Title	Time to the doctorate and research career: some evidence from Spain
Journal	Research in Higher Education
Manuscript Type	Research Paper
Author	Antonio Caparrós Ruiz (University of Málaga)
Abstract	Education and research are considered as the cornerstones of the economic growth and the job creation for the Lisbon Strategy proposed by the European Union. Therefore, understanding the transmission channels of the educational investments to the society is important to inform policymakers and students about the benefits and opportunities associated with the acquisition of human capital. In this context, PhD programs play a significant role to reach the European research goals. The current study contributes to shed empirical evidence about the determinants of the time to the doctorate in Spain and its influence on the probability of carrying out an innovate activity (for example, working as a researcher), in both cases the PhD program's academic field is included as regressors. One of the main hypothesis to verify is whether a prolonged time to complete the doctoral studies is a negative signal about the individual's capacity to develop research skills. If this is the case, longer time to doctorate would imply less probability of working as a researcher. The methodology applied consists in estimating a Cox model to analyse the determinants of the time to the doctorate, and a probit model to examine the probability of being a researcher considering time to the doctorate as an endogenous regressor. Data used in this study come from the 2009 Survey on Human Resources in Science and Technology, provided by the Spanish National Institute of Statistics (INE in 2009 survey on human resources in science and technology, INE, Madrid, 2010).
Keywords	PhD program, Research career, Endogenous regressor
Acknowledgments	This is a post-peer-review, pre-copyedit version of an article published in the review Research in Higher Education (Vol. 60(1), pp. 111-133). The final authenticated version is available online at: "https://doi.org/10.1007/s11162-018-9506-2".

Time to the doctorate and research career: some evidence from Spain

Introduction

Measuring the socio-economic impact of financial and human resources on research may assist and encourage policymakers to implement strategies to improve the cost-effective use of public education expenditure. On the other hand, understanding the transmission channels of educational investments to society may inform students about the benefits, returns and opportunities associated with the acquisition of human capital that could help them in their educational choices. In Spain, public education spending represents a significant proportion of the Gross Domestic Product (GDP) and of government expenditure, with percentages that exceed 4.5% and 11%, respectively, during the period 2000-2013 (World Bank 2013). Moreover, expenditure on higher education (tertiary education) represents 25% of Spanish education spending and stands at more than 1% of GDP (1.3% in 2010), which is quite similar to the percentage observed in European Union (EU) countries (1.4% in 2010), although lower than the average rate (1.6% in 2010) for the countries of the Organisation for Economic Cooperation and Development (OECD 2013).

The analysis of tertiary education effects on students' professional careers can be a major source of information, shedding light on its costs and benefits and helping individuals with the process of decision-making about their future investments in human capital. In that respect, the research literature on educational and economic topics has documented the positive impact of higher education on labour markets from different perspectives. For example, most studies (e.g. Altonji *et al*. 2012) prove that the wages received by individuals with higher education outweigh the costs of the degree (tuition plus earnings lost while attending school). On the other hand, more highly educated individuals have more opportunities of finding jobs and experiencing upward labour market mobility (e.g. McCormick 1997). Moreover, some empirical research has verified that higher education is a mechanism to achieve upward intergenerational occupational mobility (e.g. Caparrós 2016). Higher education also plays an important role in the development of lifelong learning strategies, since skilled workers are more able and motivated to finance or to receive training, with the corresponding benefits for their productivity and the competitiveness of their firms (Brooks and Everett 2008). However, there are some drawbacks that must be taken into account when assessing the effectiveness and benefits of higher education. For example, there is a rising skills gap between higher education and the labour market. Employers are having difficulties finding applicants to perform qualified tasks, which generates a problem of credential inflation; that is, there is a mismatch between the knowledge acquired in the educational system and the productive system's needs, causing a high rate of over-education among highly educated people (Walker and Zhu 2008). One explanation of this credential inflation could be the existence of asymmetric information, which causes individuals to have problems identifying and communicating their transferable and generic skills to potential employers. For example, in the case of the PhD holders, they have difficulties reflecting the value of their PhD program and demonstrating the transferability of their competences and academic experiences to employers (Cryer 1998).

From a general socio-economic point of view, there is no doubt that higher education can play an important role in the current knowledge society due to the increasing demand of high skills generated by the productive system. In addition, note should be made of the rising use of information and communication technologies (ICT) in the economy, which represents a powerful boost for economic growth: for example, ICT explain 25% of EU GDP growth and 40% of productivity growth, according to the European Commission (European Commission 2005). In this context, the completion of a PhD program represents a qualitative impulse to acquire the specialization and necessary skills to be employed as highly-qualified workers, to manage complex professional tasks or even to become competent researchers in the different academic fields. This way, the knowledge acquired in PhD programs can help fill the skills gap between academic degrees and labour market needs. Moreover, it should be noted that PhD programs have acquired more relevance within the Bologna Process framework, and are considered as a crucial element to reach the European research goals. In particular, according to the Bologna Process framework, PhD programs should play an essential role to promote and sustain knowledge-based society through the generation and transmission of R&D knowledge. Furthermore, PhD programs should promote interdisciplinary training, contribute to the development of skills that can be transferable to the labour market, and train PhD students as early stage researchers. Definitely, PhD programs should be the link between the European Higher Education Area (EHEA) and the European Research Area (ERA). In this sense, it is worth highlighting the role of the ERA to support EU funding of doctoral training, even with specific programs such as the Marie Sklodowska Curie Actions (MSCA) that invests directly in PhD's (European Commission 2016).

The educational community is aware of this new reality, with the increase in participation in PhD programs; for example, the number of doctoral degrees rose 38% during the past decade in the OECD countries. Moreover, this has been motivated by the long-standing increase of investment efforts in R&D on the part of developed countries, generating an increase of up to 21% in the number of researchers since 2007. This effort has been particularly significant in EU countries, which are aiming to invest 3% of their GDP in R&D by 2020 and where 25% of the world's researchers are working (UNESCO 2015).

Focusing on the Spanish labour market, in 2009 doctorates totalled 1%, which was still far from the figures registered in other European countries (for example, in Switzerland, doctorate holders total 3.4% of the population). It is interesting to note that the influence of PhD programs on workers' careers is quite positive, as workers with a PhD enjoy a privileged situation in relation to other employees. For example, statistical evidence for 2016 reveals that the unemployment and inactivity rates for doctorate holders were around 5%, whereas these rates reached 20% for the population as a whole (OECD 2018). On the other hand, there is a wide wage differential between doctorate holders and the rest of wage earners (for example, they earn 60% more than workers with only primary education). However, in spite of their importance, to date the careers of doctorate holders in the Spanish labour market have not received much attention from economic researchers, probably due to the lack of suitable databases containing information to address this topic. Nonetheless, it is possible to highlight, for example, the articles of Canal (2013), Canal and Wall (2014) or Canal and Rodríguez (2016) that use data from the 2006 Survey on Human Resources in Science and Technology (INE 2007). Canal (2013) examines the job satisfaction determinants of employees with doctoral degrees, finding that Spanish doctorate holders' job satisfaction is located at medium-high levels. Canal and Wall (2014) analyse the factors that affect PhDs' career success and conclude that job satisfaction is the variable with the greatest influence: in particular, aspects related to satisfaction with job promotion or the degree of responsibility. Finally, Canal and Rodríguez

(2016) assess how the type of doctoral training has an influence on the time to obtain a suitable job. Their results show that men are more likely to find a suitable job faster, and that this depends on their academic field.

