

Title:

Self-employment or paid-employment as first job: Evidence from Spain by nationality

Abstract

In period of recession or zero economic growth and increasing unemployment in most industrialized countries, self-employment emerges as an alternative to paid-employment. Moreover, the positive effects of self-employment on economic growth and job creation have been widely documented in economic literature. Recent economic literature has focused on longitudinal studies of self-employment concentrating on transitions into self-employment and self-employment survival (see Georgellis et al. 2005, for a literature review). This paper follows this line of investigation and sheds new empirical evidence for Spain about the transitions into self-employment or paid-employment of potential entrants to the labour market, and their duration in both labour market states. The econometric methodology consists in estimating discrete choice model and survival models controlling by personal, job characteristics and unobserved heterogeneity. In all models, individual nationality is included between the regressors. Data used come from the longitudinal data of the Spanish Social Security (*Muestra Continua de Vidas Laborales, MTAS 2008*). One of the main results of this paper is, on one hand, that the immigrants are a disadvantaged group on the entry into self-employment as first job. Moreover, once they enter self-employment state are more likely to exit from it. On other hand survival analysis in the first job shows that self-employment relationships are more stable than those for salaried worker. Moreover, in self-employment and paid-employment, unobservable heterogeneity is a relevant factor to explain the job's duration.

JEL Classification: J15, J41, J44

Author: Antonio Caparrós Ruiz

Universidad de Málaga

antonio@uma.es; tlf: 952 13 11 64; fax: 952 13 72 62

Postal adress

Facultad de Ciencias Económicas y Empresariales

Departamento de Estadística y Econometría

Campus de El Ejido, 29071 Málaga

Self-employment or paid-employment as first job: Evidence from Spain by nationality*

**Antonio Caparrós Ruiz
Universidad de Málaga**

1. Introduction

The positive effects of self-employment on economic growth and job creation has been widely documented in economic literature [for instance, Andretsch and Thurik (2001); Thurik *et al.* (2008)]. Moreover, nowadays, in period of recession or zero economic growth and increasing unemployment in most industrialized countries, self-employment emerges as an alternative to paid-employment (Evans and Leighton 1989). On other hand, there also relevant non-pecuniary benefits associated with self-employment, such as more flexible working schedules and higher levels of job satisfaction [see, for example, Blanchflower and Oswald (1998), OECD (2000)].

Within the EU25, self-employment accounted for 16% of total employment in 2005 (EUROSTAT 2006), with 19% of men self-employed, compared with 11% of women and, by activity sector, the construction industry records the highest rate of self-employment (25%). Among the member states, the highest rates in 2005 were registered in Greece (32%) and Italy (29%), while the lowest percentages were recorded in Latvia (7%) and Estonia (8%). In Spain, according to the Spanish Labour Force Survey (INE 2008) the number of self-employed workers is around 3 million people in 2008, which represents around 17% of employed people.

Recently, economic literature has focused on longitudinal studies of self-employment concentrating on transitions into self-employment and self-employment survival (see Georgellis *et al.* 2005, for a literature review). In Spain, Alba-Ramírez

* The authors gratefully acknowledge financial support from Proyecto SEJ 2007-68045-CO2-01/ECON, Ministerio de Ciencia y Tecnología (Spain).

(1994) focusing on US and Spanish displaced workers examines whether their current unemployment status affects their probability of self-employment transition. The Spanish data used in this study come from the Working and Living Conditions (INE 1985). Their main finding indicates that longer unemployment duration is positively associated with self-employment entry, with the effect being stronger in the US than Spain. In addition, Carrasco (1999) using the Spanish Continuous Family expenditures Survey (INE 1985-1994) investigates the influence of individual characteristics and business cycle on the probability self-employment duration for men. One of their main results is that hazard rate decreases with duration in self-employment.

This paper sheds new empirical evidence about the transitions into self-employment and duration in this state for Spain. In particular, it has a twofold objective. First, under a cross-sectional framework, it analyses the personal factors determining the choice facing the potential entrants to the labour market between self-employment and paid-employment. In such a case, the econometric methodology consists in estimating discrete choice model. Second, under a longitudinal framework, it studies the dynamic and duration of both labour market states, estimating survival models and controlling by personal, job characteristics and unobserved heterogeneity. In all models, individual nationality is included between the regressors. Data used come from the longitudinal data of the Spanish Social Security (*Muestra Continua de Vidas Laborales, MTAS 2008*). One of the advantages of this survey is that their panel structure allows constructing job duration, with accurately, for new workers entering self-employment or paid-employment.

The paper is structured as follows. Section 2 gives a description of the data used. Econometric models and empirical results are presented in section 3. Finally, section 4 provides conclusions.

2. Data

The data used come from the administrative data set of the Spanish Social Security (*Muestra Continua de Vidas Laborales*, MTAS 2008). This data set is representative of the employed people in Spain, and is formed on approximately 1.1 million people who have had an affiliation in the Social Security in the year 2004, containing information on their entire labour market history. For this study, the advantages of these data set with respect to others data set such as the Spanish Labour Force Survey or the European Union Household Panel for Spain, is that offers a more exact measure of the employment spells than the others. In particular, the exact data when each job begins and ends is known.

The initial sample used in this paper included information about individuals who have had their first labour relationship between the period 2004- 2007¹, with a job duration for at least 3 months. Thus seasonal or causal jobs, that are expected to last only until the end of a season (for example, summer jobs in the case of students), are not considered. Moreover, similar to Carrasco (1999), the agriculture sector is not included owing to the special characteristics of self-employment in this sector. Finally, this gives a total number of 57.120 individuals. Table 1 shows descriptive statistics of the variables used in the estimation of the model that determines the individual choice between paid-employment and self-employment.

First, the results highlight than self-employed people accounts for 7.2% of employed people, being Spanish workers the collective more represented (73%) followed by EU workers (14%) and South American workers (8%). Within paid-employment workers, Spanish people constitute 59% of the total, and South American people is the immigrant's collective more relevant within the total of paid-employment workers (20%). Second, self-employed workers are on average 6 years older than their

¹ The *Muestra Continua de Vidas Laborales* is only representative from year 2004, because the random sample (4% of the total population) has been drawn in this year.

employed counterparts (32 years *versus* 26 year, respectively). Thirdly, for gender, the female self-employed workers amounts to 49%, 6 percentage points higher than to the figure for paid-employment workers. On other hand, the distribution of workers by educational levels shows than the rate of employed with more than primary education is higher for self-employed than for paid-employed (37% versus 25%).

