KNOWLEDGE OF CATALAN, PUBLIC/PRIVATE SECTOR CHOICE AND EARNINGS: EVIDENCE FROM A DOUBLE SAMPLE SELECTION MODEL

Antonio Di Paolo

Universitat Autònoma de Barcelona & IEB

Abstract:
This paper explores the earnings return to Catalan knowledge for public and private workers in Catalonia. In doing so, we allow for a double simultaneous selection process. We consider, on the one hand, the non-random allocation of workers into one sector or another, and on the other, the potential self-selection into Catalan proficiency. In addition, when correcting the earnings equations, we take into account the correlation between the two selectivity rules. Our findings suggest that the apparent higher language return for public sector workers is entirely accounted for by selection effects, whereas knowledge of Catalan has a significant positive return in the private sector, which is somewhat higher when the selection processes are taken into account.
Keywords: Language, Sector Choice, Earnings, Simultaneous Selection, Catalonia
JEL classifications: J24, J45, J70, C31

Corresponding author: Antonio Di Paolo; e-mail: antonio.dipaolo@uab.cat, tel.: +34935813415; fax: +34935812292. Departament d'Economia Aplicada, Universitat Autònoma de Barcelona (UAB); Campus de Bellaterra, Edifici B 08193 Bellaterra (Cerdanyola), Spain. Institut d'Economia de Barcelona, Universitat de Barcelona.

1. Introduction

The period of democracy in Spain, which started with the end of the dictatorship (1939-1975), has been characterized by large-scale regional devolution. The economic, legal and political decentralization has brought with it a significant degree of independence from the State, especially for regions such as Catalonia. An additional feature of the democratization process is the general expansion of the Welfare State, which has led to a huge increase in public sector employment. In Catalonia, the combination of these two factors has meant a great expansion of local government in the last thirty years, which has progressively gained importance with respect to some of the existing centralized institutions. Another important aspect of the democratization process is the recognition of regional culture and language, which were severely repressed during the dictatorship (the Franco regime had prohibited the use of Catalan in public and strongly disapproved of its use in private). In the case of Catalonia, this "cultural devolution" has been spearheaded by two major public linguistic

1

policies aimed at promoting and enhancing the use of the Catalan language among the population.

Specifically, the first linguistic policy implemented by Catalonia's Autonomous Government (the *Generalitat*) was the "Linguistic Normalization Act" of 1983 (*Llei de Normalització Lingüística*), which aimed to reinstate the public use of Catalan and to stimulate its learning and its use in private. This law established not only that Catalan was to be the official language of the Catalan government and of local public administrations, but also the main language used in primary and secondary education[1]. In order to stimulate and improve private study, the *Generalitat* also began to organize language courses directed to the adult population, offered completely free of charge. More relevant today for the economic value of knowledge of Catalan was the "Linguistic Policy Act" of 1998 (*Llei de Política Lingüística*), which attempted to reassert the presence of Catalan (versus Castilian) by a) increasing fluency requirements for public sector employees, and b) introducing major incentives (and in some case requirements) for increasing the use of Catalan in private business and other socioeconomic and cultural domains[2] (Solé and Alarcón 2001).

These two public policies have increased the economic value of knowledge of Catalan in the local labour market, to the extent that proficiency in the language is believed to be very important today. In fact, using Census data of 1991 and 1996 (that is, even before the implementation of the linguistic reform of 1998), Rendon (2007) shows that knowledge of Catalan has a substantial impact on the probability of employment among native-born individuals and immigrants from the rest of Spain. Moreover, Quella and Rendon (2009) suggest that the skills of speaking and writing Catalan have a significant effect on employment prospects, especially in white-collar occupations, services, government and education. Finally, a recent study by Di Paolo and Raymond (2010) stresses the existence of significant earnings returns to Catalan proficiency (defined as the ability to speak and write in Catalan) among first- and second generation immigrants in Catalonia. In general, the potential return to Catalan proficiency exists because, in spite of the efforts at institutional level to raise Catalan fluency among the population, full functional knowledge is far from widespread.

This study attempts to make another step forward, by simultaneously considering the role of knowledge of Catalan in sector choice and the relationship between language proficiency

---

[1] In fact, Spanish (or Castilian) is taught as a second language in pre-university education. At university the language used is not determined by law, and is established by the professor.
[2] With the 1998 Act, private suppliers of public services have been subjected to almost the same linguistic requirements as the public sector. Moreover, the Catalan government has introduced economic incentives for "normalizing" Catalan in private firms and stimulating active learning of this language by workers; finally, the government also promoted the language through the mass media by introducing incentives in the use of Catalan on radio, TV and also in the newspapers and written publications in general.

and monthly earnings. Applying a level of detail that goes beyond our previous study, the general aim of this paper is to determine whether there is an earnings/productivity effect of the knowledge of Catalan, distinguishing between the public and the private sectors. We also aim to establish whether the statistical association between knowledge of Catalan and earnings in the two sectors is affected by the potential relationship between language proficiency and private/public sector selection. In other words, we want to examine whether knowledge of Catalan is considered an "advantage" in the two sectors (and consequently increases an individual's earnings), or whether it merely represents a "requirement" for working in the Catalan public sector. We hypothesize that with the strict regulations imposed by the legislation on language mentioned above, and the progressive contraction of public institutions of the State (which are not regulated by the regional legislation), Catalan proficiency will emerge as a prerequisite to enter the public sector, and may represent an advantage in the private sector.

With these purposes in mind, this paper proceeds as follows. The next section offers a brief review of the relevant literature and situates the study; section 3 presents the data used in the empirical analysis and some descriptive statistics. Section 4 describes the econometric methodology for dealing with this particular issue; section 5 contains the empirical results, and section 6 concludes.