This paper aims to contribute to the previous debate, providing some empirical evidence regarding the determinants of the time to the doctorate in Spain and its influence on the probability of carrying out an innovative activity (for example, working as a researcher). In both cases the area of knowledge of the PhD program is included as a regressor. To the best of our knowledge, these topics are unprecedented in the literature on Spain and across Europe and are associated with interesting issues for students and agents responsible for the educational system. In particular, a prolonged time to complete the PhD program may have negative effects for individuals and society. On the one hand, it can be associated with higher institutional costs, fewer career prospects, fewer years of productive work-life or even with loss of job promotion opportunities if the teaching and research career is pursued at a university (Evangelauf 1989). On the other hand, knowledge about delays in the completion of the dissertation can discourage other students to access PhD programs, which would have negative long-term consequences on scientific and technological development and competitiveness. This way, this study can help understand the factors encouraging the success to achieve the doctorate and a career as researcher.

Data used in this study was obtained from the 2009 Survey on Human Resources in Science and Technology, provided by the Spanish National Institute of Statistics (INE 2010), which offers some advantages for our research since it focuses on doctorates (classified by field of knowledge) at Spanish universities in general from 1990 to 2009 and provides information about the labour situation of holders in 2009 (which allows us to know if they are working as researchers). This survey was carried out before the implementation of the Bologna Process and contains the most recent information to analyse the issues previously mentioned. The Bologna Process implied significant legislative changes of the process to obtain a doctoral degree, aiming to increase the professional insertion of doctorate holders in the productive system. In Spain, the current maximum duration of PhD programs depends on whether the student is enrolled full-time or part-time (3 and 5 years, respectively), although it is possible to request a time extension. Moreover, the students' scientific production during the PhD program is supervised by committees. The changes introduced by the Bologna Process tried to push student performance and to speed up completion of doctoral studies. This paper supports with empirical evidence the potential effectiveness of these legislative changes, verifying whether there is a negative relationship between the time to the doctorate and the probability of working as researcher.

The remainder of the paper is organised as follows. Section 2 outlines the research design showing the data set and the variables used in the analysis. Sections 3 and 4 present the econometric framework and the estimation results. Section 5 provides some concluding remarks and, finally, there is an appendix with some statistical results.

Research design

The lack of information in Europe about the professional situation of PhD holders prompted the EU (regulation 753/2004) to implement harmonised statistics about science, technology and innovation through the development of community surveys concerning doctorates. These statistics were aimed at supporting EU actions and policies on human resources in the context of the Lisbon Strategy, in which research is the cornerstone of economic growth and job creation. As a result of this initiative, the Spanish Office for National Statistics (INE) conducted the 2009 Survey on Human Resources in Science and Technology (INE 2010). The primary objective of this survey was to quantify the research capacity of doctorate holders, to learn about their professional activity and to follow their national or international migratory trajectories. For the purposes of this study, this data source is appropriate because it contains information about personal and employment characteristics of doctorate holders residing in Spain who obtained their PhD qualification between 1990 and 2009 from a public or private Spanish university. In addition, there is information about the field of knowledge and the duration of the PhD program which is required to attain the study goals.

The number of individuals who were surveyed totals 4,123, but the sample used in the analysis consists of 4,103 individuals, once the anomalous values related to the time to doctorate have been removed (for example, those individuals who provide unrealistic answers regarding the duration of the PhD program, such as having completed the PhD program in 6 months). The selected individuals are classified in the following research fields: Natural Sciences, Engineering and Technology, Medical Sciences, Agricultural Sciences, Social Sciences and Humanities. The distribution of these individuals according to the area of knowledge and the summary of statistics of the time to the doctorate appear in Table 1.

[Insert Table 1]

The most representative scientific field in the sample is Natural Sciences, with 33% of the doctorate holders, followed by Medical Sciences and Social Sciences with percentages around 20%. On the contrary, the areas of knowledge with the smallest proportion of doctorate holders are Engineering and Technology and Agriculture Sciences, with percentages that do not exceed 1%. In respect to the summary of statistics corresponding to the time to the doctorate, there are also important differences by scientific discipline. In particular, it is worth mentioning the presence of two groups, one comprising Social Sciences and Humanities and the other academic fields, since the first group has a longer time to doctorate than the second one.

In order to obtain an idea of the duration of doctoral programs, the Kaplan-Meier (KM) estimator is used to obtain a non-parametric estimation of the survivor functions. The KM estimate of the survivor function is given by the following expression:

$$\hat{S}_t = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{1}$$

where d_i is the number of individuals who completed a doctorate during period t_i and n_i is the number of individuals at risk at the beginning of t_i. One of the advantages of the KM estimator is that it allows obtaining the probability that a student will complete the doctoral thesis at time t conditional to the time in the PhD program up to this moment. Figure 1 plots the KM estimates, distinguishing the survive functions according to the academic field.

[Insert Fig.1]

A look at the graphs corroborates the previous descriptive results concerning the time to the doctorate, since the PhD programs in Social Sciences and Humanities have higher survival rates than the other academic fields.

To statistically verify the different patterns observed among the survival functions, the log-rank test (Mantel-Cox method), whose statistics follow a distribution $X^2_{(1)}$, is applied and proves the null hypothesis of no differences between two survival curves (Lawless 1982). Table 2 displays the results associated with this test when comparing two survival curves for each pair of academic fields. The results obtained verify, on the one hand, that the patterns of time to the doctorate in the fields of Social Sciences and Humanities are statistically different than the other areas of knowledge. On the other hand, Natural Sciences, Engineering Technology and Agriculture Sciences have KM survival curves that are statistically similar, but they differ from the one corresponding to Medical Sciences. Moreover, the long-rank test also reveals differences between the survival functions associated with Social Sciences and Humanities.

[Insert Table 2]

From a multivariate point of view, factors affecting the completion of the doctorate have been analysed by the literature with different methodological statistical approaches (e.g. Seagram *et al.* 1998). Other studies have explored the temporal dimension of completion of the doctorate (that is, the timing of the event) rather than whether the students complete or not their doctoral studies (e.g. Sheridan and Pyke 1994; Ehrenberg and Mavros 1998; Siegfried and Stock 2001; Wao 2010). This paper is included in this branch of the literature, since it analyses how the time to the doctorate is affected by a set of explanatory variables. The selection of the regressors is based on the theoretical foundations of those studies, examining the evolution of the educational process and, in particular, the time to the doctorate. Thus, the set of explanatory variables is composed of two subgroups of regressors. The first one contains personal and family characteristics such as gender, parents' educational attainment or the starting age of the doctoral studies. Thus, the inclusion of gender is aimed at capturing the possible existence of inequalities or the existence of gender-role stereotypes (e.g. Sheridan and Pyke 1994 or King 2008). Parents' educational attainment is derived from the conceptual framework postulated by the Integration Theory (Tinto 1975). According to this theory, family background variables are essential to the social and academic integration of students, since they shape their motivation and self-determination in order to fulfil the commitments acquired with the education system. Finally, the PhD program starting age is a proxy of human capital accumulation and the acquisition of competences between the end of the undergraduate degree studies and the beginning of the doctoral studies. Moreover, the starting age shows the time horizon of the returns associated with educational investment, and implies that older individuals are more motivated to complete their doctoral studies. The second subgroup of regressors are related to the characteristics of the PhD program: the academic field (that is, the knowledge areas mentioned above), the types of research conducted during the PhD program (non-oriented research or not, applied research or not, experimental development or not) and the doctoral studies' sources of financing (Spanish grant, foreign grant, earned income working as research assistant or not, loans or savings). With the inclusion of the knowledge area and the type of research as regressors, we are trying to capture the influence of institutional factors, which has been verified by the literature on this topic (e.g. Golde 2005). In relation to financial support, the empirical research has verified its relevance to boost completion rates and to decrease the time to the doctorate (Ehrenberg and Mavros 1995). Enrolling in a PhD program can be an option that is riskier than entering the labour market following completion of the undergraduate degree so financial aid such as grants or earned income working as research assistant contribute to diminish the uncertainty and opportunity costs of starting a PhD program. Furthermore, the type of support could also affect the time to the doctorate: for example, students with a grant have more available time to advance in their dissertation research than those working as research assistants. On the other hand, grants are assigned according to student performance, so it is expected that the recipients of fellowships complete their PhD studies in less time.

