

# Self-employment as first job choice in Spain: Evidence by nationality

ANTONIO CAPARRÓS RUIZ



Centro de Estudios Andaluces  
CONSEJERÍA DE LA PRESIDENCIA

El Centro de Estudios Andaluces es una entidad de carácter científico y cultural, sin ánimo de lucro, adscrita a la Consejería de la Presidencia de la Junta de Andalucía.

El objetivo esencial de esta institución es fomentar cuantitativa y cualitativamente una línea de estudios e investigaciones científicas que contribuyan a un más preciso y detallado conocimiento de Andalucía, y difundir sus resultados a través de varias líneas estratégicas.

El Centro de Estudios Andaluces desea generar un marco estable de relaciones con la comunidad científica e intelectual y con movimientos culturales en Andalucía desde el que crear verdaderos canales de comunicación para dar cobertura a las inquietudes intelectuales y culturales.

**Las opiniones publicadas por los autores en esta colección son de su exclusiva responsabilidad**

© 2010. Fundación Centro de Estudios Andaluces. Consejería de Presidencia. Junta de Andalucía  
© Autores

Ejemplar gratuito. Prohibida su venta.

---



Centro de Estudios Andaluces  
CONSEJERÍA DE LA PRESIDENCIA

E2010/03

## Self-employment as first job choice in Spain: Evidence by nationality

Antonio Caparrós Ruiz\*

### RESUMEN

Los efectos positivos del autoempleo sobre el crecimiento económico y la creación de empleo han sido contrastados ampliamente en la literatura económica. Recientemente, el fenómeno del autoempleo se ha analizado desde una perspectiva longitudinal, centrándose en las transiciones y permanencia en dicho estado laboral. Este trabajo sigue esta línea de investigación, ofreciendo evidencia empírica sobre la elección entre el empleo autónomo o asalariado que realizan los individuos al iniciar su vida laboral, y la duración que presentan en ambas situaciones laborales. La metodología econométrica consiste en la estimación de modelos de elección discreta y de supervivencia, incluyéndose entre los regresores a la nacionalidad de los individuos. Los datos utilizados en el análisis provienen de la Muestra Continua de Vidas Laborales (MTAS, 2008).

Palabras clave: Autoempleo, empleo asalariado, transiciones, nacionalidad

### ABSTRACT

The positive effects of self-employment on economic growth and job creation have been widely documented in the economic literature. Recent economic literature has focused on longitudinal studies of self-employment concentrating on the transition into and survival in self-employment. This paper follows this line of investigation and offers new empirical evidence for Spain about the transition into self-employment or paid employment of potential entrants to the labor market and their duration in both segments of the market. The econometric methodology consists of estimating discrete-choice models and survival models, controlling for personal characteristics, job characteristics, and unobserved heterogeneity. In all models, nationality is included in the regressors. The data used come from the longitudinal data of the Spanish Social Security Survey (*Muestra Continua de Vidas Laborales*, MTAS 2008).

Keywords: Self-employment, paid-employment, transitions, nationality

JEL classification: J15, J41, J44

\* Departamento de Estadística y Econometría  
Universidad de Málaga  
Campus de El Ejido s/n, 29071 Málaga  
Teléfono: 952 13 11 63; e-mail: antonio@uma.es

## 1. Introduction

The positive effects of self-employment on economic growth and job creation have been widely documented in the economic literature (for instance, Andretsch and Thurik, 2001; Thurik et al., 2008). Moreover, in the current period of recession or zero economic growth and increasing unemployment that is evident in most industrialized countries, self-employment is emerging as an alternative to paid employment (Evans and Leighton, 1989). In addition, relevant nonpecuniary benefits are 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 and 11% of women being self-employed; by sector of activity, the construction industry had 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 were recorded in Latvia (7%) and Estonia (8%). In Spain, according to the Spanish Labor Force Survey (INE 2008), the number of self-employed workers was around 3 million in 2008, or around 17% of the labor force.

Recently, the economic literature has focused on longitudinal studies of self-employment concentrating on the transition into self-employment and survival in self-employment (see Georgellis et al., 2005 for a literature review). Alba-Ramírez (1994), examines whether the current unemployment status of displaced workers affects their probability of entering self-employment. The Spanish data used in this study come from the Working and Living Conditions Survey (INE 1985). The main finding indicates that longer duration of unemployment is positively associated with entry into self-employment. In addition, Carrasco (1999), using data from the Spanish Continuous Family Expenditures Survey (INE 1985–94), investigates the influence of individual characteristics and business cycle on the duration in self-employment for men.