Table 1. Sample Descriptives of variables used in the discrete choice model

Variables	Paid-employment		Self-employment		Total	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Geographical origin						
Spain	0.59	0.49	0.73	0.45	0.61	0.49
European Union	0.11	0.32	0.14	0.35	0.12	0.32
Rest of Europe	0.02	0.14	0.01	0.11	0.02	0.14
Africa	0.05	0.23	0.02	0.14	0.05	0.22
Asia	0.02	0.14	0.02	0.15	0.02	0.15
South America	0.20	0.40	0.08	0.27	0.19	0.39
Age	26.17	8.83	32.48	11.73	26.62	9.22
Gender						
Female	0.43	0.50	0.49	0.50	0.44	0.50
Male	0.57	0.50	0.51	0.50	0.56	0.50
Education						
Less than primary education	0.45	0.50	0.31	0.46	0.44	0.50
Primary Education	0.30	0.46	0.32	0.46	0.30	0.46
More than primary education	0.25	0.43	0.37	0.48	0.26	0.44
Activity						
Manufacturing	0.12	0.32	0.09	0.29	0.12	0.32
Construction	0.19	0.40	0.14	0.34	0.19	0.39
Wholesale, retail trade...	0.21	0.40	0.27	0.44	0.21	0.41
Hostels and restaurants	0.14	0.35	0.13	0.33	0.14	0.34
Transport	0.03	0.17	0.02	0.15	0.03	0.17
Financial intermediation	0.01	0.09	0.02	0.12	0.01	0.09
Real estate, renting and busines activities	0.16	0.36	0.21	0.40	0.16	0.36
Education	0.04	0.20	0.01	0.12	0.04	0.19
Health	0.04	0.20	0.04	0.20	0.04	0.19
Other social and personal service activities	0.06	0.24	0.06	0.24	0.06	0.24
Spanish Regions						
Andalusia	0.16	0.36	0.20	0.40	0.16	0.36
Galicia	0.05	0.22	0.07	0.26	0.05	0.22
Castilla-La Mancha	0.05	0.22	0.05	0.20	0.05	0.22
Extremadura	0.02	0.13	0.02	0.15	0.02	0.13
Valencia	0.11	0.31	0.13	0.33	0.11	0.31
Murcia	0.03	0.17	0.03	0.18	0.03	0.17
Canary Islands	0.05	0.21	0.05	0.22	0.05	0.21
Asturias	0.02	0.13	0.03	0.15	0.02	0.14
Cantabria	0.01	0.10	0.01	0.10	0.01	0.10
Basque Country	0.04	0.20	0.05	0.21	0.04	0.20
Navarra	0.01	0.11	0.01	0.09	0.01	0.11
Aragón	0.03	0.17	0.03	0.16	0.03	0.17
La Rioja	0.01	0.07	0.01	0.08		
Madrid	0.17	0.37	0.10	0.30	0.17	0.37
Castilla-León	0.04	0.20	0.05	0.22	0.04	0.20
Balearic Islands	0.03	0.16	0.03	0.17	0.03	0.16
Catalonia	0.18	0.38	0.12	0.32	0.17	0.37
GDP growth rate	3.62	0.22	3.60	0.24	3.62	0.22
Sample size	53018		4102		57120	

Source: MCVL (MTAS 2008).

Concerning to activity sector, self-employment is focused on wholesale (27%) and real estate activities (21%), followed by construction (14%) and hostels and restaurants (13%). The distribution of paid-employed by activity sector is qualitatively similar than for self-employed. By Spanish regions, two-third of paid-employed people is concentrated in Catalonia, Andalusia, Madrid and Valencia, being the first region the more represented within the distribution with a rate of 18%. With respect to self-employment, more than 50% of people reside on some of these four regions, representing Andalusia the 20% of self-employed. Finally the variable GDP growth rate, in the year in which the individuals start their job, is included to consider the influence of the business cycle. For this variable the mean value is 3.62% for both self-employment and paid-employment.

In table 2, descriptives for the resulting spell data set can be found. First, it's observed that 76.6% of self-employed spells are right censored, in contrast with the rate of 30.3% for paid-employed spells. Moreover, average duration for completed spells is around fourth months higher for self-employed than for paid-employment. These results show evidence that, in this sample, self-employment relationships are more stable than those for salaried workers.

Table 2. Descriptives for employment spells (%)

	Paid-employment	Self-employment	Total
Completed spells	69.7	23.4	66.2
<i>Mean duration in days</i>	248	359	250
Right censored spells	30.3	76.6	33.8
Total	100	100	100

Source: MCVL (MTAS 2008).

In order to get an idea on the shape of distributions of self-employment and paid-employment duration, the latter evidence is completed obtaining non-parametric estimation of the survivor functions relying on Kaplan-Meier product-limit estimators.

Let $t_1 < t_2 < \dots < t_j < \dots < t_k < \infty$ represent the survival times that are observed in the data set. Let d_j be the number of completed spells at t_j , let m_j be the number of observations censored in the interval $[t_j, t_{j+1}]$, and let n_j the number of observation ending their spell immediatly prior to t_j , which is made up of those who have a censored or completed spell of lenght t_j or longer:

$$(1) \quad n_j = \sum_{i \geq j}^k (m_i + d_i)$$

Then the Kaplan-Meier estimate of the survivor function is given by the product of one minus the number of exits (d_j) divided by the number of persons at risk of exit just before t_j (n_j):

$$(2) \quad \hat{S}(t) = \prod_{j/t_j \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