The descriptive statistics corresponding to the set of covariates are displayed in Table 3. The summary statistics reveal that the average starting age of a PhD program is around 29, which is 2 years higher than the average age corresponding to completion of the undergraduate degree in Spain (OECD 2014). This fact reveals the importance of finishing doctoral studies within a reasonable time to avoid a late entry into the labour market. Second, the proportion of male PhDs is slightly higher than female (55% versus 45%), in contrast to the gender distribution in undergraduate degree studies, where the percentage of female students rose to 60% in 2010. Concerning family background, the main result is that the percentage of parents with at most primary studies is overwhelming (49% for fathers and 68% for mothers), whereas individuals with a PhD father or mother are a minority (only 7% and

2%, respectively). Third, regarding the type of research, it is worth mentioning that most individuals are engaged in non-oriented or applied research (72% and 63%, respectively). On the contrary, the proportion of individuals in a PhD program that includes experimental research is only 37%. In relation to the ways to finance doctoral studies, the most common is a Spanish grant (38%) or working as an assistant researcher (24%). These results show the important role of the public administrations in providing financial support to the PhD programs, since the percentages of students that finance their studies with loans, savings or working as research assistants are less representative.

[Insert Table 3]

The second objective of this study is to detect the factors determining the probability that doctorate holders are working as researchers or not at the end of the period observed in the survey (year 2009). With this approach, it is possible to approximate the effectiveness of PhD programs to boost aspects related to innovation in the productive system. That being said, the analysis focuses on wage earners, with a sample of 3,706 individuals (62% employed as researchers). The group of regressors used to explain the probability of working as a researcher or not comprises the following variables: gender, age, area of knowledge of the PhD program, type of research during the doctorate studies, time to the doctorate, labour market status (type of contract and working day), economic sector, and regional dummy variables. First, in relation to personal characteristics, the empirical research has verified that gender is a potential source of bias for PhD holders, since women have more difficulties accessing research resources and pursuing a scientific career (e.g. Bormann and Enders 2004). We seek to verify whether this gender inequality prevails when focusing on the group of doctorate recipients working as researchers. On the other hand, the inclusion of the variable age tries to capture how the PhD holders' preferences to follow a research career evolve with their life cycle (e.g. Thune 2009). Second, the area of knowledge and the type of research are included to approach the career prospects associated with the scientific field, which has been found to be a relevant factor in studies by Fox and Stephan (2001), Robin and Cahuzac (2003), Lee et al. (2010) and Canal and Rodríguez (2013). Fox and Stephan (2001) show that the preferences of PhD recipients and their prospects for scientific careers differ by field of science. However, they verify that there is a gap between prospects and realities, since the differences based on the area of knowledge diminish when the individuals report on their actual employment and wages. Robin and Cahuzac (2003) determine that post-doctoral positions held by PhD holders (defined as temporary research stays in a foreign country) represent a stepping stone toward permanent employment. Lee et al. (2010) observe that the competences and skills associated with each area of knowledge of a PhD program are valued differently for each occupation in the labour market. Finally, Canal and Rodríguez (2013) analyse wage differences among employees with doctoral studies and observe that PhD holders in the field of sciences achieve higher earnings than their counterparts in the knowledge area of Humanities and Social Sciences.

Regarding the selection of the time to the doctorate as a regressor, its theoretical foundation is based on the predictions of the Screening Theory (Spence 1973). According to this theory, the completion of a PhD program in less time would be a positive information signal for employers about the quality of workers. First of all, with respect to job characteristics, the type of contract and the working day are included to control how stable the job relationship is. Second, the economic sector assesses how the demand for PhD holders as researchers depends on the industrial structure and, in consequence, on its R&D activities (García-Quevedo *et al*. 2012). Finally, as usual in the labour market literature, the regional dummy variables are included to take account of the peculiarities of each regional labour market (e.g. Caparrós 2014).

The summary statistics corresponding to these variables appear in Table 4, which allows us to highlight some conclusions. First, male workers are more represented within the group of researchers, with a proportion of 57%, which contrasts with the 43% observed for their female counterparts. Second, by area of knowledge, the group of individuals with a doctorate in Natural Sciences or Social Sciences stand at 62% within the group of researchers, whereas this proportion is only 42% in the other group. It is also worth noting that 35% of individuals not working as researchers are doctorate in Medical Sciences (these professionals only represent 12% of the research workers). On the other hand, doctorate holders in Humanities or Agricultural Sciences have approximately the same percentage in the two groups considered (around 13% and 3%, respectively). Third, there are no significant differences according to the type of research; only the percentage of individuals with non-oriented research is higher in the group of research workers than in the other group (74% versus 70%). In relation to employment characteristics, it is noted, first of all, that the rate of open-ended full contracts is above 75% in all cases, although the presence of part-time contracts is 4 percentage points higher for the doctorate holders working as non researchers (8% versus 4%). Second, it is remarkable that most researchers are employed by higher education (61%), whereas only 9% of them have a job in the private sector (firm or private non-profit institution). This is consistent with the statistics from other European countries (e.g. Auriol *et al*., 2013) and could show that the training received by the doctorate holders is mainly oriented to the academia. This fact is a challenge for the European doctoral funding policy; thus, for example, providing transferable skills that enhances the mobility of doctorate holders between the academic and non-academic sectors is an objective included in the EC Innovate Doctoral Training Principles (European Commission 2011) and in the Marie Sklodowska-Curie Actions (European Commission 2016). On the other hand, 60% of doctorate holders not working as researchers have a labour relationship with the public administration, whereas the proportion of individuals with jobs in the private sector rises to 20%. Finally, the most represented regions in the sample are the East, Madrid and the South, with percentages of about 20%, although there are some difference when splitting doctorate holders between those working as researchers and those not working as researchers. For example, in the South region, there is a difference of 4 percentage points between the group of individuals who are not pursuing a research career (20% versus 16%, respectively).

[Insert Table 4]

Econometric specification

This section focuses on developing the econometric methodology used to reach the proposed objectives. The first objective of this article is to analyse how the time to the doctorate depends on personal and family characteristics and on aspects related to the type of PhD program. Therefore, the most suitable methodology to achieve this goal is the specification and estimation of hazard-based duration models that analyse the likelihood of ending an activity given the time spent on it. In other words, these models control the occurrence of an event and its timing as well as estimating the explanatory variables' effects on duration of the activity (in this case, time to the doctorate).

In this setting, the distribution of the hazard may be assumed as parametric, but there are some shortcomings when using parametric forms: for example, the lack of theoretical support to propose a particular function, which could induce the formulation of an incorrect parametric form whose estimation would be inconsistent. To avoid these problems, it is usual to follow a semi-parametric approach such as the Cox proportional hazard model (Cox 1972) whose hazard is assumed to be:

$$h_i(t \mid X_i) = h_0(t)\exp(X_i'\beta) \tag{2}$$

where $h_0(t)$ is the non-parametric part of the model and $X_i'\beta$ is the parametric component. X_i is a vector containing characteristics observed for the individual, and β is a vector containing unknown parameters. Each β_j coefficient of the β vector represents the increment in log hazard associated with a 1 unit increase in X_j; therefore, $(e^{\beta_j}-1)*100$ can be interpreted as the percentage variation in hazard with one unit increase in X_j. The unknown parameters β_j are estimated by maximizing the partial log-likelihood given by:

$$\log L = \sum_{j=1}^{D}\left[\sum_{i\in D_j} X_i'\beta - d_j \log\left\{\sum_{k\in R_j}\exp(X_k'\beta\right\}\right] \tag{3}$$

where D is the set of occurrences, D_j represents the set of all individuals completing the PhD program at the same time t_j, d_j contains the total number of events at t_j and R_j includes the individuals who are at risk at time t_j. The formulation of the partial log-likelihood depends on how the ties are treated; that is, those individuals who report the same time to the doctorate. In this case, equation 2 is expressed following the Breslow approximation (Breslow 1974).