2. Related Literature

This paper is based on two main groups of previous research. The first is the vast recent literature on the economic return to language proficiency (see Chiswick and Miller 2007 and Chiswick 2008 for a comprehensive overview). In general, this literature highlights two well-established findings. On the one hand, language proficiency is in general associated with higher earnings among the immigrant population, and language deficiencies explain a significant part of the migrants' earnings gap with respect to the native population; on the other, the statistical association between individual labour market outcomes and language proficiency is likely to be biased by unobserved individual heterogeneity[3], which requires the implementation of more sophisticated econometric techniques than a simple OLS estimation,

---

[3] Moreover, given that the information on language knowledge is habitually self-reported in the surveys, misclassification/measurement error is also a problematic issue in the empirical literature which introduces additional methodological complications; see Dustmann and Van Soest (2001, 2002) for further details.

including Instrumental Variables, Self-Selection models, Statistical Matching, Differentiation with Panel Data, depending on data availability.

However, it seems that the relationship between language proficiency and earnings cannot be considered in isolation from occupation-/sector choice-related decisions. In fact, the papers by Berman et al. (2003) and Lang and Siniver (2009), based on panel data, suggest that knowing Hebrew in Israel has a positive value only in high-skill occupations, and that the apparent language return for low-skilled workers is due entirely to unobserved heterogeneity (ruled out by taking the first difference from the longitudinal estimates). However, neither of these papers explicitly considers that occupation is the intervening activity that links income to education, and the other forms of Human Capital — including language proficiency. This would mean that language knowledge may play an important role in determining the type of occupation the individual can enter; therefore, more fluent individuals would tend to be self-selected into occupations with higher language requirements. This possibility has been explicitly considered in a recent paper by Chiswick and Miller (2010), who exploited the O*NET database that contains information on occupation-specific language requirements in the US. The authors find a positive sorting of more proficient workers into occupations that require a higher level of language fluency; moreover, they argue that the positive effect of language proficiency is higher the higher the level of English required in the occupation. Finally, Aldashev et al. (2009) explore the effect of language knowledge on immigrants' earnings in Germany, considering the potential effect of language fluency on the simultaneous selection process into economic sector and occupation type. After estimating a two-step model with multiple simultaneous source of selection, they find that when the positive effect of language proficiency in the simultaneous selection of economic sector and occupation type is taken into account, there is no evidence to suggest a pure productivity language effect. In other words, it seems that the earnings return to language knowledge is only indirect, because more proficient workers are more likely to be selected into high-paid works.

As briefly mentioned above, in this paper we investigate the potential earnings return to proficiency Catalan in the public and private sectors. Given the strict regulation of language requirements in public sector occupations, we suspect that the economic value of knowledge of Catalan may differ radically between these domains. Nevertheless, we believe that in the case of Catalonia, language proficiency and the public/private sector choice can be taken as a joint simultaneous decision, and that this simultaneity must be taken into account in order to obtain a correct estimate of the earnings return to Catalan in the two sectors. Therefore, as

references we also take some of the studies of the earning differentials across the public and private sectors.

In particular, we refer to studies that consider the existence of a potential selectivity process behind the sector choice decision, which is taken into account in the estimation of earning determinants for predicting earning differentials in the two sectors — normally using Endogenous Switching Models (see van der Gaag and Vijverberg 1988, Hartog and Oosterbeek 1993, Dustman and van Soest 1998, Adamchik and Bedi 2000, Bender (2003) among others). We also refer to two other studies that consider the existence of other sources of selection apart from the public/private sector choice, which are treated in much the same way as in Aldashev et al. (2009). Specifically, the paper by Christofides and Pashardes (2002) takes into account the existence of a double selection problem in the choice between self- and paid-employment and public/private sector selection. Nevertheless, their results indicate that these two underlying choices are not interrelated, and they eventually estimate the wage equations controlling for two independent selection correction terms — one for the type of employment and the other for the desired sector. Moreover, Heitmueller (2006) treats labour force participation and sector choice as simultaneous decisions, and he takes account of this potential double selection mechanism for computing the earning gap between public and private sector employees.

Following these two strands of the literature, we start by modelling a Bivariate Probit Equation to explain on the one hand the propensity to be proficient in Catalan or not, and on the other hand the decision to work in the public or the private sector. Subsequently, as explained in detail in what follows, from this bivariate estimation we construct two selectivity correction terms — one for sector choice and the other for knowledge of Catalan respectively — that take into account the simultaneity of the two self-selection mechanisms when estimating the earnings equations. In general, the inclusion of the correction terms would adjust the biases in the estimates caused by the non-random assignment of workers in the economic sectors, and by the potential self-selection into knowledge of Catalan. However, if the likelihood of working in one sector or the other and the propensity to be fluent in Catalan are interrelated variables (this is a realistic possibility in the case of Catalonia), neglecting this simultaneity would lead to inconsistent estimates of the economic value of knowledge of Catalan in the two sectors. As illustrated below, the empirical results indicate that the two selection processes are positively related, and this relationship must be taken into account for a correct estimation of the economic value of knowledge of Catalan in the public and private sectors. Moreover, the results indicate that the apparently higher return to knowledge of

Catalan in the public sector estimated by simple OLS is accounted for entirely by the fact that proficient workers are more likely to be selected into the public sector (and vice versa). In contrast, there exists a significant return to proficiency in Catalan for individuals who work in the private sector, which is significantly higher when selection into language knowledge and economic sector is accounted for. This empirical evidence could be taken as an indication that the economic value of Catalan proficiency for public sector workers is only represented by a higher chance of working in this sector, but (once entered) language knowledge will have little effect on improving earning opportunities[4] — i.e. knowledge of Catalan is merely an entry requirement. However, in the absence of a strict regulation regarding language capabilities in the private sector, fluency in Catalan may represent an advantage for proficient (private) workers, which is reflected by higher expected monthly earnings.