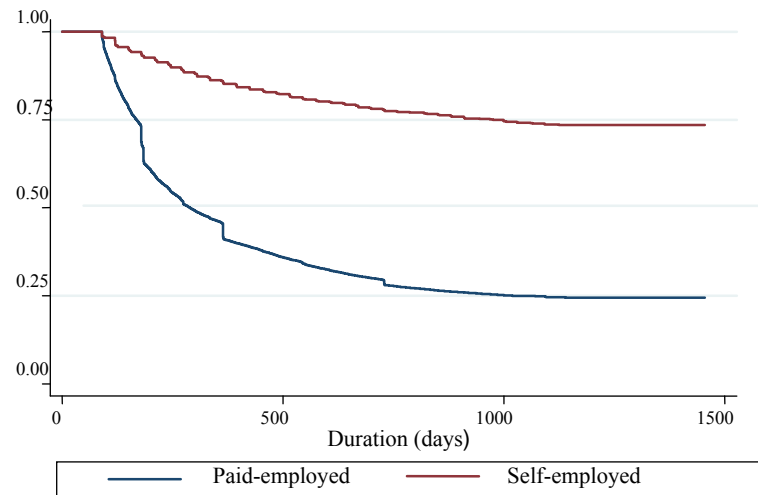
In this case, the Kaplan-Meier estimates is the nonparametric hazard estimated by the number of exits from the self-employment (paid-employment) in each period divided by the population still in this labour state at the beginning of that period. Thus, this nonparametric survival technique allows us to examine the probability of individual leaving the self-employment (paid-employment) state conditional of having survived in the origin state for the whole interval.

The Kaplan-Meier estimates, which appear in figure 1, suggest that the self-employed have higher survival rates than do salary workers. Moreover, looking at the survivor function, the hazard rates are stable since two years.

In figures 2 and 3, the empirical survivor functions of self-employed and salary workers by nationality are plotted. For self-employed, the hazard rate of exit is higher for immigrants than for natives in all periods. In the other case, survivor functions in both collectives, immigrants and natives, are quite similar. Therefore, the application of the logrank test (Mantel-Cox method) rejects the equality of survival distribution

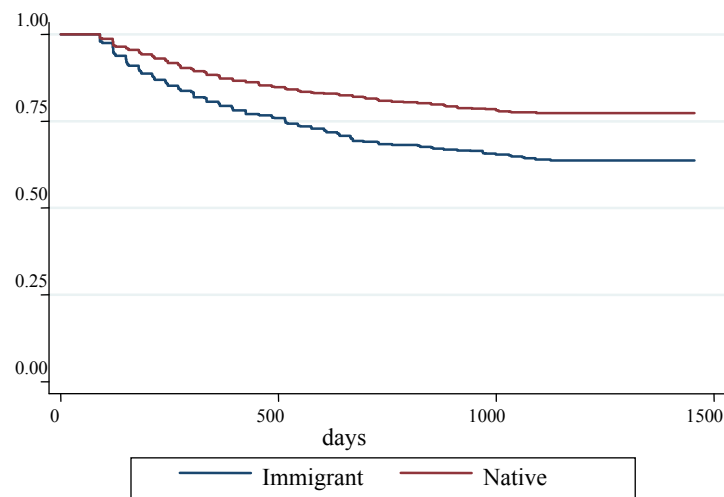
between native and immigrant workers both to the paid-employment and for self-employment.

Figure 1. Survivor functions for self-employed and paid-employed



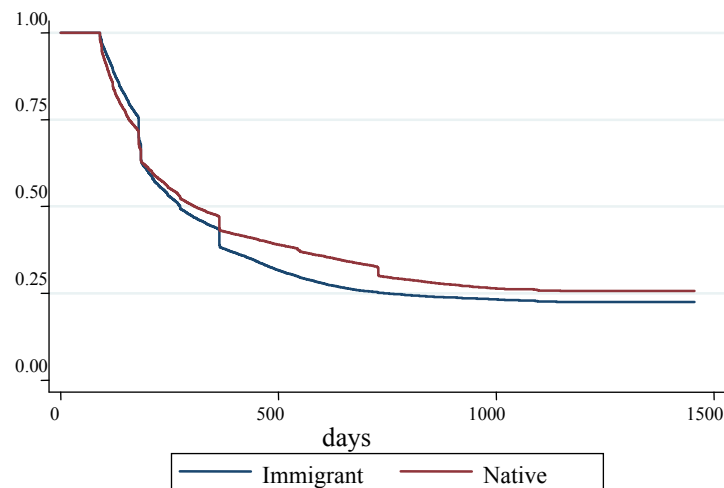
Source: MCVL (MTAS 2008).

Figure 2. Survivor functions by self-employed and nationality



Source: MCVL (MTAS 2008).

Figure 3. Survivor functions by paid-employed and nationality



Source: MCVL (MTAS 2008).

3. Econometric specifications and results

This section is structured as follows. In subsection 3.1, first, the econometric model used to estimate the probability of being self-employed compared to being paid-employed is specified. Second, the model is estimated and the results are interpreted. On other hand, subsection 3.2 is devoted to the study of the dynamics of self-employment and paid-employment by the estimated duration models.

3.1 Probability of being self-employed compared to being paid-employed

An individual decides whether or not to become self-employed in his first job, comparing the market wage he would earn as a salaried worker, W_i , with the expected net income from self-employment, R_i . In this way, it's possible to define the following index function:

$$(1) \quad I_i = R_i - W_i = X_i' \beta + \varepsilon_i$$

(2)

where X_i is the set of exogenous variables that collects personal and socioeconomic characteristics which affect R_i and W_i . In such a case, the regressors considered are: nationality, age, gender, education, activity sector, Spanish region of residence and GDP growth rate in the year to initiate the employment relationship. β is a vector of unknown parameters and ε_i is a disturbance term that is supposed that follows a logistic distribution. I_i is not observed but is known if a person is self-employed or paid-employed, so a binary variable Y_i is built taking the value of 1 if the individual is employed and 0, otherwise. Then, the probability of self-employment is given by:

$$(2) \quad \text{Prob}(Y_i = 1) = \text{Prob}(I_i > 0) = F(X_i' \beta)$$

where F is the logistic cumulative density function.

Table 3 reproduces the estimate of the logit model of entry into self-employment. We first note that a large number of significant variables are shown to explain the probability of being self-employed compared to being paid-employed.