An extension of the Cox model is to consider the presence of shared frailty within a subgroup of observations. In particular, shared frailty is an unobservable variable or latent group-specific random effect that is identical for the individuals within the same group but differs among subgroups. If shared frailty is denoted by v_j and it is assumed that it enters multiplicatively on the hazard function, the Cox regression with shared frailty is expressed as:

$$h_{i,j}(t \mid X_i) = h_0(t)\exp(X_{ij}^{'}\beta + v_j) \tag{4}$$

j indexes the subgroups generated with the sample. In this study, the criterion for grouping individuals has been the time elapsed between the end of the university studies and the start of the PhD program (less than a 1 year, from 1 to 3 years, and more than 3 years). This way, v_j would capture the unobservable and omitted characteristics that could affect students' capacity to complete the PhD program. These unobservable factors could be related to the motivation to undertake the doctoral studies (professional reasons or no reasons), the obsolescence of the knowledge acquired in the undergraduate degree studies (becoming an obstacle to finishing the doctoral studies) or the influence of other activities carried out by the individuals during the time gap between the university studies and the PhD program. The estimates for the Cox model with shared frailty are computed by maximizing the following log-likelihood function:

$$LogL = \log L_{Cox}(\beta, v_1, v_2, v_3) + \sum_{m=1}^{3} \left[\begin{array}{l} \frac{1}{\vartheta}\{v_m - \exp(v_m)\} + (\frac{1}{\theta} + D_m)\left\{1 - \log(\frac{1}{\vartheta} + D_m)\right\} \\ -\frac{\log\theta}{\theta} + \log\Gamma(\frac{1}{\vartheta} + D_m) - \log\Gamma(\frac{1}{\vartheta}) \end{array} \right] \tag{5}$$

where $\log L_{Cox}(\beta, v_1, v_2, v_3)$ is the standard Cox partial log likelihood, and D_m is the number of occurrences in group m.

The second objective of this study is to analyse the determinants of performing an innovative professional activity (working as a researcher) once individuals have completed their PhD program. One of the main issues of interest is the inclusion of the time elapsed until the doctorate as an endogenous continuous regressor. In particular, this allows us to verify whether prolonged time to complete doctoral studies could indicate to the labour market negative signals about the individual's capacity to develop research skills. If this is really the case, longer time to doctorate would imply less probability of working as a researcher. Obviously, to verify this assumption, it is also necessary to include the scientific fields of the PhD program as determining factors of the research activity. This way, the differences in the duration of the doctoral studies previously observed for each area in the descriptive analysis would be controlled in the econometric analysis.

The formulation of the structural model necessary to reach the second aim of this study is as follows:

$$y_{1i}^{*} = \alpha y_{2i} + z_{1i}^{'}\gamma_1 + u_{1i} \tag{6}$$

$$y_{2i} = z_{2i}^{'}\gamma_2 + u_{2i} \tag{7}$$

y_{1i}^{*} is the unobserved propensity of performing innovative tasks on the job. It is assumed that it depends on the endogenous continuous regressor y_{2t} that measures the duration of a PhD program until completing the doctoral thesis, and on a set of exogenous regressors denoted by the vector z_{1t} that includes personal and labour characteristics. z_{2t} is the vector containing the covariates that explains y_{2t} and is formed by a subset of z_{1t} and by instrumental variables. In particular, the dummy variables generated to capture the fathers' and mothers' educational levels are used to instrument y_{2t} in equation (6). Initially, these variables can be appropriate instruments as the family characteristics are related to the time to doctorate, but they are not associated with the individual's own unobserved skills. Anyway, their validity as instrumental variables will be tested using the Amemiya-Lee-Newey test (Amemiya 1978; Newey 1987; Lee 1992). The error terms u_{1i} and u_{2i} are supposed to be distributed as $N(0,\Sigma)$ where Σ is not a block diagonal matrix. As y_{1i}^{*} is an unobservable variable, it is necessary to define the following indicator variable y_{1i}:

$$y_{1i} = \begin{cases} 0 & if \quad y_{1i}^{*} \leq 0 \\ 1 & if \quad y_{1i}^{*} > 0 \end{cases} \tag{8}$$

that shows if the individuals perform innovative activities in their job. Assuming that the error terms are normally distributed, it leads to the following probit model:

$$\Pr ob(y_{1i} = 1) = \Pr ob(\alpha y_{2i} + z_{1i}^{'}\gamma_1 + u_i > 0) = \Phi(\alpha y_{2i} + z_{1i}^{'}\gamma_1) \tag{9}$$

where Φ is the normal cumulative density. The parameters of the model could be estimated by the usual probit maximum likelihood estimator, but this estimator is inconsistent with endogenous regressors. An estimator that produces efficient estimates of the parameters is Amemiya's Generalized Least Squares (AGLS). Moreover, the

AGLS estimator is useful to test hypotheses about the parameters, as it allows us to obtain consistent estimators of the standard errors (Newey 1987). Once the coefficients are estimated, we obtain the average partial effects (APEs) by taking differences (if the regressor is a dummy variable) or derivates (if the regressor is a continuous variable) from the following expression (Wooldrige 2002):

$$N^{-1}\sum_{i=1}^{N}\Phi(\hat{\alpha}\, y_{2i} + z_{1i}^{'}\hat{\gamma}_1 + \hat{\vartheta}\hat{u}_{2i}) \tag{10}$$

where the unknown coefficients have been replaced by consistent estimators resulting from applying the AGLS estimator. The standard errors of the APEs are obtained using the bootstrap method (Guan 2003).

Results

The Cox model assumes the proportional-hazards assumption, so the test proposed by Grambsch and Therneau (1994) that is based on the scaled Schoenfeld residuals is applied to verify that the log hazard-ratio function is constant over time. The results for each covariate appear in Table A1 of the appendix and show that the variable age and the area of knowledge of medical sciences have time-varying effects. Therefore, the interaction of the latter variables with time has been included in the Cox model, and the results corresponding to the estimates of the hazard ratios for each regressor are set out in Table 5. First, it is verified that the academic field has a significant effect on the duration of the PhD program, once the other regressor effects have been controlled. In particular, at each point of time, the incidence of the event "completion of the PhD program" for students in the Natural Sciences or Engineering and Technology is higher than the one corresponding to the Social Sciences and Humanities (more than 35% and 55%, respectively), which contrasts with the fact that students in the Social Sciences and Humanities take on average less time to complete their undergraduate degree (Lassibille and Navarro 2011). This result can open a debate about the efficiency and productivity of PhD programs in Spain, and the possible sources of inequality that PhD programs could be generating. The main question is why the time to the doctorate is longer in the Social Sciences and Humanities in relation to other academic fields, if the complexity of the PhD programs is not greater than in the other areas of knowledge. The answer to this question falls outside the scope of this research, since more detailed statistical information about the process of completion of the doctoral thesis would be necessary, such as for example, the quality of the students or the difficulties that they faced to fulfil the PhD program related to the choice or change of the thesis topic, supervision of the thesis tutor or the availability of resources to undertake the research (computer equipment, software, data). In this sense, the positive bias of funding programs to PhD students in the fields of Natural Sciences and Engineering and Technology is well known, as it allows them to devote themselves full-time to their doctoral studies. Moreover, another difference between the Humanities and Social Sciences versus the Natural Sciences and Engineering and Technology is that, in the first case, the PhD students work more independently while, in the second case, the PhD students are included in research groups and have a clear roadmap for the completion of their doctoral studies.