3. Data and Descriptive Statistics

The empirical analysis is based on the data from the 2006 "Survey of Living Conditions and Habits of the Catalan Population (ECVHP06)" carried out by the Statistical Institute of Catalonia (IDESCAT). The original sample comprises 10,358 observations of individuals aged 16 or more residing in Catalonia. The survey aims to obtain socioeconomic and demographic information of the overall population, focusing on individuals and their families. The data were collected between the fourth trimester of 2005 and the third trimester of 2006; therefore, the information on individual labour market status and monthly earnings reflects the situation in 2005-2006. Analyzing this period is very attractive for our purposes, since the unemployment rate was exceptionally low (6.6%)[5]; this means that we can focus only on the employed population, as we consider that neglecting the potential self-selection into employment would not be problematic during this period[6]. Moreover, the relatively high rate of female activity (52.3%) allowed us to include women in the analysis. The final sample consisted in 5,019 observations of all the individuals aged 16 to 65 in regular employment, with valid information on earnings (recorded in brackets).

---

[4] Consistently, individual earnings in the public sector are determined on the basis of educational attainment, professional category and seniority. This means that two identical individuals who are working in the same occupation cannot be discriminated in terms of earnings on the basis of language proficiency; however, during the hiring stage, there may be significant discrimination in favour of the candidate with knowledge of Catalan.

[5] The value reported is the mean unemployment rate between the fourth trimester of 2005 and the third trimester of 2006. The information is taken from the EPA (*Encuesta de la Población Activa*, Active Population Survey (INE)) for Catalonia. Similar values for the unemployment rate in Catalonia had not been recorded since 1978.

[6] We also tried to estimate a Probit model for employment, but it performed very poorly due to the extremely low number of zeros (unemployed individuals).

The information on Catalan knowledge contained in the survey is reported in four categories: namely, an individual may claim he/she "does not understand", "understand but is unable to speak", "is able to speak but not to write", and "is able to speak and write" Catalan. Only individuals who could speak and write Catalan were considered fully proficient. This restrictive definition of language proficiency might help to minimize the potential measurement error in the self-reported language knowledge variable[7]. Table 1 contains a basic description of the selected sample, in terms of the two relevant dimensions of analysis — i.e. proficiency in Catalan and economic sector. Specifically, each cell of this 2×2 table contains the expected frequencies of every possible outcome, and the row-percentages in italics inside brackets.

Table 1: Expected Frequencies according to proficiency in Catalan and public/private sector

		CATALAN PROFICIENCY		
		Yes	*No*	*Total*
ECONOMIC SECTOR	*Public*	688 *(83%)*	141 *(17%)*	829 *(100%)*
	Private	2,542 *(61%)*	1,648 *(39%)*	4,190 *(100%)*
Total		3,230	1,789	5,019

Source: ECVHP06, our own calculations.
Proficiency: ability to speak and write Catalan.

Slightly more than 16% of the selected sample work in the public sector, and 83% of them are fully proficient in Catalan. In all likelihood, the remaining 17% comprise individuals who work in the institutions of the central government, where knowledge of Catalan is not strictly considered a requirement. In the private sector, which represents 83% of the final sample, the proportion of fully proficient workers falls to 61%, reflecting the lack of any strict public regulation concerning knowledge of Catalan in this sector. These differences in the distribution of language capital in the Catalan labour market may be reflected in different rates of return to language knowledge in the two sectors.

Given that knowledge of Catalan is significantly more widespread in the public than in the private sector, one might expect to observe higher return among private sector workers. Table 2 illustrates the means of log earnings[8] in the two sectors according to Catalan proficiency. This descriptive evidence is the exact opposite of what we might have expected.

---

[7] This means that we may be estimating the lower bound of the true return to Catalan knowledge; in fact, it is quite reasonable to assume that individuals tend to over-report their true language abilities. Therefore, we believe that individuals who claim to be able to speak and write Catalan have at least an acceptable functional knowledge of the language.

[8] The information on individual earnings is presented in brackets in the ECVHP06 survey. We adopt the standard solution of creating a continuous earnings variable over the mid-points of each earnings interval; see table 1A in the Appendix for details.

Table 2: Means and Equality Tests for Monthly Earnings by Economic Sector and Language Knowledge

EXPECTED LOG-MONTHLY EARNINGS	ECONOMIC SECTOR		Difference in mean t-statistic
	PUBLIC	PRIVATE	
TOTAL SAMPLE	7.17 *(0.018)*	6.94 *(0.008)*	11.872
PROFICIENT	7.20 *(0.019)*	6.98 *(0.01)*	10.41
NOT PROFICIENT	7.00 *(0.043)*	6.88 *(0.012)*	2.672
Difference in mean t-statistic	*4.34*	*6.35*	

Note: standard errors in italics inside brackets.

As commonly reported in the literature, public sector workers earn significantly more than those in the private sector, but the positive statistical association between monthly earnings and language proficiency is higher in the former (0.2 log points). Moreover, the positive earnings premium in the public sector is significantly more pronounced among proficient workers. Even so, the interest lies in the *ceteris-paribus* earnings return to knowledge of Catalan, whereas these simple mean differences could be confounded by the effect of other earnings determinants. We may start with the estimation of two OLS regressions (one for each sector), which include the typical earnings covariates plus a variable indicating whether the individual is proficient in Catalan. However, there are at least three reasons for believing that the OLS estimates of the language return in the two sectors may be biased. First, individuals may be selected into public or private sector jobs according to unobservable characteristics that also affect their earnings potential — i.e. the allocation of the labour force into the private or the public sector is not random. Second, the relationship between Catalan proficiency and monthly earnings may be affected by unobserved individual heterogeneity. Third, the unexplained propensity to speak and write in Catalan and the likelihood of being selected in the public sector are correlated.