One of Carrasco's main results is that the risk of leaving self-employment decreases with the duration in self-employment.

This paper offers new empirical evidence about the transition into and duration in self-employment in Spain. In particular, it has two objectives. First, using a cross-sectional framework, it analyzes the personal factors determining the choice between self-employment and paid employment facing potential entrants to the labor market. The econometric methodology consists of estimating a discrete-choice model. Second, using a longitudinal framework, it studies the dynamic and duration of both labor market states, estimating survival models and controlling for personal characteristics, job characteristics, and unobserved heterogeneity. In all models, individual nationality is included in the regressors. Data used come from the longitudinal data of the Spanish Social Security Survey (Muestra Continua de Vidas Laborales, MTAS 2008). One of the advantages of this survey is that the panel structure makes it possible to construct job duration, with accuracy, for new workers entering self-employment or paid employment.

The paper is structured as follows. Section 2 describes the data used. Section 3 presents the econometric models and empirical results. Finally, section 4 provides conclusions.

## 2. Data

The data used come from the administrative data set of the Spanish Social Security Survey. This data set is formed of approximately 1.1 million people affiliated with the social security system in 2004, containing information on their entire labor market history. For this study, the advantage of this data set with respect to other data sets, such as the Spanish Labor Force Survey or the European Union Household Panel for Spain, is that it offers a more exact measure of the duration of employment than the others. In particular, the exact date when each job began and ended is known.

The initial sample used in this paper includes information about individuals who had their first job in the period 2004–07,<sup>1</sup> with a job duration of at least three months. Thus seasonal or casual jobs, which 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. This yields a total of 57,120 individuals. Table 1 shows descriptive statistics of the variables used to estimate the model determining the individual choice of paid employment or self-employment.

---

<sup>1</sup> The *Muestra Continua de Vidas Laborales* is only representative for 2004, because the random sample (4% of the total population) is from this year.

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

Variables	Paid employment		Self-employment		Total	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<b>Geographic origin</b>						
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
<b>Age</b>	26.17	8.83	32.48	11.73	26.62	9.22
<b>Gender</b>						
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
<b>Education</b>						
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
<b>Activity</b>						
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 and retail trade	0.21	0.40	0.27	0.44	0.21	0.41
Hotels 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 business 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
<b>Spanish regions</b>						
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

First, the results indicate that self-employed people account for 7.2% of all employed people, with Spanish workers being more represented (73%), followed by European Union (EU) workers (14%) and South American workers (8%). Within salaried workers, Spanish workers constitute 59% of the total, and South American workers are the most common immigrant group (20%). Second, self-employed workers are, on average, six years older than their salaried counterparts (32 and 26 years, respectively). Third, by gender, females account for 49% of self-employed workers, 6 percentage points higher than the figure for salaried workers. The distribution of workers by educational level shows that the percentage of self-employed workers who have more than a primary education is higher than the percentage of salaried workers (37% and 25%, respectively).

Concerning sector of activity, self-employment is focused on wholesale (27%) and real estate activities (21%), followed by construction (14%) and hotels and restaurants (13%). The distribution of paid employment by sector of activity is qualitatively similar to that of self-employment. By Spanish regions, two-thirds of people in paid employment are found in Catalonia, Andalusia, Madrid, and Valencia. Catalonia has the highest share of paid employment, with a rate of 18%. With respect to self-employment, more than 50% of people reside in one of these four regions, with Andalusia representing 20% of the 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 3.6% for both self-employment and paid employment.

Table 2 presents data on period of employment. First, 76.6% of self-employment spells are right censored in contrast with 30.3% of paid-employment spells. Moreover, average duration of completed spells is around four months longer for self-employment than for paid employment. These results show evidence that, in this sample, relationships are more stable for self-employed workers than for salaried workers.

Table 2. Employment spells, by type of employment (%)

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

Source: MTAS (2008).