With respect to the area of knowledge of the Medical Sciences, this variable has time-varying effects. It is initially observed that the hazard of completing the doctoral studies is 63% higher than the one corresponding to the PhD in Social Sciences, but this percentage decreases 7 percentage points per year. Another relevant result is that the type of research conducted during the PhD program has an impact on the hazard of completing the doctoral studies. For example, the probability of completing the doctoral studies (conditioned to the time spent on the PhD program) is approximately 8% higher for individuals who carry out non-oriented and experimental research.

In relation to the effects of personal and family characteristics, first of all gender is not a significant variable, so there are no discernible differences between males and females students. This result differs from the findings obtained in previous studies that report that men finish their doctoral studies more quickly than women (e.g. Tuckman *et al.* 1990), and proves that in the Spanish case there are no gender differences to complete the doctoral studies. On the one hand, this may mean that the resources and the supervisory/mentoring process is similar for male and female students. On the other hand, it is possible that women can spend similar time to study and complete their doctoral thesis than men in the early stages of their careers when they have fewer family responsibilities. Second, the starting age of the PhD program is negatively related to the time to the doctorate, but this negative effect decreases slightly over time. This variable could be a proxy of the effect of knowledge accumulation on the success of the PhD program. Unfortunately, the data do not allow us to measure the training received by students between the undergraduate degree and the start of the PhD program. Third, family background has an asymmetric effect on the time to the doctorate, since only the maternal educational level has a positive influence on the timing for the completion of the doctoral studies. Thus, students whose mothers have higher education or doctoral studies show a hazard of completing their PhD program that is 26% higher than the one corresponding to the reference category (individuals whose mothers have not completed primary studies). This is hard evidence of intergenerational transmission of education, which has also been widely documented by the literature (e.g. Behrman and Rosenzweig 2002), and reveals that mothers are the strongest models for children's education, having a greater impact on learning and children's academic success than fathers.

Concerning financial support, the estimates indicate significant effects that are particularly important if the individual is a beneficiary of a Spanish or foreign grant. For example, the hazard of completing doctoral studies for students with a Spanish grant is 66% higher than the one corresponding to those who work as research assistants (79% higher if the comparison is made with students working as non-research assistant). These results reveal that the completion of doctoral studies is an activity that requires full commitment and is difficult to combine with paid work, which supports the measures adopted with the new doctorate legislation of assigning different maximum lengths to the PhD program to full-time and part-time students.

[Insert Table 5]

Table 6 provides the estimates corresponding to the probit model associated with the probability of working as a researcher, where the variable time to the doctorate is considered as an endogenous continuous regressor. The first step is to test the endogeneity of the variable time to the doctorate, which is verified with the result obtained from applying the Wald test of exogeneity. Given the previous findings obtained with the estimates of Cox model, we have decided to only consider the mother's education as an instrument of the endogenous regressor. The result associated with the test of overidentifying restrictions corroborates this decision, since the instrument's validity is proved; that is, the mother's education is not correlated with the error term of equation 6. This way, it is possible to confirm that the econometric methodology applied is appropriate.

[Insert Table 6]

In relation to the regressors' influence, we can highlight several interesting findings. First, the average marginal effect corresponding to the time to the doctorate indicates a negative effect of this variable on the probability of working as a researcher: in particular, one more year to complete the doctoral studies decreases this probability by 8 percentage points. From a different point of view, Figure 2 plots the average fitted probabilities according to the time to doctorate. In this graph, for example, we can observe than those individuals who finish their PhD program in less than 2.5 years have a probability of being employed as researcher 41 percentage points higher than the one corresponding to those who complete their doctoral studies in more than 6.5 years (76% *versus* 35%). Several reasons could be suggested to explain these results. On the one hand, individuals could take more time in finishing their PhD program as a result of having poor skills, making a bad selection of the academic field or not having the appropriate supervision or training. Therefore, these facts could be perceived as negative signals by the labour market and the employers. On the other hand, the initial motivation of individuals who take more time in completing the doctorate could be different from that of being employed as a researcher. For example, they could use the doctorate to find a better job, to get a promotion, to expand their knowledge or just for personal fulfilment. These facts seem to be strong reasons for doctorates in Medical Sciences, since they have a probability of working as a researcher that is 15 percentage points lower that the one corresponding to the other areas of knowledge. Second, the results show a slight gender and age disequilibrium to access research careers: in particular, men have a probability of being employed as researchers that is higher by 3 percentage points, and an increase of age by one year diminishes this probability by 0.8 points.

[Insert Fig. 2]

Concerning labour characteristics, first of all part-time wage earners register the lowest probability of working as researchers (12 percentage points less than the one corresponding to full-time wage earners). Second, the type of sector has a significant influence on the likelihood of working as a researcher, since doctorate holders have more probability of performing research tasks in higher-education institutions. In contrast, private firms offer few opportunities: individuals working in this sector have a probability of being a researcher that is 45 percentage points lower than the one corresponding to higher-education institutions. This evidences the weak role of investment in research and development in Spanish firms. Some statistical indicators obtained by the OECD (OECD 2016) can help us understanding this situation; for example, research and development spending by firms in Spain decreases by 7% between the 2007 and 2014. This expenditure represents only 0.7% of GDP in Spain, which contrasts with the percentage reached in other countries, for example, 2.7% in Finland. Moreover, the percentage of researchers working in the private sector in relation to the total number of researchers is 45% in EU-27, whereas in Spain this ratio is only 34.5%.

Finally, the estimates associated with the regional dummy variables indicate that the probability of working as a researcher is affected by the region of residence: in particular, those doctorate holders living in the Centre, Northeast and Northwest of Spain have the lowest options.

Conclusions

This study provides empirical evidence from Spain regarding some aspects of tertiary education and its effects on students' professional careers, which are relevant for policymakers and for the education community. In particular, the analysis has focused on the determining factors that have an impact on the time to complete a PhD program, and on the relationship between the time to the doctorate and the probability of performing tasks as a researcher in the labour market. In order to meet these objectives, we have used data from the 2009 Survey on Human Resources in Science and Technology provided by the Spanish National Institute of Statistics (INE 2010). This survey is the most recent information source in Spain to analyse how doctorate holders are inserted in the labour market as researchers. To the best of our knowledge, this article addresses topics from a methodological approach that is unprecedented in the literature on Spain and Europe. In particular, the econometric method applied to analyse the time to complete the doctorate is a Cox model with shared frailty allowing time-varying effects, whereas the influence of the duration of PhD programs on the probability of pursuing a researcher career is quantified by estimating a probit model with endogenous regressor where the maternal education level is used as an instrumental variable. The findings obtained in this study can represent an element of support for the strategies and the decision-making process of students regarding their choices in human capital investment in relation to PhD programs. On the other hand, they offer empirical knowledge about the transitions from education to employment, currently one of the main concerns in the Spanish society. Moreover, the results can be a source of information that helps policymaker predicting and assessing the efficacy of some legislative changes implemented in PhD programs with the arrival of the Bologna Process.