Concerning personal characteristics, first, the results show that the probability of being entrepreneurship is lower for immigrant workers than for native counterparts. In particular, the probability of being self-employed for South-American and African workers is lower in 5.8 and 4.1 percentage points than for Spanish ones. On one hand, this result may be explained because these groups of immigrants face capital constraints that limit their access to self-employment (see, for example, Blau and Graham 1992 and Blanchflower *et al.* 1998). On other hand, immigrant may face discriminatory barriers such as in securing customer (Georgellis *et al.* 2005).

Table 3. Logit estimates of the probability of being self-employed

Variables	Coefficients^a	Marginal effects
Geographical origin		
European Union	-0.638***	-0.025***
Rest of Europe	-1.532***	-0.039***
Africa	-1.565***	-0.041***
Asia	-0.699***	-0.025***
South America	-1.929***	-0.058***
Age	0.078***	0.003***
Gender		
Male	0.196***	0.009***
Education		
Primary Education	0.372***	0.018***
More than primary education	0.673***	0.037***
Activity		
Construction	0.076	0.003
Wholesale, retail trade...	0.561***	0.030***
Hostels and restaurants	0.339***	0.017***
Transport	-0.115	-0.005
Financial intermediation	0.436**	0.024**
Real estate, renting and business activities	0.220**	0.011**
Education	-1.136***	-0.034***
Health	-0.001	-0.001
Other social and personal service activities	0.225**	0.011**
Spanish Regions		
Galicia	-0.005	-0.001
Castilla-La Mancha	-0.150*	-0.006*
Extremadura	0.080	0.003
Valencia	-0.008	-0.001
Murcia	-0.051	-0.002
Canary Islands	-0.143*	-0.006*
Asturias	-0.129	-0.005
Cantabria	-0.320**	-0.013**
Basque Country	-0.072	-0.003
Navarra	-0.193	-0.008
Aragón	-0.238**	-0.010
La Rioja	-0.027	-0.001
Madrid	-0.695***	-0.026***
Castilla-León	-0.049	-0.002
Balearic Islands	0.077	0.003
Catalonia	-0.418***	-0.017***
GDP growth rate	-0.433***	-0.020***
Constant	-3.266***	
Log-likelihood ratio test	3730.56***	
Sample size	57120	

(a) The reference is a Spanish woman without primary education, working in the industry sector and living in Andalusia.

(b) (****) Significant at 1%, (**) at 5%, (*) at 10%.

Source: MCVL (MTAS 2008).

Second, worker's age influences positively on the probability of being self-employed. This can be explained because the responsibility to face the tasks of a

business can be a variable correlated positively with the age of the individual. With respect to gender, male odd-ratio of being self-employed compared to being paid-employed is 1.2 times higher than for female workers. Self-employment is in many cases an activity which requires a large work-load. This may make it more difficult for women to be self-employed, especially for those with young children² (Andersson, P. and Wandensjo, E. 2008). Another important conclusion is that educational attainment and self-employment are positively correlated. In particular, people with more than primary education have a probability of being self-employed 1.8 and 3.7 percentage points higher than worker without primary education. On one hand this result is opposite according to the filter theory (Arrow 1973). This theory proposes the signalling hypothesis, that implies between other predictions that those planning to enter self-employment have no need to acquire formal qualification to indicate their quality to potential employers. But, on other hand, Lucas (1978) stated that more able and creative individuals become entrepreneurs and the rest become wage workers. This leads to an expectation that the more educated is an individual is the more likely he will become entrepreneur, as education could serve as a filter such that the more educated tend to be better informed, implying that they are more efficient at assessing self-employment opportunities. Moreover, for more educated workers are more likely to find external finance such as equity and loans for their business (see, for example, Bates 1990 and Cressy 1996).

Concerning job's characteristics, on one hand, self-employed workers are more likely to be employed in wholesale or financial intermediation sector with a probability higher in 3 and 2.4 percentage points higher than in manufacturing sector. On other hand, education sector is the least likely to have self-employed (3.4 less than in manufacturing sector). For Spanish regions, Madrid and Catalonia that are the most

² Unfortunately, the dataset used in this paper lacks of family characteristics, such as marital status or number of children, which could be important in this issue.

developed are where workers have less likely to be self-employed with respect to Andalusia. In these former regions, the greatest opportunities for wage employment discourage self-employment. Relating business conditions, the GDP growth rate exerts a negatively influence on self-employment and is a push factor to paid-employment.

3.2 Duration of self-employment versus paid-employment

In this subsection the join effect of various personal, job and labour market characteristics affecting the probability of exit from self-employment or paid-employment are analysed.

So, let T_i be the length of the employment spell of individual i . The hazard for individual i at time t , $\lambda_i(t)$ can be defined by the following equation:

$$(3) \lim_{h \rightarrow 0} \frac{\text{prob}(t+h > T_i \geq t / T_i \geq t)}{h} = \lambda_i(t)$$

A widely used way to allow for the presence of observed variables affecting the duration in each labour market state (self-employment or paid-employment) is the Cox proportional hazard model (Cox 1972). In this model, the instantaneous hazard rate of exit from self-employment or paid-employment at time t , conditional on survival to time t is:

$$(4) \lambda_i(t; x_i) = \lambda_0(t) \exp(x_i' \beta)$$

where X_i is a vector of covariates for individual i , β is the vector of unknown coefficients and $\lambda_0(t)$ denotes the baseline hazard function. An advantage of this model is that it's not necessary to assume a particular form for the baseline hazard function.

Moreover, the Cox model assumes that the hazard functions of any two individuals with different values on one or more covariates differ only by a factor of proportionality:

$$(5) \quad \frac{h(t; x_i)}{h(t; x_j)} = e^{(x_i - x_j)\beta}$$

where X_i and X_j denote vectors of characteristics for two persons i and j at some t . The possible interpretation of the estimated coefficient β_k in the proportional hazards specification is that each coefficient summarizes the proportional effect of the absolute changes in the corresponding variable X_k on the hazard rate:

$$(6) \quad \beta_k = \partial \ln \lambda(t, X_{ik}) / \partial X_{ik}$$

Estimations of this model are shown in table 4. In the second column appear the estimated coefficients that show the risk of exit for self-employment. First, it's observed that all immigrant collectives, with exception of Asian workers, have higher risk of exit than Spanish workers. In particular, EU workers have a hazard that is more than double that of the Spaniards. Second, age is found to have a positive and significant effect on duration. On one hand, it's possible that this variable is a proxy of the acquisition of non formal human capital necessary for the good performance of the business. On other hand, it could also be explained by the job shopping theory (Jonhson 1978) which states that younger individuals tend to go into riskier occupation. With respect to gender, male workers have less risk of exit from self-employment than women. By activity sector, the biggest job instability arises within the education, financial intermediation, construction and tourism sector. For the first two cases, the hazard of exit from self-employment is 1.6 times higher than the sectors not cited. For the last two sectors, is 1.3 times higher than the rest of sectors. By Spanish regions, it's observed little variability in the risk of leaving self-employment, only Extremadura shows a lower risk of exit (0.6 times lower

than the rest). Finally, the economic cycle approximated from GDP growth rate has a positive effect on the duration of self-employment.