In order to get an idea of the duration of self-employment and paid employment, nonparametric estimation of the survivor functions is obtained by the use of 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  $n_j$  be the number of observations ending their spell immediately prior to  $t_j$ , which is made up of those who have a censored or completed spell of length  $t_j$  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 1 minus the number of exits ( $d_j$ ) divided by the number of persons at risk of exit just before  $t_j$  ( $n_j$ ):

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

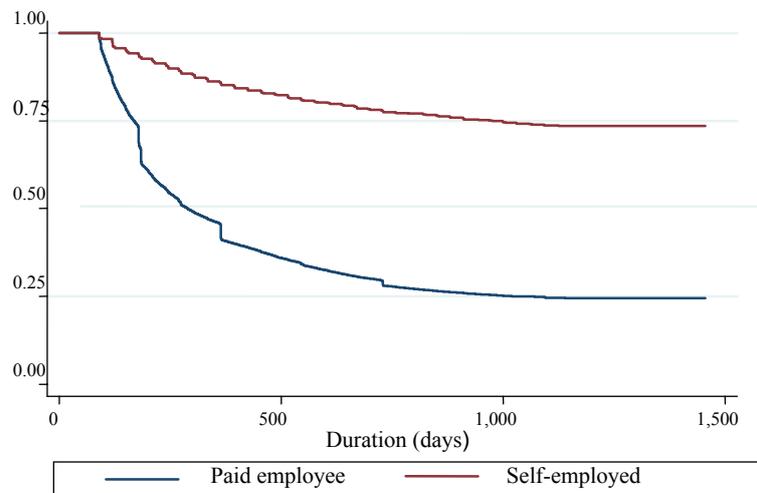
In this case, the Kaplan-Meier estimate is the nonparametric hazard estimated by the number of exits from self-employment (paid employment) in each period divided by the population still in this segment of the labor force at the beginning of that period. Thus this nonparametric survival technique allows

us to examine the probability that an individual will leave self-employment (paid employment) conditional on having survived in the original state for the whole interval.

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

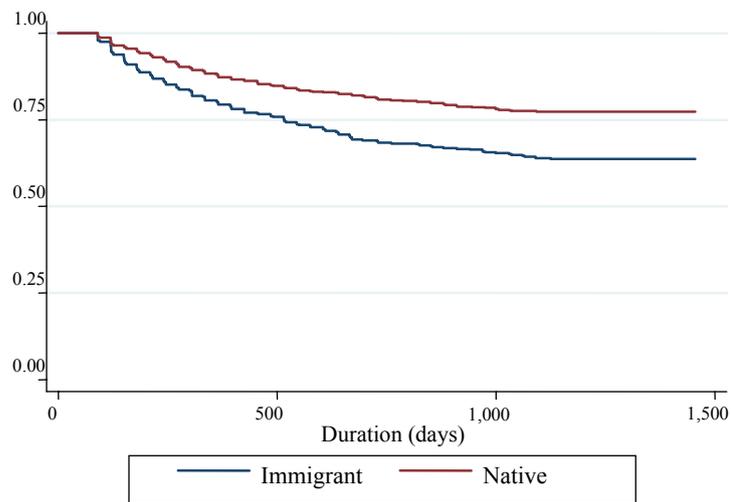
Figure 1. Survivor functions for self-employed and salaried workers

Source: MTAS (2008).



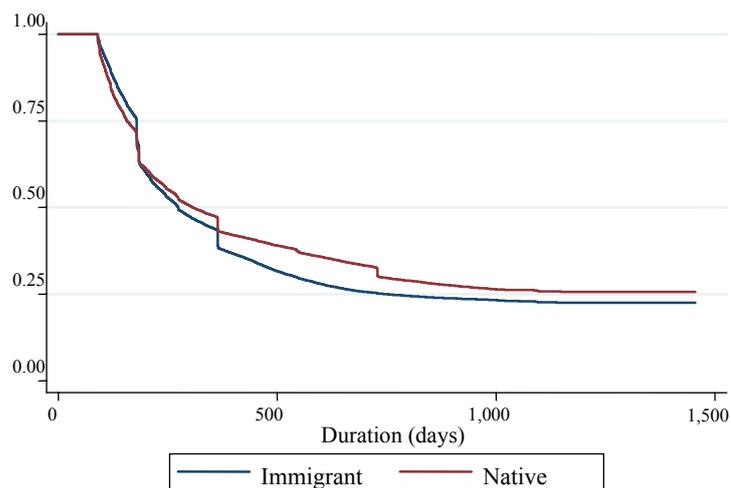
Figures 2 and 3 plot the empirical survivor functions of self-employed workers and paid employees by nationality. For self-employed workers, the risk of exit is higher for immigrants than for natives in all periods. For salaried workers, survivor functions in both collectives—immigrants and natives—are quite similar. Therefore, the application of the log-rank test (Mantel-Cox method) rejects the equality of survival distribution between native and immigrant workers both for paid employment and for self-employment.

Figure 2. Survivor functions for self-employed workers, by nationality



Source: MTAS (2008).