We deal with these potential sources of inconsistency with a bivariate selection model, allowing for a simultaneous selectivity process captured by two joint reduced-form selection equations — one for being proficient in Catalan and one for working in the public or the private sector — as explained in detail in the next section. Therefore, in order to model monthly earnings, knowledge of Catalan and sector choice, we exploit all the relevant information contained in the ECVHP06 database[9]. Table 2A in the Appendix contains the basic descriptive statistics separately for public and private workers. The sub-sample of public sector workers is somewhat older, highly feminized, and better educated than the sub-sample

---

[9] Descriptions of each explanatory variable can be found in Table 1A in the Appendix.

of private sector workers. As expected, there is a higher presence of foreign workers in the private sector. Public sector workplaces are more stable, given the higher job tenure and the higher unionization rate; finally, public sector employees work fewer hours per month than those in the private sector.

4. Empirical Strategy

We start by estimating two log-earning regressions: one for private sector employees ($PUB_i=0$) and one for public sector employees ($PUB_i=1$); the δ coefficients in (1) represent our parameters of interest, which capture the percentage earnings increase associated with Catalan knowledge ($CAT_i = 1$ if the individual is proficient in Catalan, 0 otherwise).

$$\ln Y_{i,PUB0} = \alpha_{PUB0} + X'_i\beta_{PUB0} + CAT_i\delta_{PUB0} + \varepsilon_{PUB0} \quad \text{if } PUB_i = 0$$
$$\ln Y_{i,PUB1} = \alpha_{PUB1} + X'_i\beta_{PUB1} + CAT_i\delta_{PUB1} + \varepsilon_{PUB1} \quad \text{if } PUB_i = 1 \tag{1}$$

A selectivity problem arises when the likelihood of entering the public sector and/or the propensity to achieve full language proficiency depend on unobservable individual characteristics that are potentially related to the unobservable earnings determinants. The two selection processes can be treated with the standard methods proposed by Lee (1978) and Heckman (1979), but only if the two selection rules are strictly independent.

However, in our case, the selection rules — i.e. public/private sector choice, and proficiency in Catalan — are clearly unlikely to be independent. In fact, because of the Catalan institutional setting those who work in the public sector are, in general, more likely to know Catalan, and those who are fully proficient in Catalan may be more likely to work in the public sector[10]. This means that we must deal with a joint double selection rule, which can be written as

$$CAT_i^* = Z'_i\gamma + u_i \tag{2a}$$

$$PUB_i^* = W'_i\psi + \omega_i \tag{2b}$$

with $\text{var}(u_i) = \text{var}(\omega_i) = 1;\ \text{cov}(u_i,\omega_i) = \text{corr}(u_i,\omega_i) = \rho_{u,\omega},$

---

[10] In fact, we consider two similar individuals who differ only with respect to knowledge of Catalan, because one of them is fully proficient and the other is not: the former has a clear comparative advantage over the latter for entering the public sector, which would mean a higher likelihood of working in that sector (*ceteris paribus*).

where Z_i and W_i contain the observable determinants of the latent propensity to know Catalan (*CAT**) and of the desired sector choice (*PUB**) respectively, and $\rho_{u,\omega}$ represents the correlation coefficient between the unobservable elements of the two equations. If this correlation coefficient is statistically different from zero, we must generalize the selectivity problems to a double simultaneous selection process, which can be addressed with the methodology proposed by Fishe et al. (1981), Ham (1982) and Tunali (1986), and more recently used by Heitmueller (2006) and by Aldashev et al. (2009) among others. Specifically, taking expectation of the earnings equations in (1) we obtain that,

$$E\left[\ln Y_{i,PUB0} \mid X_i, PUB_i = 0, CAT_i\right] = \alpha_{PUB0} + X_i'\beta_{PUB0} + CAT_i\delta_{PUB0} + E\left[\varepsilon_{PUB0} \mid X_i, PUB_i = 0, CAT_i\right]$$

$$E\left[\ln Y_{i,PUB1} \mid X_i, PUB_i = 1, CAT_i\right] = \alpha_{PUB1} + X_i'\beta_{PUB1} + CAT_i\delta_{PUB1} + E\left[\varepsilon_{PUB1} \mid X_i, PUB_i = 1, CAT_i\right]$$

$$(3)$$

where the last terms in both equations contain the joint double selectivity bias; following the two-step procedure proposed by Tunali (1986), we consider this generalized selectivity problem as a double simultaneous selection situation, with full information on the outcomes of the two selection rules (giving four distinct cells). That is, as shown in the descriptive analysis, the sample contains cases of proficient individuals in either the public or the private sector. Moreover, there are also public sector employees who are not proficient in Catalan, the ones who work in the central government's institutions — that do not consider knowledge of Catalan as a strict requirement for workers. Finally, we obviously observe private sector employees who do not know Catalan, given the absence of a general legal requirement regarding knowledge of Catalan, and the co-existence with Spanish. In addition, we may also reasonably assume that the two selection rules are simultaneous (rather than sequential), given that during the "linguistic normalization" process, public sector workers with limited knowledge of Catalan were allowed to improve their proficiency by attending specific language courses for public-employees, provided free of charge by the Catalan government (and normally taught during part of the working day).