Table 4. Estimation of the Cox regression model

Variables	Self-employment Coefficients^a	Paid-employment Coefficients^a
Geographical origin		
European Union	0.852***	0.071***
Rest of Europe	0.578**	0.203***
Africa	0.465**	0.338***
Asia	0.162	0.101**
South America	0.685***	0.160***
Age	-0.028***	-0.007***
Gender		
Male	-0.275***	-0.053***
Education		
Primary Education	-0.107	-0.038**
More than primary education	-0.110	-0.120***
Activity		
Construction	0.305**	0.264***
Wholesale, retail trade...	0.082	0.119***
Hostels and restaurants	0.285**	0.286***
Transport	0.146	0.103**
Financial intermediation	0.479*	-0.331***
Real estate, renting and business activities	0.008	0.189***
Education	0.497*	0.148***
Health	-0.037	-0.001
Other social and personal service activities	-0.276	0.180***
Spanish Regions		
Galicia	-0.164	-0.158***
Castilla-La Mancha	-0.265	-0.132***
Extremadura	-0.494*	-0.184***
Valencia	-0.059	-0.001
Murcia	-0.170	-0.026
Canary Islands	0.143	-0.047*
Asturias	-0.054	-0.143**
Cantabria	-0.133	-0.113**
Basque Country	0.033	-0.106***
Navarra	-0.413	-0.157**
Aragón	-0.120	-0.051*
La Rioja	-0.266	0.039
Madrid	-0.052	-0.177***
Castilla-León	-0.182	-0.166***
Balearic Islands	0.073	0.142***
Catalonia	-0.009	-0.117***
GDP growth rate	-0.135*	-0.342***
Loglikelihood ratio test	198.69	1704.58
Sample size	4.102	53.018

(a) The reference is a Spanish woman without primary education, working in the industry sector and living in Andalusia.

(b) (****) Significant at 1%, (**) at 5%, (*) at 10%.

Source: MCVL (MTAS 2008).

Concerning to paid-employment duration, we can observe that almost all explanatory variables are relevant to explain the risk of leaving paid-employment. First,

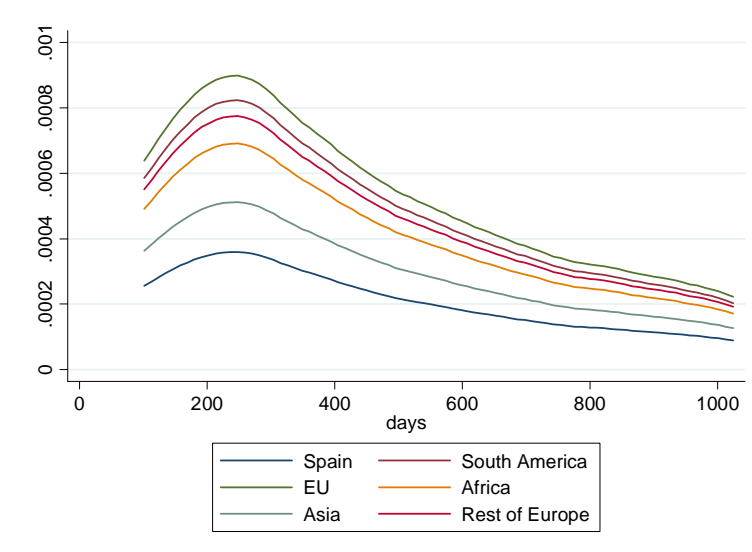
again the Spanish workers have a lower risk of leaving job to foreign workers. Therefore, in such a case the differences between among different nationalities are lower than in the case of self-employed and in a distinct order. Now, on one hand, Spanish and European Union workers present a hazard of leaving employment quite similar. On other hand, the employees of Africa and South America have the highest risk of exit from paid-employment. Second, age and gender are two relevant variables to explain the survival of employees in their jobs but, similar to the variables of nationality, their influence on the risk of exit from paid-employment is lower than in the case of self-employment. Referring to variables of human capital, there is a positive relationship between educational attainment and the probability of staying in employment. Thus, the postulates of human capital theory are verified because this theory shows a positive relationship between the accumulation of knowledge, productivity, and a successful career (Becker 1962; Black and Lynch 1996).

By activity sector, construction next to the tourism sector presents the greatest risk of leaving the job with regard to manufacturing sector. Both cases present a risk of leaving the job that is 30% higher than the industry. By contrast, workers in the sector of financial intermediation have the lowest risk of leaving the job. Coefficient relating to dummies representing Spain's regions showed that Andalusia, Valencia Murcia and La Rioja present a lower risk of leaving the employment than the rest of regions. Finally, the influence of economic cycle on the risk of exit from paid-employment has the same sign that in the case of self-employment, but the intensity is higher.