Figure 3. Survivor functions for salaried workers, by nationality



Source: MTAS (2008).

### 3. Econometric specifications and results

This section is structured as follows. The first subsection uses the econometric model to estimate the probability of being self-employed compared to the probability of being a paid employee. It then estimates the model and interprets the results. The second subsection studies the dynamics of self-employment and paid employment by the estimated duration models.

#### 3.1. Probability of being self-employed compared to being a paid employee

An individual decides whether or not to become self-employed in his first job by 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 is possible to define the following index function:

$$(3) \quad I_i = R_i - W_i = X_i' \beta + \varepsilon_i,$$

where  $X_i$  is the set of exogenous variables that collects personal and socioeconomic characteristics that affect  $R_i$  and  $W_i$ . The regressors considered are nationality, age, gender, education, sector of activity, Spanish region of residence, and GDP growth rate in the year in which the employment relationship is initiated.  $\beta$  is a vector of unknown parameters, and  $\varepsilon_i$  is a disturbance term that follows a logistic distribution.  $I_i$  is not observed but is known if a person is either self-employed or a paid employee, so a binary variable  $Y_i$  is constructed taking the value of 1 if the individual is employed and 0 otherwise. Then the probability of self-employment is given by the following:

$$(4) \quad \text{Prob}(Y_i = 1) = \text{Prob}(I_i > 0) = F(X_i' \beta),$$

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 a paid employee.

Second, the worker's age has a positive influence on the probability of being self-employed. This may occur because the ability to conduct a business can be correlated positively with the age of the individual. With respect to gender, the odds of being self-employed compared to being a salaried worker is 1.2 times higher for male than for female workers. Self-employment often requires a heavy workload. This may make it more difficult for women to be self-employed, especially for those with young children (Andersson and Wandensjo, 2008).<sup>2</sup> Another important conclusion is that educational attainment and self-

---

<sup>2</sup> Unfortunately, the data set used in this paper does not include family characteristics, such as marital status or number of children that could be important in this issue.

employment are positively correlated. In particular, men and women with more than a primary education have a probability of being self-employed that is 1.8 and 3.7 percentage points higher, respectively, than men and women without a primary education. On the one hand, this result is opposite that of the filter theory (Arrow, 1973). This theory proposes the signaling hypothesis, which implies that those planning to enter self-employment have no need to acquire formal qualifications to indicate their quality to potential employers. On the other hand, Lucas (1978) finds that more able and creative individuals become entrepreneurs, and the rest become wage workers. This leads to an expectation that the more educated an individual is, the more likely he or she is to become an entrepreneur, as education could serve as a filter such that the more educated tend to be better informed and thus more efficient at assessing self-employment opportunities. Moreover, more educated workers are more likely to obtain external finance such as equity capital and loans for their business (see, for example, Bates, 1990; Cressy, 1996).

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

Variables	Coefficients	Marginal effects
Geographic 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 and retail trade	0.561***	0.030***
Hotels 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	3,730.56***	
Sample size	57,120	

Note: The reference is a Spanish woman without primary education, working in industry, and living in Andalusia.

\*\*\* Significant at 1%, \*\* at 5%, \* at 10%.

**Source: MTAS (2008).**

Concerning job characteristics, on the one hand, self-employed workers are more likely to be employed in the wholesale retail or the financial intermediation sector, with a probability that is 3.0 and 2.4 percentage points higher, respectively, than in the manufacturing sector. On the other hand, they are less likely to be in the education sector (3.4 percentage points lower than in the manufacturing sector). By Spanish regions, workers are less likely to be self-employed in Madrid and Catalonia, which are the most developed regions in Spain, than in Andalusia. In Madrid and Catalonia, the greater opportunities for wage employment discourage self-employment. With regard to business conditions, the GDP growth rate exerts a negative influence on self-employment and is a push factor for paid employment.

### 3.2. Duration of self-employment versus paid employment

This subsection analyzes the joint effect of various personal, job, and labor market characteristics on the probability of exiting from self-employment or paid employment.