Therefore, Tunali (1986) shows that, assuming a joint normal distribution of the error terms *(ε_j, u, ω)* with zero mean and variance-covariance matrix[11]

---

[11] Notice that since the covariance between ε_{PUB1} and ε_{PUB0} is not directly identifiable, the variance-covariance matrix has been split into two matrices.

$$\sum{}_{j} = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} & \sigma_{\varepsilon\omega} \\ & \sigma_u^2 & \sigma_{u\omega} \\ & & \sigma_\omega^2 \end{bmatrix}, j = PUB1; PUB0,$$

the correction terms for the two selectivity processes would take the form:

$$\lambda_{i,PUB} = \begin{cases} \phi(W_i'\hat{\psi}) \times \Phi[A_i] \times F\left(Z_i'\hat{\gamma}, W_i'\hat{\psi}; \rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 1 \,\&\, PUB_i = 1 \\ \phi(W_i'\hat{\psi}) \times \Phi[-A_i] \times F\left(-Z_i'\hat{\gamma}, W_i'\hat{\psi}; -\rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 0 \,\&\, PUB_i = 1 \\ -\phi(W_i'\hat{\psi}) \times \Phi[A_i] \times F\left(Z_i'\hat{\gamma}, -W_i'\hat{\psi}; -\rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 1 \,\&\, PUB_i = 0 \\ -\phi(W_i'\hat{\psi}) \times \Phi[-A_i] \times F\left(-Z_i'\hat{\gamma}, -W_i'\hat{\psi}; \rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 0 \,\&\, PUB_i = 0 \end{cases}$$

(4)

$$\lambda_{i,CAT} = \begin{cases} \phi(Z_i'\hat{\gamma}) \times \Phi[B_i] \times F\left(Z_i'\hat{\gamma}, W_i'\hat{\psi}; \rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 1 \,\&\, PUB_i = 1 \\ -\phi(Z_i'\hat{\gamma}) \times \Phi[B_i] \times F\left(-Z_i'\hat{\gamma}, W_i'\hat{\psi}; -\rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 0 \,\&\, PUB_i = 1 \\ \phi(Z_i'\hat{\gamma}) \times \Phi[-B_i] \times F\left(Z_i'\hat{\gamma}, -W_i'\hat{\psi}; -\rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 1 \,\&\, PUB_i = 0 \\ -\phi(Z_i'\hat{\gamma}) \times \Phi[-B_i] \times F\left(-Z_i'\hat{\gamma}, -W_i'\hat{\psi}; \rho_{u,\omega}\right)^{-1} & \text{if } CAT_i = 0 \,\&\, PUB_i = 0 \end{cases}$$

where

$$A_i = \frac{W_i'\hat{\psi} - \rho_{u,\omega} Z_i'\hat{\gamma}}{\left(1 - \rho_{u,\omega}^2\right)^{1/2}}; B_i = \frac{W_i'\hat{\psi} - \rho_{u,\omega} Z_i'\hat{\gamma}}{\left(1 - \rho_{u,\omega}^2\right)^{1/2}},$$

and $F(\cdot)$ stands for the Bivariate Normal Distribution of the predicted probabilities computed from the joint estimation of (2a) and (2b) with a Bivariate Probit model. This means that the conditional expected earnings in (3) can be written as

$$E\left[\ln Y_{i,PUB0} \mid X_i, PUB_i = 0, CAT_i\right] = \alpha_{PUB0} + X_i'\beta_{PUB0} + CAT_i\delta_{PUB0} + \sigma_{\varepsilon 0}\rho_{\varepsilon 0u}\lambda_{i,PUB} + \sigma_{\varepsilon 0}\rho_{\varepsilon 0\omega}\lambda_{i,CAT}$$

$$E\left[\ln Y_{i,PUB1} \mid X_i, PUB_i = 1, CAT_i\right] = \alpha_{PUB1} + X_i'\beta_{PUB1} + CAT_i\delta_{PUB1} + \sigma_{\varepsilon 1}\rho_{\varepsilon 1u}\lambda_{i,PUB} + \sigma_{\varepsilon 1}\rho_{\varepsilon 1\omega}\lambda_{i,CAT}$$

(5)

where λ_{PUB} and λ_{CAT} represent two additional variables that must be included in the earnings equation for the two sectors, in order to correct the estimation of the parameters of interest (the δ coefficients) for the potential bias caused by the double simultaneous selectivity problem described above. Note that if the correlation between the error terms of the two selection rules $\rho_{u,\omega}$ is equal to zero, the λ terms reduce to two independent correction terms, as in Heckman's standard method. On the other hand, if $\rho_{u,\omega} \neq 0$, neglecting the statistical relationship between the selection rules would still lead to inconsistent estimates.

4.1 Identification

In general, standard selectivity models *à la* Heckman require the presence of at least one exclusion restriction to ensure that the parameters are identified not only because of the non-linearity of the selectivity-correction term. This means that at least one variable that appears in the selection equation can be reasonably assumed to be excludable from the outcome equation(s) of the second stage. However, as pointed out by Tunali (1986, pp. 245), the bivariate selectivity model requires additional exclusion restrictions for identifying the correlation coefficient parameter of the error term of the simultaneous selection equations. That is, at least one determinant of each selection process must not be related with the unexplained earnings component. Moreover, for complete identification at least one variable included in the sector choice equation must not influence Catalan knowledge and *vice versa*, and this variable must not appear in the earnings equations.

In the case of the Catalan proficiency equation, like Rendon (2007) we assume that having received schooling in Catalan (i.e. after the 1983 reform) affects knowledge of Catalan but does not directly affect individual earnings. We also assume that after controlling for the years since migration in the earnings equations, having arrived at the age of 10 or younger only affects individual's earnings through Catalan proficiency. Moreover, we consider that language use with the children acts as a determinant of language proficiency and can be excluded from the earnings equation. Finally, regional origins of national immigrants only appear in the Catalan knowledge equation.

In order to identify the sector choice equation, we consider that father's occupation influences sector choice but does not directly affect individual earnings (following Dustmann and van Soest 1998). Moreover, in accordance with Christofides and Pashardes (2002), the sector choice equation contains a variable indicating whether the individual's spouse or partner is an employer (capturing a potential reduction in the cost of job-search in the private sector), and another variable indicating whether the individual perceives income from other sources (non-labour income). The number of children is assumed to influence only sector choice, and does not directly affect earnings. In order to reinforce identification, years of completed schooling are included in the knowledge of Catalan and earnings equations, but completed education is included as a dummy variable in the sector choice equation (Hartog and Oosterbeek 1993, among others, adopt a similar strategy). This reflects that earnings and language knowledge depend on the length of completed studies (human capital view), whereas the sector choice decision is more closely related to the legal value of educational

certificates. Moreover, following the same logic, we also assume that the type of University studies only affects the decision to enter the public sector (and does not directly affect earnings).