Figures 4 and 5 show the predicted hazard functions evaluated at mean values of covariates, for self-employed and paid-employed workers. In these graphs, it's observed that the exit rates from the current job increases to a maximum near the beginning of new employment and declines over time (the bell-shaped pattern). This can be explained, because in the initial phase of a new job, it's relatively easy for a worker to

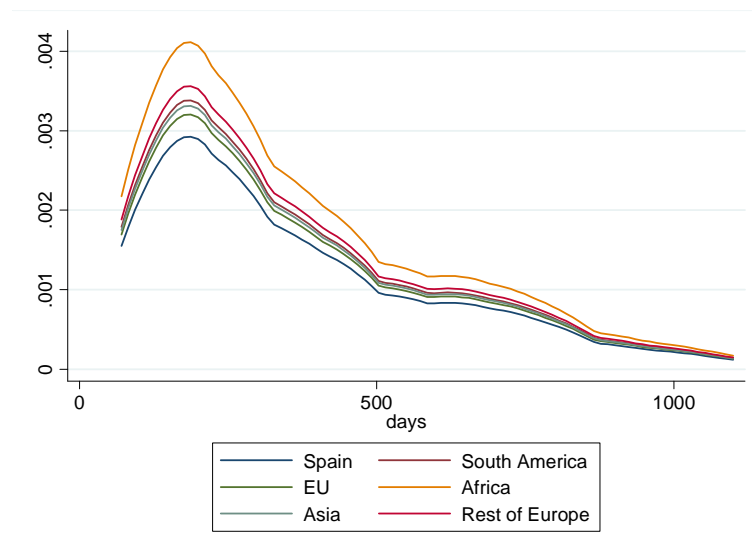
move out of his job searching for a better job, if the current job reward is lower than his expectation. But after the match has lasted for a while, workers are less likely to have the job because their experience is a valuable resource associated with current and future reward. In the case of self-employed workers, the uncertainty is prevalent during the first stages of business formation. As time elapses entrepreneurs learn of their own abilities by participating in the actual running of business and noting how well they perform (Jovanovic 1982).

Figure 4. Predicted hazard of exit from self-employment (Cox model)



Source: MCVL (MTAS 2008).

Figure 5. Predicted hazard of exit from paid-employment (Cox model)



Source: MCVL (MTAS 2008).

The later Cox duration model implies that the exit from the self-employment (paid-employment) occurs in continuous time and is quite suitable for these data because the spells are measure in days. Nevertheless, one of the assumption of the continuous time Cox model presented is that the inter-individual heterogeneity (differences existing among individuals) can be explained by the set of covariates, that is to say by the set of observed characteristics. However, a common issue in estimating duration models is the possibility of allowing for unobserved individual heterogeneity to be incorporated in the model, because uncontrolled heterogeneity can bias the effects of included explanatory variables.

The control of unobserved heterogeneity requires the use of grouped-data because in these data the exit from self-employment (paid-employment) may occur at any particular day. In this case, the variable duration in the self-employment (paid-employment) state is grouped into number of months in such state³.

Following Meyer (1990), we assume that the hazard rate is parameterized using a proportional hazard form. Then, the corresponding discrete-time hazard rate for individual i of exit in the interval (a_{j-1}, a_j) , in this case the length of interval is a month is:

$$(7) \quad h_j(x_i) = 1 - \exp[-\exp(x_i' \beta + \gamma_j + \log(\varepsilon_i))]$$

where ε_i is a positive-valued random variable with density $f_\varepsilon(\varepsilon)$ that control by the individual heterogeneity. The choice of f density for ε is then required to enable the calculation of the log-likelihood. In this model, it's assumed that ε is a Gamma distributed random variable with mean one and finite variance θ^4 . On other hand, γ_j is a function describing duration dependence in the discrete time hazard. To allow a flexible

3 Although the estimation of discrete time duration model with group data allows to control unobservable heterogeneity, it's necessary consider that the measure of the employment duration are less precise than in the initial Cox model with spells expressed in days.

⁴ The likelihood-ratio test of $H_0: \theta = 0$ can be used to test the existence of unobserved heterogeneity.

specification of the baseline hazard, we do not impose any restrictions on how γ_j vary from month to month, then a specific parameter for each time interval is considered, this implies the introduction in the model of time dummies for each month. Estimation model is carried out via Jenkins “pghaz routine” in Stata (Jenkins 1997)⁵.

Table 5 displays the results. First, it’s observed that the variance of the unobservable heterogeneity is significantly different from zero as for exits from self-employment as for exist from paid-employment implying that frailty effect is found. In particular, this result shows the importance of factors such as the ability, motivation or social contacts on survival employment (self-employment and paid-employment). Another consequence of the presence of unobservable heterogeneity is the existence of bias in estimated regression coefficients from Cox models. Thus, we firstly observe that the new results report that the positive difference between the hazard rates of exit from self-employment or paid-employment for foreigner workers, compared with their Spanish counterparts entrepreneurs, increase with respect to the estimates of the Cox model.

In particular, in exit from self-employment, EU workers have a hazard of exit that triples than of Spaniards, respectively. On other hand, South American and workers from rest of Europe shows a hazard that more than doubles the existing one for Spanish. In the case of paid-employment, the hazard of exit of foreigner workers is also higher than of Spaniards, although the difference between Spanish workers and the rest of national groups diminishes in relation with the existing ones for self-employment. Moreover, now, African workers show the highest hazard rate (one and a half times larger than the corresponding to the Spanish employees). On the contrary, it should be emphasized that EU and Asian workers have a hazard of exit only an 8% higher than Spaniards.

⁵ Details about the specification of the log-likelihood function can be found also in Jenkins (1997).