As a summary, the following main findings should be highlighted. First of all, from a statistical point of view it is observed that the duration of PhD programs differs significantly according to the fields of research, the areas of the Social Sciences and Humanities being the ones where students take more time to complete their PhD program. This has also been tested with the analysis of the survival curves and the application of the log-rank test. Second, the majority of doctorate holders are working as researchers, but there is still room for improvement since 40% are not performing research tasks in their jobs; moreover, in the Medical Sciences the opposite occurs (65% of individuals are not working as researchers). From an econometric point of view, the main conclusions associated with the first objective of the study are the following: first, the doctorate students in the Natural Sciences and Technology and Engineering have a probability of completing their doctoral studies (conditioned to the time elapsed in the PhD program) that is higher than the one corresponding to the Social Sciences and Humanities (more than 35% and 55%, respectively). This result suggests that it is necessary to obtain more information about the internal process leading to the completion of a doctoral thesis, since the complexity of the contents in the Natural Sciences and Engineering and Technology is not *a priori* higher than the one corresponding to the Social Sciences and Humanities. Moreover, another factor explaining this result could be that the first group of areas of knowledge has more availability of financial aid, which allows PhD students to devote themselves full-time to their PhD. For example, in the Europe Horizon 2020 program only 10% of the funds goes to subjects corresponding to the Humanities and Social Sciences. On the other hand, most PhD students in the Natural Sciences and Engineering and Technology are integrated in research groups with a clear plan for completion of the dissertation thesis. Therefore, one policy implication derived from this could be to increase the weight of the Humanities and the Social Sciences in funding programs. Second, students with PhD mothers have a hazard of finishing their PhD program that is 25.5% higher than the one corresponding to those whose mothers have not completed primary studies. This highlights the mother's role as an instrument to preserve educational background. Third, grants are a significant factor for success in a PhD program, since they provide financial support and allow students to carry out their doctoral studies on a full-time basis.

In respect to the second objective, it has been verified that time to the doctorate has a negative effect on the probability of working as a researcher: in particular, those students who complete their PhD program in less than 2.5 years have a probability of being employed as researcher that is 41 percentage points higher than the one corresponding to those who complete their doctoral studies in more than 6.5 years. Hence, the measure of imposing a maximum duration of 3 years for doctoral studies proposed in Spain (Royal Decree 99/2011) would have a positive effect on the transfer of research from the education system to the productive system. In this sense, it is worth noting that the Bologna process does not impose a 3 year limit on the doctorate, although it is a common practice since most European funding programs limit their support to 3 years. The legal limit to the duration of Spanish doctoral studies should be accompanied by other measures to ensure the quality of PhD programs. Thus, it would be necessary to develop an efficient mechanism for recruiting PhD students that ensures their quality and their success in the PhD program; this should be accompanied by transparent information campaigns that pursue a good fit between students and PhD program, and report the competences that may be acquired with a PhD program or about the professional success of former doctorate recipients. Second, it is necessary to reinforce monitoring of the scientific production of PhD students during the completion of their doctoral studies. This could be achieved by different means: on the one hand, by implementing regular meetings where PhD students report their progress; on the other hand, by increasing the incentives to supervise or mentor PhD dissertations since, in most cases, the mentors must combine their supervision of dissertations with their scientific and academic activity. Other policy recommendations would be to increase student support, especially in the knowledge areas of the Humanities and Social Sciences, which are less favoured in the funding programs, and to reward those universities or faculties whose PhD students complete their doctoral studies in the shortest amount of time. Finally, taking more time than average to complete doctoral studies does not only depends on the quality of the students but also on the area of knowledge; hence, it would be interesting to encourage financial aid for hiring PhD holders in the Humanities and Social Sciences to perform research activities.

With respect to other results associated with the second objective of the study, it is noted, first of all, that the time to the doctorate is an endogenous regressor, so the unobservable factors that have an influence on this variable and those corresponding to the propensity of being a researcher are correlated. Second, there are some slight gender disparities in relation to access to a research career in favour of men (men have a probability of being researcher that is higher by 3 points than the one corresponding to women), which contrast with the sharp gender inequalities observed in other aspects of the Spanish labour market. Finally, the employability of doctorate holders as researchers in private firms in Spain is very low, with a probability 44 percentage points lower than the one corresponding to doctorate holders working in higher-education institutions, after controlling for the remaining covariates. This implies that educational and political strategies must be addressed to increase the research role of doctorate holders in the private sector.

References

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: high school curriculum, college major, and careers. *Annual Review of Economics* 4(September), 185-223.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica* 46(5), 1193-1205.

Auriol, L., Misu, M. and Freeman, R. A. (2013). Careers of Doctorate Holders: Analysis of Labour Market and Mobility Indicators. OECD Science, Technology and Industry Working Papers, 2013/04. Paris: OECD.

Behrman, J. & Rosenzweig, M. (2002). Does increasing women's schooling raise the schooling of the next generation?. *American Economic Review* 91(1), 323-334.

Bormann, L. & Enders, J. (2004). Social origin and gender of doctoral degree holders. *Scientometrics* 61(1), 19-41.

Breslow, N. (1974). Covariance analysis of censured survival data. *Biometrics* 30(1), 89-99.

Brooks, R. & Everett, G. (2008). The impact of higher education on lifelong learning. *International Journal of Lifelong Learning* 27(3), 239-254.

Canal, J. F. (2013). Ingresos y satisfacción laboral de los trabajadores españoles con título de doctor. *Revista Española de Investigaciones Sociológicas* 144(octubre), 49-72.

Canal, J.F. & Rodríguez, C. (2013). Wage differences among PhDs by area of knowledge: are science areas better paid than humanities and social ones? The Spanish case. *Journal of Education and Work* 26(2), 187-218.

Canal, J.F. & Rodríguez, C. (2016). Doctoral training and labour market needs. Evidence in Spain. *Research Evaluation* 25(1), pp. 79-93.

Canal, J. F. & Wall, A. (2014). Factors determining the career success of doctorate holders: evidence from the Spanish case. *Studies in Higher Education*, 39(10), 1750-1773.

Caparrós, A. (2014). Geographical mobility and potential wage gain of immigrants within Spain. *Regional Studies*, 48(4), pp. 680-690.

Caparrós, A. (2016). The impact of education on intergenerational occupational mobility. *Journal of Vocational Behavior* 92(February), 94-104.

Cox, D. R. (1972). Regression models and life-tables. *Journal of Royal Statistical Society*, Series B 34(2), 187-220.

Cryer, P. (1998). Transferable skills, marketability and lifelong learning: the particular case of postgraduate research students. *Studies in Higher Education* 23(2), 207-216.

Ehrenberg, R. & Mavros, P. (1995). Do doctoral students' financial support patterns affect their times to degree and completion probabilities?. *The Journal of Human Resources*, 30(3), 581–609.

European Commission (2005). *i2010 – A European information society for growth and employment. COM (2005) 229 final*. Brussels: European Commission.

European Commision (2011). *Report of mapping exercise on doctoral training in Europe "Towards a common approach"*. Brussels: European Commission.

European Commission (2016). *Horizon 2020—Work Programme 2016–2017–3. Marie Sklodowska-Curie Actions. Report No.: C(2016)1349*. Brussels: European Commission.

Evangelauf, J. (1989, March 15). Lengthening of time to earn a doctorate causes concern. *Chronicle of Higher Education* 35, 13-14.

Fox, M. & Stephan, P. (2001). Careers of young scientists: preferences, prospects and realities by gender and field. *Social Studies of Science* 31(1), 109-122.

García-Quevedo, J., Mas-Verdú, F. & Polo-Otero, J. (2012). Which firms want PhDs? An analysis of the determinants of the demand. *Higher Education* 63(5), 607-620.

Golde, C. M. (2005). The role of the department and discipline in doctoral student attrition: Lessons from four department. *The Journal of Higher Education* 76(6), 669-700.

Grambsch, P. M. & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81: 515–526.