5. Results

5.1. OLS Earnings Equations

The analysis of the empirical results starts with the estimation of (1) by simple OLS, which is reported in Table 3; the high R^2 indicates that the covariates included have satisfactory power for explaining the log of monthly earnings. Comparing the estimates across the two sectors we observe that native-born individuals of Catalan origin (i.e. with at least one parent born in Catalonia) earn somewhat less than second-generation immigrants in the private sector; moreover, immigrants from elsewhere in Spain present a clear earnings advantage in both sectors. There is no clear penalization for European immigrants, whereas private sector workers proceeding from Africa, Asia and other countries earn significantly less than native-born immigrants, even accounting for Catalan knowledge. Immigrants who arrived many years ago are paid less than recent immigrants with similar characteristics, but only in the private sector. Females earn less than males, although the earnings gap is significantly lower in the public sector; and, as commonly found in the literature, married individuals tend to earn more than their unmarried counterparts,.

The return to one additional year of schooling is considerably higher in the public sector, while an additional year of job tenure has practically the same impact on monthly earnings in both sectors. Previous experience shows a positive linear effect on earnings in the public sector[12] and an inverse U-shaped effect in the private sector; monthly hours of work has a significant positive effect on earnings, with a higher impact in the public sector. Union members earn more than non-union members, and the earning effect of union membership is significantly higher in the private sector. Among private workers, those who work in a large firm and those who are self-employed earn more than the mean.

---

[12] The coefficient estimate for the quadratic previous experience is not statistically different from zero and it has been dropped from the equation. Its exclusion does not modify the rest of results.

Table 3: OLS Estimations

Dependent Variable: Ln(Earnings)	PUBLIC SECTOR	PRIVATE SECTOR
Constant	5.411 (48.05)	5.772 (128.8)
Native-(Catalan Origin)	*Ref. Cat.*	*Ref. Cat.*
Native-(2nd gen Immigrant)	-0.028 (-0.86)	0.028 (1.81)
Spain	0.089 (1.58)	0.077 (2.50)
Europe	-0.063 (-0.49)	-0.001 (-0.02)
Latin America	-0.081 (-0.82)	-0.032 (-1.14)
Africa	-0.163 (-1.73)	-0.089 (-2.63)
Asia and other countries	0.035 (0.36)	-0.084 (-1.83)
YMS/10	-0.003 (-0.25)	-0.014 (-2.46)
Female	-0.235 (-8.89)	-0.311 (-24.1)
Married	0.069 (2.56)	0.087 (6.32)
Years of Schooling	0.064 (15.2)	0.046 (22.81)
Job Tenure (in months)/10	0.013 (11.75)	0.011 (14.83)
(Previous) Experience/10	0.084 (4.40)	0.131 (6.88)
Experience^2/100	—	-0.022 (-4.82)
Hours of work (per month)/10	0.034 (8.16)	0.025 (16.76)
Union membership	0.083 (2.42)	0.116 (8.22)
#Workers>500	—	0.055 (3.33)
Self-Employed	—	0.060 (2.85)
Living in Barcelona	0.016 (0.55)	0.023 (1.59)
Proficiency in Catalan	0.116 (2.70)	0.064 (3.78)
R^2	0.51	0.42
Nº. Observations	829	4.190

Note: Robust Standard Errors; t-Statistics in Parentheses.

Moreover, proficiency in Catalan has a significant and positive effect on monthly earnings in both sectors and, consistent with the descriptive evidence presented above, the return to knowledge of Catalan seems to be higher in the public sector; indeed, point-estimates indicate a 6.5% (= $exp(\delta)-1$) return in the private sector, whereas the language premium for public

14

workers is reflected in extra earnings of 12%. However, as noted above, these OLS estimates may be seriously biased. One possible source of bias is the potential non-randomness of the mechanism that allocates workers in the public or the private sector. Unobserved individual heterogeneity may represent another source of bias, if individuals opt to learn Catalan on the basis of their unobservable attributes — potentially related to unexplained earnings components. Finally, a third source of bias may be the correlation between the unobservable determinants of the two selection mechanisms (Catalan proficiency and sector choice).

5.2. Bivariate Selection Equations

In order to deal with these multiple sources of bias, we implement the double simultaneous selection correction with the methodology presented in the previous section. The first step is the joint estimation of the selection equation (2a) and 2(b) to explain Catalan proficiency and sector choice respectively. Table 4 shows the maximum likelihood estimates of the resulting Bivariate Probit.

The results of the estimation of the knowledge of Catalan equation indicate that females are somewhat more likely to speak and write Catalan than males with similar characteristics. As expected, the propensity to be proficient in Catalan decreases with age, indicating that older individuals have more difficulty in assimilating the language. Second-generation immigrants are clearly less likely to speak and write Catalan than native-born individuals of Catalan origin. Individuals born outside Catalonia are also clearly penalized, except for those from eastern Spain (Valencia and Balearic Islands); this result is no surprise, since Catalan is also spoken in these regions of Spain (even though it is less institutionalized). Moreover, the disadvantage is even higher for those individuals who were born outside Spain, especially for Latin American immigrants; in all likelihood, this is because their mother-tongue is Spanish, and the incentives for learning Catalan are lower for them (*ceteris paribus*). However, the positive and statistically significant coefficient for time in Catalonia (years since migration) indicates that a longer exposure to the local language favours its assimilation[13].