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:

$$(5) \quad \lim_{h \rightarrow 0} \frac{\text{prob}(t+h > T_i \geq t / T_i \geq t)}{h} = \lambda_i(t)$$

A widely used way to allow for the presence of observed variables affecting the duration in each state labor market (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 as follows:

$$(5) \quad \lambda_i(t; x_i) = \lambda_0(t) \exp(x_i' \beta),$$

where  $X_i$  is a vector of covariates for individual  $i$ ,  $\beta$  is a vector of unknown coefficients, and  $\lambda_0(t)$  denotes the baseline hazard function. An advantage of this model is that it is 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:

$$(7) \quad \frac{h(t; x_i)}{h(t; x_j)} = e^{(x_i - x_j)\beta},$$

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

$$(8) \quad \beta_k = \partial \ln \lambda(t, X_{ik}) / \partial X_{ik}.$$

Estimations of this model are shown in table 4. The second column presents the estimated coefficients that show the risk of exit for self-employment. First, all immigrant collectives, with the exception of Asian workers, are found to have a higher risk of exit than Spanish workers. In particular, EU workers have a risk that is more than double that of Spanish workers. Second, age is found to have a positive and significant effect on duration. On the one hand, it is possible that this variable is a proxy for the acquisition of nonformal human capital necessary for the good performance of the business. On the other hand, this could also be explained by the job-shopping theory (Johnson, 1978), which states that younger individuals tend to enter riskier occupations. With respect to gender, men have less risk of leaving self-employment than women. By sector of activity, the biggest job instability arises within the education, financial intermediation, construction, and tourism sectors. For the first two sectors, the risk of exit from self-employment is 1.6 times higher than in the sectors not cited. For the last two sectors, it is 1.3 times

higher than in the rest of the sectors. With regard to Spanish regions, little variability is observed in the risk of leaving self-employment, with only Extremadura showing a lower risk of exit (0.6 times lower than the rest). Finally, the economic cycle approximated by the 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	Paid employment Coefficients
Geographic 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 and retail trade	0.082	0.119***
Hotels 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***
Log likelihood ratio test	198.69	1,704.58
Sample size	4,102	53,018

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

\*\*\* Significant at 1%, \*\* at 5%, \* at 10%.

Source: MTAS (2008).

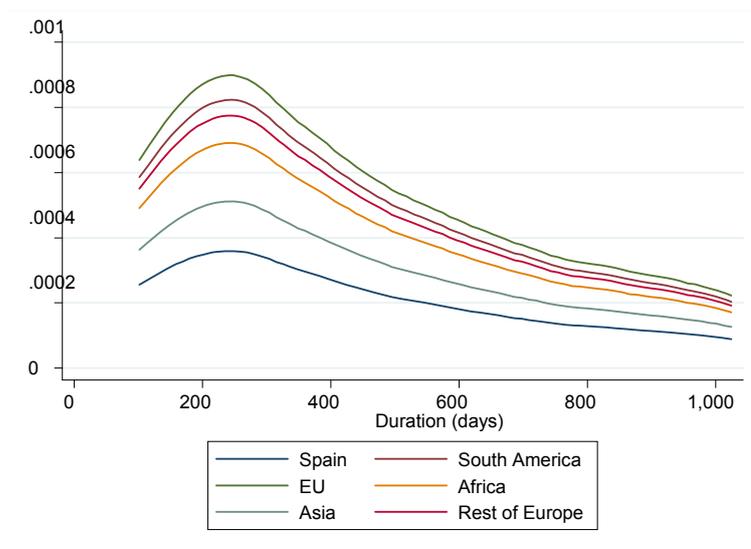
Concerning the duration of paid employment, almost all explanatory variables are relevant. First, Spanish workers once again have a lower risk of leaving their job than do foreign workers. In the case of paid employees, the differences among nationalities are lower—and in a distinct order—than in the case of self-employed workers. On the one hand, Spanish and EU workers have a similar risk of leaving employment. On the other hand, African and South American workers have the highest risk of leaving 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, they influence the risk of exit from paid employment less than the risk of exit from 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 sector of activity, workers in the construction and tourism sectors have the greatest risk of leaving the job. In both sectors, the risk of leaving the job is 30% higher than in the industry sector. By contrast, workers in the financial intermediation sector have the lowest risk of leaving the job. Coefficients relating to dummies representing Spain's regions show that Andalusia, Valencia, Murcia, and La Rioja present a lower risk of leaving employment than the rest of regions. Finally, the influence of the economic cycle on the risk of exit has the same sign for paid employment as for 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 salaried workers. In these graphs, the exit rates from the current job increase to a maximum near the beginning of new employment and decline over time (the bell-shaped pattern). In the initial phase of a new job, it is relatively easy for a worker to leave his job and to search for a better job if the current job reward is lower than expected. But after