Table 5. Estimation of discrete-time model

Variables	Self-employment Coefficients^a	Paid-employment Coefficients^a
Geographical origin		
European Union	1.161***	0.080**
Rest of Europe	0.805**	0.161***
Africa	0.505*	0.333***
Asia	0.129	0.077**
South America	0.933***	0.129***
Age	-0.036***	-0.008***
Gender		
Male	-0.384***	-0.066***
Education		
Primary Education	-0.156	-0.057***
More than primary education	-0.165	-0.156***
Activity		
Construction	0.5191**	0.287***
Wholesale, retail trade...	0.166	0.135***
Hostels and restaurants	0.429**	0.321***
Transport	0.156	0.133***
Financial intermediation	0.682*	-0.314***
Real estate, renting and business activities	0.061	0.228***
Education	1.014**	0.184***
Health	0.100	0.012
Other social and personal service activities	-0.361	0.208***
Spanish Regions		
Galicia	-0.161	-0.159***
Castilla-La Mancha	-0.276	-0.133***
Extremadura	-0.544*	-0.199***
Valencia	-0.057	-0.006
Murcia	-0.254	-0.026
Canary Islands	0.219	-0.046*
Asturias	-0.035	-0.152***
Cantabria	-0.065	-0.109**
Basque Country	0.060	-0.096**
Navarra	-0.345	-0.164**
Aragón	0.058	-0.043
La Rioja	-0.533	0.039
Madrid	-0.189	-0.195***
Castilla-León	-0.170	-0.170***
Balearic Islands	0.227	0.156***
Catalonia	-0.024	-0.127***
GDP growth rate	-0.056	0.025
Time dummies		
4 months	-0.056	-0.334***
5 months	-0.332*	0.017
6 months	-0.382*	0.138***
7 months	-0.301	-0.688***
8 months	-0.203	-0.708***
9 months	-0.252	-0.717***
10 months	-0.276	-1.140***
11 months	-0.358	-1.283***
12 months	-0.290	0.122***
13 months	-0.452	-1.489***
14 months	-0.731*	-1.382***
15 months	-0.489	-1.291***
16 months	-0.779**	-1.383***
17 months	-0.227	-1.495***
18 months	-0.518	-1.109***
19 months	-0.926**	-1.620***
20 months	-0.944**	-1.567***
21 months	-0.577	-1.644***
22 months	-0.402	-1.823***
23 months	-0.900**	-1.962***
24 months	-0.378	-0.690***
25 months	-1.164**	-2.180***
26 months	-2.108**	-2.245***
27 months	-1.074**	-2.270***
28 months	-0.773*	-2.364***
29 months	-0.803*	-2.614***
30 months	-0.643	-2.475***
31 months	-1.785**	-2.608***
32 months	-0.870*	-2.601***
33 months	-0.526	-2.705***
34 months	-1.022*	-3.934***
35 months	-1.442**	-3.838***
Constant	-2.485***	-1.745***
Loglikelihood ratio test of unobservable heterogeneity (H₀: θ = 0, no unobservable heterogeneity)	4.599***	15.490***
Sample size	90288	554372

(a) The reference is a Spanish woman without primary education, working in the industry sector and living in Andalusia.

(b) (****) Significant at 1%, (**) at 5%, (*) at 10%.

Source: MCVL (MTAS 2008).

With respect to age, the estimated coefficients show that the positive influence of this variable on the survival employment rises in relation to the obtained with the Cox model. Thus, for example, the increase in a year at the age of the worker causes a decrease of 3.6% in the hazard rate of exit from self-employment. Likewise, it occurs with the estimated coefficient of the gender dummy variable. Now, on one hand, male workers have a risk of exit from self-employment and paid-employment that is smaller than that of women by 38.4% and 6.6%, respectively. On other hand, the hazard rate for salary workers with more than primary education is 12% lower than the corresponding to salary workers without primary education.

In relation to activity dummy variables, the greatest difference with respect to the Cox model appears in the hazard rate of exit from self-employment. In particular, the estimated coefficients report an increase in the hazard rate of self-employed in education, financial intermediation, construction and hostels and restaurants with respect to the existing for workers in the industry sector.

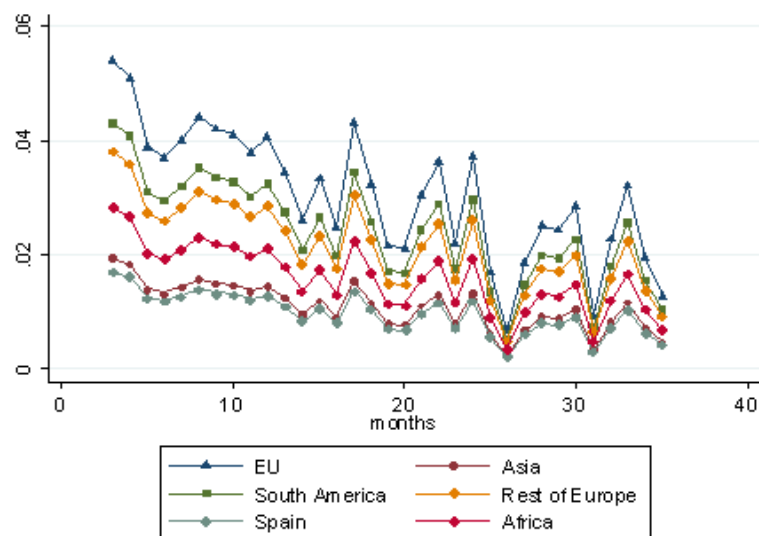
With reference the dummy variables that display the Spanish region where work the individual, it's remarkable that there are few differences with the estimates of the Cox model, thus the significance and magnitude of the coefficients is similar.

About the variable that denotes the GDP growth rate, in contrast with the estimates of the Cox model, this regressor is not relevant to explain the exit from self-employment or paid-employment. This result can be explained because the GDP growth rate is correlated with the set of monthly dummy variables introduced in the model to capture duration dependence.

With regard to the coefficients attached to the monthly dummy variables, in the case of paid-employment, these coefficient show a decreasing pattern with respect to the hazard achieved between the third and fourth month. In the model that explains self-employment duration, the behaviour of the set of coefficients attached to dummies

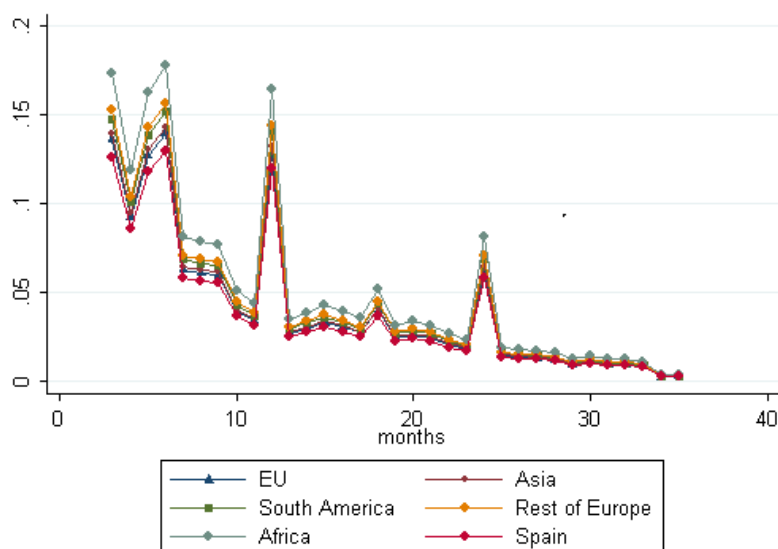
which capture duration dependence is not as homogenous, although the smaller hazards of exit from self-employment appear in the first two years with exception of some months. Finally, in Figures 5 and 6 are plotted the predicted hazard functions evaluated at mean values of covariates, for self-employed and paid-employed workers. In these graphs, we can observe that the predicted hazard for all period is lower for the Spanish workers with respect to the predicted hazard for the rest of national groups. Moreover, in the case of paid-employment, two highlighted peaks are detected in the twelfth and twenty-fourth months. This can be explained because in the Spanish labour market a large proportion of fixed-term contract end in these both periods.