---

[13] Even so, the negative sign of the interaction (Born in Spain)×YSM indicates that immigrants from the rest of Spain are less likely to be proficient in Catalan as the length of their stay increases; therefore, the advantage of individuals who were born in Spain with respect to foreigners decreases with time spent in Catalonia. This shows that individuals who came from the rest of Spain in the past may have had fewer incentives to learn Catalan, since they may well have arrived when the use of this language was still restricted to oral communication.

Table 4: Bivariate First-Stage Estimation (Catalan Proficiency/Sector Choice) — Robust Standard Errors

Dependent Variable: Being Proficient in Catalan (0-1)	Coefficients	z-Statistic
Constant	-0.73	-0.38
Female	0.168	3.42
Age/10	-0.235	-6.89
Native-(Catalan Origin)	ref. cat.	—
Native-(2nd gen Immigrant)	-0.610	-10.02
Eastern Spain	-0.351	-1.47
Southern Spain	-1.430	-10.74
Central Spain	-1.404	-10.72
North-West Spain	-1.184	-6.08
North-East Spain	-1.262	-6.96
Europe	-2.181	-10.70
Africa	-2.051	-12.29
Latin America	-2.665	-17.35
Asia and other countries	-2.406	-10.86
YSM/10 (= 0 for Natives)	0.215	3.34
(Born in Spain)×YSM	-0.019	-2.84
Years of Education	0.137	19.37
Arrival at age 10 or younger	0.478	4.29
Partial Normalization	0.584	7.25
Complete Normalization	0.851	7.78
Speaks Catalan with children	0.654	10.98
Dependent Variable: Working in the public sector (0-1)		
Constant	-2.832	-8.68
Female	0.213	4.56
Number of Children	0.023	0.92
Age/10	0.531	3.36
Age^2/100	-0.049	-2.67
Native	ref. cat.	—
Spain	0.017	0.27
Europe	-0.588	-2.81
Extra-Europe	-0.535	-5.59
No Education – Primary Education	ref. cat.	—
Lower-Secondary Education	-0.029	-0.28
Vocational Secondary Education	0.304	3.09
General Upper-Secondary Education	0.336	3.31
Tertiary Education	1.399	13.46
University: Humanities	ref. cat.	—
University: Exact Sciences	-0.667	-6.55
University: Social Sciences	-0.632	-7.13
Partner is Employer	-0.325	-2.81
Father Skilled White-Collar	0.174	2.20
Father Skilled Blue-Collar	0.223	2.57
Non-labour income	0.104	2.27

$P_{u\omega} = 0.231$; $\chi^2[H_0: \rho_{u\omega} = 0] = 33.14$ (P-Value = 0.000)

Schooling is clearly one of the most important determinants of the probability of speaking and writing Catalan, with a positive and highly significant estimated coefficient. As found by Rendon (2007), linguistic assimilation is easier for immigrants who arrived at a young age (even controlling for the years since migration); moreover, individuals affected by the 1983

language legislation are more likely to be able to speak and write Catalan, with a stronger effect for those who were schooled entirely in Catalan after the 1983 reform. Our results also show that the individual's environment plays an important role in explaining the chances of achieving language proficiency. In fact, use of the language with the children[14] significantly increases the probability of speaking and writing Catalan.

The estimates of the sector choice equation reveal that females are significantly more likely to work in the public sector than males. As commonly found in the literature, the number of children does not significantly affect the probability of working in one sector or in the other. The likelihood of being selected in the public sector increases with age but at a decreasing rate; foreigners are less likely to enter the public sector, and those who were born in Europe are even less likely to do so than their non-European immigrants.

Individuals with a post-compulsory education certificate have a higher chance of working in the public sector than individuals with lower-secondary education or less; among tertiary educated workers, those who studied exact sciences or social sciences at University are significantly less likely to work in the public sector than those who studied humanities. As expected, the relative cost of searching for a public sector job is higher for those individuals whose partner is an employer; moreover, being the child of a skilled white-collar or skilled blue-collar father increases the likelihood of having a public occupation. Surprisingly, those individuals who have some non-labour income are more likely to work in the public sector.

Finally, the correlation coefficient $\rho_{u\omega}$ is positive and statistically different from zero, which means that the two selection rules are not independent. Specifically, this result indicates that individuals who are more likely to be proficient in Catalan are also more likely to work in the public sector and *vice versa*; notice that in this Bivariate Probit model the positive relationship between the two selectivity mechanisms is indirect — i.e. it is captured by the correlation between the unobservable of the two equations. The evidence of significant correlation between the disturbances of the selection equations also suggests that joint estimation provides more efficient results than independent estimation of the two selection rules. In addition, it implies that this correlation should be taken into account in order to obtain a consistent estimate of the return to knowledge of Catalan in each sector, because controlling for the two selectivity rules assuming that they are independent may not entirely eliminate selection bias(es).

---

[14] We also tried to include the number of children, but its coefficient is not statistically different from zero (a common result in the literature); moreover, including or excluding this variable does not modify the overall results obtained.

5.3 Selectivity-Corrected Earnings Equations

As noted above, the OLS estimation of the language return in the private and public sectors could be biased on the one hand by the non-randomness of the allocation of workers into one sector or another, and on the other hand by the self-selection into knowledge of Catalan. We should also take into account the positive correlation between these two selectivity rules, as suggested by the previous results. In order to obtain a consistent estimate of the δ parameters for each sector, we implement the bivariate selection correction as presented in section 4. Under the assumption of validity of the identification conditions, Table 5 reports the bivariate selectivity-corrected earnings equations for public and private sector workers, which contain the consistent estimates of the return to knowledge of Catalan in the two sectors. The bivariate estimation of the two selection rules enables us to construct the selectivity-correction terms in (4), which have been inserted into the earnings equation as additional regressors (eq. 5). Notice that the coefficients' standard errors have been obtained through bootstrapping[15], given that the calculation of the correct variance-covariance matrix obtained by Ham (1982) and by Tunali (1986) is cumbersome.