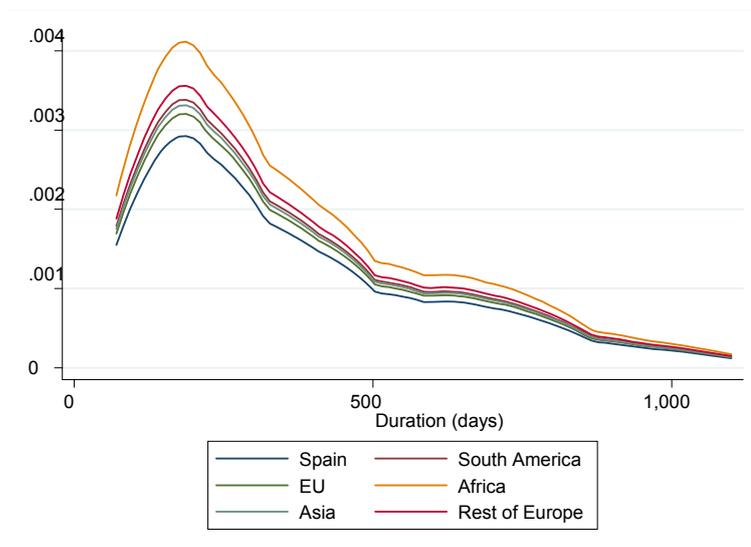
the match has lasted for a while, workers are less likely to leave the job because their experience is a valuable resource associated with current and future reward. In the case of self-employed workers, the uncertainty is prevalent during the first stages of business formation. As time elapses, entrepreneurs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

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



Source: MTAS (2008).

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



Source: MTAS (2008).

The later Cox duration model implies that the exit from self-employment (paid employment) occurs in continuous time and is quite suitable for these data because the spells are measured in days. Nevertheless, one of the assumptions of the continuous-time Cox model is that the inter-individual heterogeneity (differences existing among individuals) can be explained by the set of covariates—that is, by the set of observed characteristics. However, a common issue in estimating duration models is the possibility of allowing 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 on any particular day. In this case, the variable duration in the state of self-employment (paid employment) is grouped into the number of months.<sup>3</sup>

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 as follows:

$$(9) \quad h_j(x_i) = 1 - \exp[\exp(x_i' \beta + \gamma_j + \log(\varepsilon_i))],$$

where  $\varepsilon_i$  is a positive-value random variable with density  $f_\varepsilon(\varepsilon)$  that is controlled by the individual heterogeneity. The choice of  $f$  density for  $\varepsilon$  is then required to enable the calculation of the log likelihood. In this model, it is assumed that  $\varepsilon$  is a Gamma distributed random variable with mean 1 and finite variance  $\theta^4$ ;  $\gamma_j$  is a function describing the duration dependence in the discrete-time hazard. To allow a flexible specification of the baseline hazard, we do not impose any restrictions on how  $\gamma_j$  varies from month to month and then consider a specific

---

<sup>3</sup> Although estimating the discrete-time duration model with group data makes it possible to control unobservable heterogeneity, it is necessary to consider that the measure of employment duration is less precise than in the initial Cox model with spells expressed in days.

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

parameter for each time interval; this implies the introduction in the model of time dummies for each month. The model is estimated via Jenkins “pgmhaz routine” in Stata (Jenkins, 1997).<sup>5</sup>

Table 5 displays the results. First, the variance of the unobservable heterogeneity is significantly different from zero both for exits from self-employment and for exits from paid employment, implying a frailty effect. In particular, this result shows the importance of factors such as the ability, motivation, or social contacts for survival in employment (self-employment and paid employment).

---

<sup>5</sup> 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-empl. Coefficients	Paid empl. Coefficients
<b>Geographic origin</b>		
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***
<b>Age</b>	-0.036***	-0.008***
<b>Gender</b>		
Male	-0.384***	-0.066***
<b>Education</b>		
Primary education	-0.156	-0.057***
More than primary education	-0.165	-0.156***
<b>Activity</b>		
Construction	0.5191**	0.287***
Wholesale and retail trade	0.166	0.135***
Hotels 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***
<b>Spanish regions</b>		
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***
<b>GDP growth rate</b>	-0.056	0.025
<b>Time dummies</b>		
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***

Note: The reference is a Spanish woman without primary education, in the industry sector, and living in Andalusia.

\*\*\* Significant at 1%, \*\* at 5%, \* at 10%.

Source: MTAS (2008).

Another consequence of the presence of unobservable heterogeneity is the existence of bias in the estimated regression coefficients from Cox models. Thus we first observe that the positive difference between the risk of exit from self-employment or paid employment for foreigner workers, compared with Spanish entrepreneurs, increases with respect to the estimates of the Cox model.

