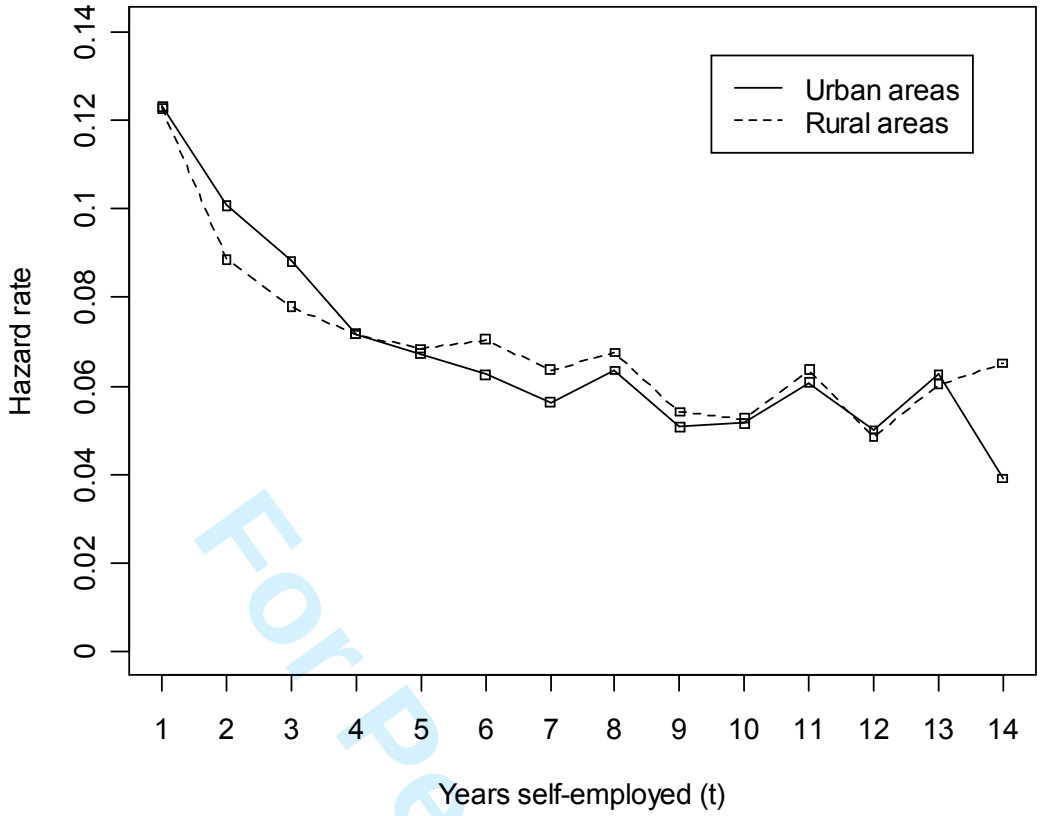


Figure 1. Predicted hazard rates in urban and rural areas (Model 2 in Table 4)