Guan, W. (2003). From the help desk: Bootstrapped standard errors. *The Stata Journal* 3(1), 71-80.

INE (Instituto Nacional de Estadística). (2007). *2006 Survey on Human Resources in Science and Technology*. Madrid: INE.

———. (2010) *2009 Survey on Human Resources in Science and Technology*. Madrid: INE.

———. (2015) *Labour Force Surv*ey. Madrid: INE.

King, M.F. (2008). *PhD completion and attrition: analysis of baseline demographic data from the PhD completion project*. Washington, DC: Council of Graduate Schools.

Lassibille, G. & Navarro, L. (2011). How long does it take to earn a higher education degree in Spain?. *Research in Higher Education* 52(1), 63-80.

Lawless, J. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.

Lee, L. (1992). Amemiya's generalized least squares and tests of overidentification in simultaneous equations models with qualitative or limited dependent variables. *Econometric Reviews* 11(3), 319-328.

Lee, H., Miozzo, M. & Laredo, P. (2010). Career pattern and competence of PhDs in science and engineering in the knowledge economy: The case of graduates from a UK research-based university. *Research Policy* 39, 869-881.

McCormick, B. (1997). Regional unemployment and labour mobility in the UK. *European Economic Review*, 41(3-5), 581-589.

Newey, W. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36(3), 231-250.

OECD (Organization for Economic Co-operation and Development) (2013). *Panorama de la Educación, Indicadores de la OCDE 2013. Informe español*. Madrid: Ministerio de Educación, Cultura y Deporte.

———.(2014). *Panorama de la Educación, Indicadores de la OCDE 2014. Informe español*. Madrid: Ministerio de Educación, Cultura y Deporte.

———.(2016). *Main Science and Technology Indicators full database*, http://www.oecd.org/sti/msti.htm.

———.(2018). *Education at a glance: Educational attaintment and labour-force status*. OECD Education Statistics (database).

Robin, S. & Cahuzac, E. (2003). Knocking on academia's doors: an inquiry into the early careers of doctors in life science. *Labour* 17(1), 1-23.

Seagram, B. C., Gould, J., & Pyke, W. (1998). An investigation of gender and other variables on time to completion of doctoral degrees. *Research in Higher Education* 39(3), 319-335.

Sheridan, P. M. & Pyke, S. W. (1994). Predictors of time to completion of graduate degrees. *Canadian Journal of Higher Education* XXIV-2, 68-87.

Siegfried, J. J. & Stock, W. A. (2001). So you want to earn a Ph.D. in Economics: How long do you think it will take? *Journal of Human Resources* 36(2), 364-378.

Spence, M. A. (1973). Job market signalling. *Quarterly Journal of Economics* 87(3), 355-374.

Thune, T. (2009). Doctoral students on the university-industry interface: a review of the literature. *Higher Education* 58, 637-651.

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research* 45(1), 89-125.

Tuckman, H., Coyle, S., & Bae, Y. (1990). *On Time to the Doctorate: A study of the increased time to complete doctorates in Science and Engineering*. Washington, DC: National Academy Press.

UNESCO (United Nations Educational, Scientific and Cultural Organization) (2015). *UNESCO Science Report: Towards 2030*. Paris: UNESCO.

Walker, I. & Zhu, Y. (2008). The college wage premium and the expansion of higher education in the UK. *The Scandinavian Journal of Economics* 110(4), 695-709.

Wao, H.O. (2010). Time to the doctorate: Multilevel discrete-time hazard analysis. *Educational Assessment. Evaluation and Accountability* 22(3), 227-247.

Wooldrige, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

World Bank (2013). *World development indicators 2013*. Washington, D.C: World Bank.

Appendix

[Insert Table A1]

Table 1 Time to the doctorate by area of knowledge

Area of knowledge	Mean	Std. Dev.	%	N
Natural Sciences	63.93	26.86	32.78	1,345
Engineering and Technology	65.97	27.83	0.08	343
Medical Sciences	70.79	39.75	20.91	858
Agricultural Sciences	66.71	36.29	0.03	134
Social Sciences	77.43	38.45	20.00	821
Humanities	82.65	42.39	14.67	602
Total	71.07	35.77	100	4,103

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

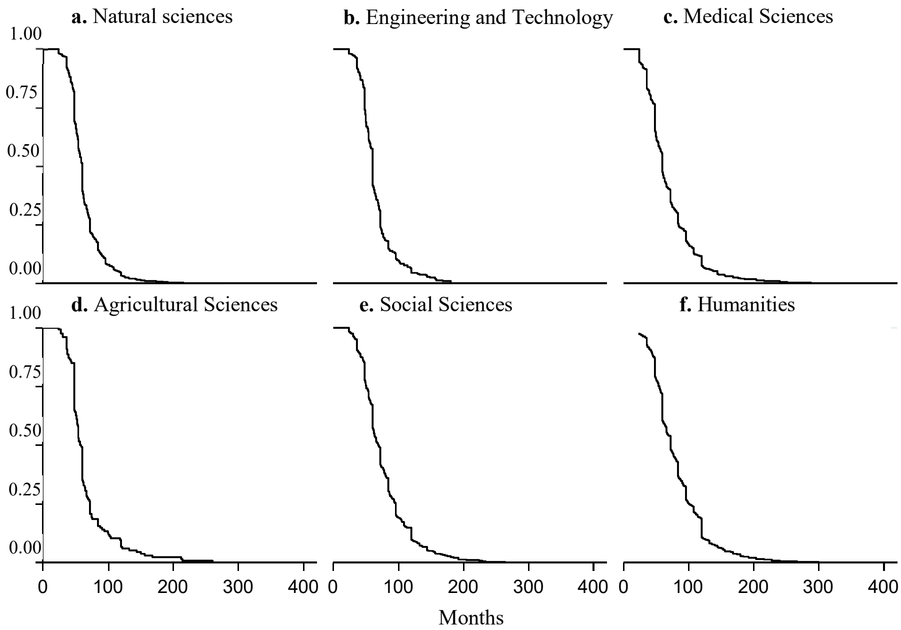

Fig. 1 Kaplan-Meier survival estimates by area of knowledge

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table 2 Long-rank test for each pair of academic field

Pair of academic field	Long-rank statistic	
Natural Sciences *versus* other areas of knowledge:		
Engineering and Technology	1.84	
Medical Sciences	32.00	***
Agricultural Sciences	0.73	
Social Sciences	98.76	***
Humanities	135.41	***
Engineering and Technology *versus* other areas of knowledge:		
Medical Sciences	6.53	**
Agricultural Sciences	0.00	
Social Sciences	33.19	***
Humanities	55.02	***
Medical Sciences *versus* other areas of knowledge:		
Agricultural Sciences	2.13	
Social Sciences	10.13	***
Humanities	27.07	***
Agricultural Sciences *versus* other areas of knowledge:		
Social Sciences	12.93	***
Humanities	22.78	***
Social Sciences *versus* Humanities	6.17	**

** $p<0.05$; *** $p<0.001$.
Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table 3 Descriptive statistics of the variables affecting the time to the doctorate