In particular, in exiting from self-employment, EU workers have a risk of exit that triples that of Spaniards. In contrast, workers from South America and the rest of Europe have a risk of exit that more than doubles the risk for Spanish workers. In the case of paid employment, the risk of exit is higher for foreign workers than for Spaniards, although the difference between workers from Spain and workers from the other countries diminishes in relation to the difference for self-employment. African workers show the highest risk (one and a half times higher than the corresponding risk for Spanish employees). On the contrary, European Union and Asian workers have a risk of exit only 8% higher than Spaniards.

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

In relation to activity dummy variables, the greatest difference with respect to the Cox model appears in the risk of exit from self-employment. In particular, the estimated coefficients report an increase in the risk for self-

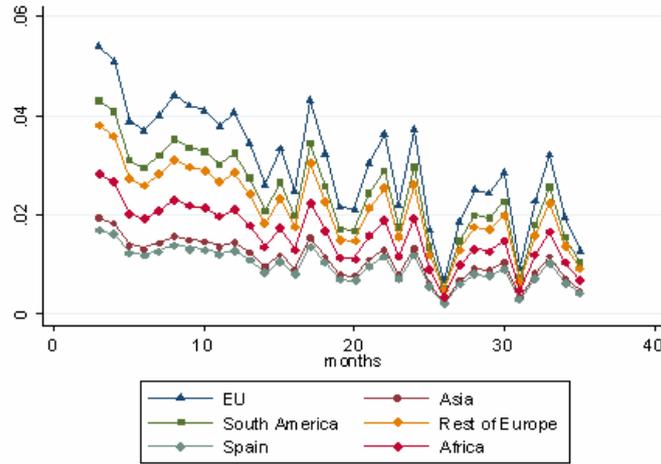
employed workers in education, financial intermediation, construction, and hotels and restaurants with respect to that for workers in the industry sector.

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

Regarding 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 may occur because the GDP growth rate is correlated with the set of monthly dummy variables introduced in the model to capture duration dependence.

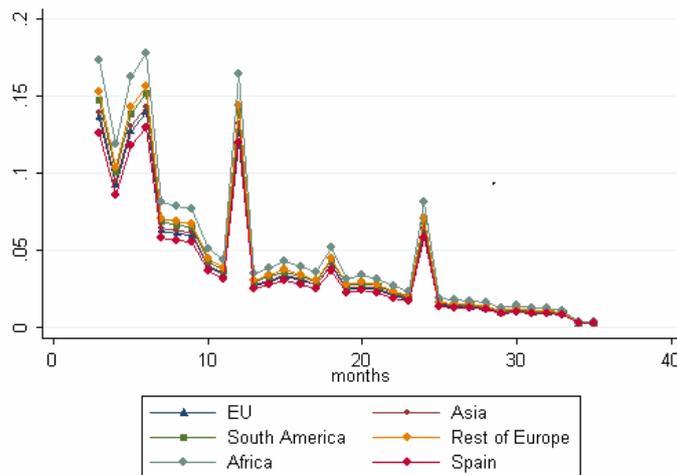
With regard to the coefficients attached to the monthly dummy variables, in the case of paid employment, these coefficient show a decreasing pattern with respect to the risk evident between the third and fourth month. In the model that explains the duration of self-employment, the behavior of the set of coefficients attached to dummies that capture duration dependence is not as homogeneous, although lower risks of exit from self-employment appear in the first two years, with the exception of some months. Finally, figures 5 and 6 plot the predicted hazard functions evaluated at mean values of covariates, for self-employed and salaried workers. In these graphs, the predicted risk in all periods

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



Source: MTAS (2008).

Figure 6. Predicted risk of exit from paid employment (discrete-time model with unobservable heterogeneity)



Source: MTAS (2008).

is lower for Spanish workers than for the rest of the national groups. Moreover, in the case of paid employment, two highlighted peaks are detected in the twelfth and twenty-fourth months, probably because in the Spanish labor market a large proportion of fixed-term contracts end in these two periods.

#### 4. Conclusions

This paper has investigated two issues. First, it has analyzed the determinants of the individual choice between self-employment and paid employment as the first job. Second, it has studied the factors influencing the duration in employment in both labor market states. Data come from the Muestra Continua de Vidas Laborales. These data are especially useful for the second objective because their panel structure provides an exact measure of the duration of employment.