Figure 5. Predicted hazard of exit from self-employment (Discrete-time model with unobservable heterogeneity)



Source: MCVL (MTAS 2008).

**Figure 6. Predicted hazard of exit from paid-employment
(Discrete-time model with unobservable heterogeneity)**



Source: MCVL (MTAS 2008).

4. Conclusions

In this paper, we have investigated two issues. First, the determinants in the individual choice between self-employment and paid-employment as first job have been analysed. Second, we have studied the factors influencing the duration in the employment spells in both labour market states. Data come from the *Muestra Continua de Vidas Laborales* (MTAS 2008). These data are specially useful in the second objective because their panel structure provides an exact measure of the employment spells.

One of the main results of this paper is that the immigrants are a disadvantaged group on the entry into self-employment as first job. Moreover, once they enter self-employment state are more likely to exit from it. Given the positive effects of self-employment on economic growth and job creation, this result should be present to policy makers when it comes to economic policies to promote the assimilation of immigrants into the Spanish labour market. Therefore, the immigrants are not an

homogenous collective. Thus, the probability of being self-employed for South-American and African workers is lower in 5.8 and 4.1 percentage points than for Spanish ones, while for EU and Asian workers is lower only in 2.5 percentage points. Moreover, EU workers have a hazard of exit that triples than of Spaniards whereas Asian workers have the same hazard rate that Spanish ones.

Other interesting results about the probability of being self-employed are, on one hand, that more aged and male people show higher probability of being entrepreneurs. On other hand, the individual educational level is correlated positively with the probability of being self-employed, which opposes the filter theory (Arrow 1973) that proposes education is not a relevant variable to entry into self-employment.

Considering survival analysis in the first job, results show that self-employment relationships are more stable than those for salaried workers. Moreover, in both labour market states (self-employment and paid-employment) unobservable heterogeneity is a relevant factor to explain the duration in the first job; that is, motivation, ability and social contacts are important elements to initiate a successful and stable career.

References

- Alba-Ramírez, A. (1994): “Self-employment in the midst of unemployment: the case of Spain and the US”. *Applied Economics*, vol. 26, issue 3, pp.189-204.
- Andersson, P. and Wandensjö, E. (2008): “A gender perspective on self-employment entry and performance as self-employed”. *IZA DP*, n°. 358.
- Andretsch, D. and Thurik, A. (2001): “What is new about new economy: sources of growth in the managed and entrepreneurial economies. *Industrial and Corporate Change* 10 (1), pp. 267-315.
- Arrow, K. (1973): “Higher education as a filter”. *Journal of Public Economics* 2 (3), pp. 193-216.

- Becker, G. (1962): "Investment in human capital: a theoretical analysis". *Journal of Political Economy*, 70, n° 5, pp. 9–49.
- Black, S. and Lynch, L. (1996): "Human capital investment and productivity". *American Economic Review*, 86, n° 2, pp. 263–67.
- Blanchflower, D. G. and Oswald, A. J. (1998): "What makes an entrepreneur?" *Journal of Labour Economics*, vol. 16, issue. 1, pp. 26-60.
- Blanchflower, D. G., Levine, P. B. and Zimmermann, D. J. (1998): "Discrimination in the small business credit market". *NBER Working Paper* n°. 68440.
- Blau, F. D. and Graham, J. W. (1999): "Black-white difference in wealth and asset composition". *The Quarterly Journal of Economics*, vol. 105, issue 2, pp. 321.339.
- Bates, T. (1990): "Entrepreneur human capital inputs and small business longevity". *The Review of Economics and Statistics*, vol. 72, issue 4, pp. 551-559.
- Carrasco, R. (1999): "Transitions to and from self-employment in Spain: an empirical analysis". *Oxford Bulletin of Economics and Statistics*, vol. 61, issue 3, pp. 315-341.
- Cox, D. (1972): "Regression models and life-tables". *Journal of the Royal Statistical Society*, 34 (may/august), pp. 187-220.
- Cressy, R. (1996): "Are business start-ups debt-rationed?" *The Economic Journal*, vol. 106, pp. 1253-1270.
- Evans, D. and Leighton, L. (1989): "Some empirical aspects of entrepreneur-ship". *The American Economic Review*, vol. 79, pp. 519-35.
- EUROSTAT (2006): *New Release* n° 133.
- Georgellis, Y., Sessions, J. and Tsitsianis N. (2005): "Self-employment longitudinal dynamics: a review literature". *Economic Issues*, vol.10, n° 2.

- Jenkins, S. P. (1997): “Discrete time proportional hazard regression”. *Stata Technical Bulletin*, n° 39, pp. 22-32.
- Johnson W. R. (1978): “A theory of job shopping”. *Quarterly Journal of Economics*, vol. 92, pp. 261-278.
- Jovanovic, B. (1982): “The selection and evolution of industry”. *Econometrica*, vol. 50, pp. 649-670.
- Light, I. (1980): “Disadvantaged minorities in self-employment” *International Journal of Comparative Sociology*, vol. 20, pp. 31-45.
- Lucas, R. E. (1978): “On the size distribution of business firms”. *Bell Journal of Economics* n° 9, pp. 508-523.
- Meyer, B. (1990): “Unemployment insurance and unemployment spells”, *Econometrica*, n° 58, pp. 757-782.
- OECD (2000): *Employment Outlook*. Paris: OECD (sept).
- Thurik *et al.* (2008): “Does self-employment reduce unemployment? *Journal of Business Venturing* (forthcoming).