Variables	Mean	St. Dev.
Time to the doctorate (months)	71.07	35.77
Age	28.61	6.18
Gender		
Male	0.55	0.50
Female	0.45	0.50
Maternal educational level		
Less than primary education	0.38	0.46
Primary education	0.30	0.45
Low technical education	0.02	0.13
Secondary education	0.10	0.29
Upper technical education	0.01	0.11
Short Higher education	0.11	0.31
Long Higher education	0.07	0.25
Doctoral studies	0.02	0.13
Paternal educational level		
Less than primary education	0.28	0.45
Primary education	0.21	0.40
Low technical education	0.03	0.17
Secondary education	0.08	0.27
Upper technical education	0.05	0.20
Short higher education	0.12	0.32
Long higher education	0.16	0.36
Doctoral studies	0.07	0.26
Area of knowledge		
Natural Sciences	0.32	0.46
Engineering and Technology	0.08	0.27
Medical Sciences	0.21	0.40
Agricultural Sciences	0.03	0.17
Social Sciences	0.20	0.39
Humanities	0.14	0.35
Type of research		
Non-oriented research	0.72	0.44
Applied research	0.63	0.48
Experimental research	0.37	0.48
Studies' financing sources		
Spanish grant	0.38	0.48
Foreign grant	0.02	0.10
Earned income working as research assistant	0.24	0.42
Earned income working as not research assistant	0.17	0.37
Loans or savings	0.19	0.39
Observations	4,103	

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table 4 Descriptive statistics of the variables affecting the probability of working or not as a researcher

Variables	Working as a researcher		Not working as a researcher	
	Mean	St. Dev.	Mean	St. Dev.
Time to the doctorate (months)	68.24	31.78	74.08	37.11
Age	37.17	6.77	38.95	7.30
Gender				
Male	0.57	0.49	0.53	0.50
Female	0.43	0.49	0.47	0.50
Area of knowledge				
Natural Sciences	0.40	0.35	0.25	0.40
Engineering and Technology	0.10	0.30	0.07	0.23
Medical Sciences	0.12	0.32	0.35	0.47
Agricultural Sciences	0.03	0.17	0.03	0.17
Social Sciences	0.22	0.41	0.17	0.37
Humanities	0.13	0.34	0.13	0.35
Type of research				
Non-oriented research	0.74	0.43	0.70	0.46
Applied research	0.63	0.48	0.63	0.47
Experimental research	0.37	0.48	0.37	0.48
Labour market status				
Wage earners: Open-ended and full contract	0.77	0.41	0.80	0.39
Wage earners: Fixed-term and full contract	0.19	0.38	0.12	0.32
Wage earners: Part-time	0.04	0.18	0.08	0.27
Type of sector				
Firm	0.05	0.22	0.20	0.40
Public administration	0.29	0.45	0.60	0.49
Higher education institution	0.61	0.48	0.16	0.36
Private non-profit institution	0.04	0.19	0.04	0.19
Current of Spanish regions of residence				
Canarias	0.05	0.20	0.04	0.18
Centre	0.09	0.28	0.11	0.31
East	0.24	0.42	0.22	0.41
Madrid	0.17	0.37	0.20	0.40
Northeast	0.12	032	0.14	0.33
Northwest	0.13	0.33	0.13	0.33
South	0.20	0.40	0.16	0.36
Observations	2,309		1,397	

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table 5. Cox model with shared frailty: Hazard ratios associated with the regressors

Variables	Cox model with shared frailty[a]	
Age	1.050	***
Age*Time	0.999	***
Gender		
Male	1.024	
Maternal education		
Primary education	1.147	**
Low technical education	1.166	
Secondary education	1.083	
Upper technical education	1.402	**
Short higher education	1.129	**
Long higher education or doctoral studies	1.256	**
Paternal education		
Primary education	0.946	
Low technical education	0.928	
Secondary education	0.930	
Upper technical education	0.939	
Short higher education	0.998	
Long higher education or doctoral studies	0.949	
Area of knowledge		
Natural Sciences	1.395	***
Engineering and Technology	1.358	***
Medical Sciences	1.627	**
Medical Sciences *Time	0.995	***
Agricultural Sciences	1.178	**
Humanities	0.866	**
Type of research		
Non-oriented research	1.081	*
Applied research	1.038	**
Experimental research	1.082	**
Studies' financing sources		
Spanish grant	1.791	***
Foreign grant	1.576	***
Earned income working as research assistant	1.081	*
Loans or savings	1.095	*
Likelihood ratio test of the frailty's variance	0.240	
Overall Wald test	441.82	***
Observations	4,103	

* p.<0.1; ** p.<0.05; *** p.<0.001.

[a] The individual of reference is a woman with father and mother without primary education, doctorate in Social Sciences (type of research: oriented, non-applied and non-experimental), and her doctoral studies have been financed working as not research assistant.

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table 6 Estimates of the probit model with endogenous continuous regressor: Probability of working as a researcher

Variables	Marginal Effects[a]	
Time to the doctorate (months)	-0.007	***
Age (years)	-0.008	***
Gender		
Male	0.025	***
Area of knowledge		
Natural Sciences	-0.001	
Engineering and Technology	-0.001	
Medical Sciences	-0.152	***
Agricultural Sciences	-0.044	
Humanities	-0.002	
Type of research		
Non-oriented research	-0.012	
Applied research	0.005	
Experimental research	-0.026	**
Labour market status		
Wage earners: Fixed-term and full-time contract	0.030	
Wage earners: Part-time	-0.122	**
Type of sector		
Firm	-0.459	***
Public administration	-0.253	***
Private non-profit institution	-0.182	***
Current of Spanish regions of residence		
Canarias	-0.017	
Centre	-0.072	**
East	0.018	
Madrid	-0.010	
Northeast	-0.044	**
Northwest	-0.045	**
Wald test of exogeneity	7.190	**
Test of overidentifying restrictions	8.035	
Overall test Wald	2,904.74	***
Observations	3,706	

** p.<0.05; *** p.<0.001.

[a] The individual of reference is a woman doctorate in Social Sciences (type of research: oriented, non-applied and non-experimental), working in a higher education institution with an open-ended and full-time contract, and residing in the South region.

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

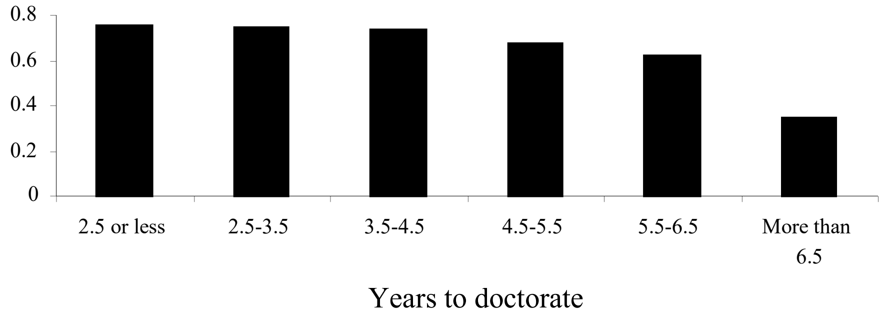

Fig. 2 Average fitted probabilities by years to doctorate

Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).

Table A1 Proportional hazard test by covariates and global test

Variable	Chi-square	
Age	16.86	***
Gender		
Male	3.08	
Maternal education		
Primary education	0.27	
Low technical education	0.32	
Secondary education	0.04	
Upper technical education	0.00	
Short higher education	0.13	
Long higher education or doctoral studies	0.24	
Paternal education		
Primary education	0.05	
Low technical education	2.44	
Secondary education	0.02	
Upper technical education	0.03	
Short higher education	0.28	
Long higher education or doctoral studies	0.27	
Area of knowledge		
Natural Sciences	0.59	
Engineering and Technology	0.02	
Medical Sciences	5.42	**
Agricultural Sciences	3.20	
Humanities	1.07	
Type of research		
Non-oriented research	0.47	
Applied research	1.34	
Experimental research	0.32	
Studies' financing sources		
Spanish grant	0.28	
Foreign grant	0.07	
Earned income working as research assistant	1.74	
Loans or savings	3.50	
Global test	64.82	***

** p. <0.05; *** p. <0.001.
Source: Own elaboration from 2009 Survey on Human Resources in Science and Technology (INE 2010).