One of the main results of this paper is that the immigrants are a disadvantaged group with regard to entry into self-employment as a first job. Moreover, once they enter self-employment, they are more likely than native Spaniards to exit from it. Given the positive effects of self-employment on economic growth and job creation, this result should be of interest to policy makers seeking to design economic policies that promote the assimilation of immigrants into the Spanish labor market. Immigrants are not a homogeneous collective, as the probability of being self-employed is lower by 5.8 and 4.1 percentage points for South American and African workers, respectively, than the probability for Spanish workers, while for EU and Asian workers it is lower only by 2.5 percentage points. Moreover, EU workers have a risk of exit that triples that of Spanish workers, whereas Asian workers have the same risk of exit as Spanish ones.

Other interesting results about the probability of being self-employed are, on the one hand, that older, male workers show a higher probability of being entrepreneurs than younger, female workers. On the other hand, the educational level of the individual is correlated positively with the probability of being self-employed; this finding does not support the filter theory (Arrow, 1973), which proposes that education is not a relevant variable for entry into self-employment.

Regarding survival analysis in the first job, results show that relationships are more stable for self-employed workers than for salaried workers. Moreover, in both labor market states (self-employment and paid employment),

unobservable heterogeneity is relevant to explaining the duration in the first job; that is, motivation, ability, and social contacts are important elements in the ability 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* 26 (3): 189–204.
- ANDERSSON, P., and WANDENSJO, E. (2008): "A gender perspective on self-employment entry and performance as self-employed." IZA Discussion Paper 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): 267–315.
- ARROW, K. (1973): "Higher education as a filter." *Journal of Public Economics* 2 (3): 193–216.
- BATES, T. (1990): "Entrepreneur human capital inputs and small business longevity." *Review of Economics and Statistics* 72 (4): 551–59.
- BECKER, G. (1962): "Investment in human capital: A theoretical analysis." *Journal of Political Economy* 70 (5): 9–49.
- BLACK, S., and LYNCH, L. (1996): "Human capital investment and productivity." *American Economic Review* 86 (2): 263–67.
- BLANCHFLOWER, D. G., LEVINE, P. B., and ZIMMERMANN, D. J. (1998): "Discrimination in the small business credit market." NBER Working Paper n° 68440. National Bureau of Economic Research, Cambridge, MA.
- BLANCHFLOWER, D. G., and OSWALD, A. J. (1998): "What makes an entrepreneur?" *Journal of Labour Economics* 16 (1): 26–60.
- BLAU, F. D., and GRAHAM, J. W. (1999): "Black-white difference in wealth and asset composition." *Quarterly Journal of Economics* 105 (2): 321–39.
- CARRASCO, R. (1999): "Transitions to and from self-employment in Spain: An empirical analysis." *Oxford Bulletin of Economics and Statistics* 61 (3): 315–41.
- COX, D. (1972): "Regression models and life-tables." *Journal of the Royal Statistical Society* 34 (May/August): 187–220.

- CRESSY, R. (1996): "Are business start-ups debt-rationed?" *Economic Journal* 106: 1253–70.
- EUROSTAT. (2006): News Release nº 133.
- EVANS, D., and LEIGHTON, L. (1989): "Some empirical aspects of entrepreneurship." *American Economic Review* 79: 519–35.
- GEORGELLIS, Y., SESSIONS, J., and TSITSIANIS, N. (2005): "Self-employment longitudinal dynamics: A review literature." *Economic Issues* 10 (2): 51–84.
- INE. (1985): Working and Living Conditions Survey.
- . (1985–94): Spanish Continuous Family Expenditures Survey.
- . (2008): Spanish Labor Force Survey.
- JENKINS, S. P. (1997): "Discrete-time proportional hazard regression." *Stata Technical Bulletin* 39: 22–32.
- JOHNSON, W. R. (1978): "A theory of job shopping." *Quarterly Journal of Economics* 92: 261–78.
- JOHNSON, W. R. (1978): "
- JOVANOVIC, B. (1982): "The selection and evolution of industry." *Econometrica* 50: 649–70.
- LUCAS, R. E. (1978): "On the size distribution of business firms." *Bell Journal of Economics* 9: 508–23.
- MEYER, B. (1990): "Unemployment insurance and unemployment spells." *Econometrica* 58: 757–82.
- MTAS. (2008): Spanish Social Security Survey [Muestra Continua de Vidas Laborales].
- OECD. (2000): *Employment Outlook*. Paris: OECD (September).
- THURIK, A., et al. (2008): "Does self-employment reduce unemployment?" *Journal of Business Venturing* (forthcoming).