In general, the estimated coefficients for both sectors are roughly identical to those estimated through OLS, and we will not describe them again for brevity reasons. Even so, we observe some interesting differences with respect to the OLS estimates, which appear to be worth analysing in more detail. The minor changes in the earnings equation estimates are consistent with the reduction in the return to schooling and potential previous experience when estimated with the double simultaneous selection correction.

Above all, the significant changes concern the estimation of the return to knowledge of Catalan in the public and private sectors. Specifically, the apparently higher return to language knowledge in the public sector estimated by OLS seems to be composed entirely by selection-bias effects. In fact, when we take into account the selection process behind sector choice and language proficiency and the correlation between the two selectivity mechanisms, the return to Catalan knowledge for public sector workers is nearly zero. In contrast, the estimated return to Catalan proficiency for private sector workers is significantly higher when the two simultaneous sources of selection are taken into account, representing almost 12% ($\approx$ *exp(δ)-1*) of extra monthly earnings[16].

---

[15] Specifically, we display the t-Statistics obtained with the Bias-Corrected standard errors, which have been computed with 1000 replications.

[16] Estimating the return to Catalan proficiency with two independent correction terms yields similar results in qualitative terms but, for both sectors, the estimated coefficients are somewhat higher than when we control for the correlation between the two selection rules (the results are not shown and are available upon request). This means that, to some extent, a part of

Table 5: Selectivity-Corrected Estimations

Dependent Variable: *Ln(Earnings)*	PUBLIC SECTOR	PRIVATE SECTOR
Constant	5.864 *(24.96)*	5.83 *(95.6)*
Native-(Catalan Origins)	*Ref. Cat.*	*Ref. Cat.*
Native-(2nd gen immigrant)	-0.041 *(-1.21)*	0.039 *(2.29)*
Spain	0.041 *(0.61)*	0.094 *(2.46)*
Europe	0.044 *(0.31)*	0.082 *(1.32)*
Latin America	-0.123 *(-0.90)*	0.059 *(1.23)*
Africa	-0.186 *(-1.59)*	-0.030 *(-0.65)*
Asia and other countries	0.007 *(0.06)*	-0.005 *(-0.09)*
YMS/10	0.001 *(0.01)*	-0.014 *(-2.38)*
Female	-0.254 *(-8.61)*	-0.336 *(-23.93)*
Married	0.067 *(2.40)*	0.085 *(6.00)*
Years of Schooling	0.054 *(6.10)*	0.033 *(9.53)*
Job Tenure (in months)/10	0.012 *(9.56)*	0.010 *(11.91)*
(Previous) Experience/10	0.059 *(2.76)*	0.117 *(5.88)*
$Experience^2/100$	—	-0.021 *(-4.62)*
Hours of work (per month)/10	0.035 *(8.10)*	0.025 *(16.94)*
Union membership	0.072 *(2.21)*	0.113 *(7.89)*
#Workers>500	—	0.057 *(3.50)*
Self-Employed	—	0.066 *(2.20)*
Living in Barcelona	0.016 *(0.54)*	0.020 *(1.40)*
Proficiency in Catalan	0.001 *(0.01)*	0.117 *(2.34)*
σ_ε	0.357	0.382
$\rho_{\varepsilon u}$	0.137 *(0.79)*	-0.144 *(-1.66)*
$\rho_{\varepsilon\omega}$	-0.315 *(-1.68)*	-0.805 *(-4.77)*

Note: Bootstrapped Standard Errors (1000 replications).
Derived t-Statistics appear in italics in brackets.

Notice also that the correlation coefficients between the unexplained earnings component and the error term of the sector choice equation are negative in both equations. This shows

---

the positive effect of knowledge of Catalan is captured by its correlation with sector choice, which must be taken into account in order to obtain an unbiased estimate of the true value of knowledge of Catalan for private and public workers.

that an individual who is selected for work in the public sector performs worse than a random individual. However, the estimated correlation coefficient is clearly statistically significant only in the private sector equation, and only slightly significant in the public sector equation (probably due to the reduced sample size). Moreover, the correlation between the earnings equation's error term and the unobservable determinants of Catalan proficiency is positive for public sector workers and negative for private sector workers. This could indicate that the higher OLS language return in public sector may only reflect the fact that those public workers who are more likely to be proficient in Catalan are also more likely to earn more, and are also more likely to be allocated in that sector; nevertheless, we do not have sufficient statistical evidence to argue that this correlation is different from zero for public sector workers. In contrast, the significant correlation between the unobservable earnings determinants and unexplained earnings is negative in the private sector, suggesting that those individuals whose propensity to know Catalan is largely determined by unobservable determinants of language knowledge earn less than the mean private sector worker.

6. Discussion and Conclusion

This paper investigates the economic value of knowledge of Catalan for private and public sector workers in Catalonia. The descriptive evidence and the results from a simple OLS estimation indicate that, apparently, the earnings return to being able to speak and write Catalan (our measure of linguistic proficiency) is positive in both sectors, but is significantly higher for public workers. However, in accordance with the main literature, we argue that both knowledge of Catalan and the decision to work in the public or in the private sector are choice variable; this represents a double selectivity process, which must be taken into account in order to obtain a consistent estimate of the return to Catalan proficiency in the two sectors. In addition, in accordance with the Catalan institutional setting, we enable the potential correlation of these two selection rules, which we control for by implementing a double simultaneous selection correction of the earnings equations.

Once this complex self-selection process is taken into account, the results are completely different and are consistent with our *ex-ante* expectation. Specifically, on the one hand, the return to knowledge of Catalan is virtually zero for public workers when we control for selection on observable and unobservable into proficiency and sector choice, and for the significant positive correlation between the unobservable determinants of the two selectivity

rules. On the other hand, when allowing for the double simultaneous selection process the return to Catalan proficiency in the private sector is still positive and is almost double the OLS estimate (rising from 6.5% to 12% of extra monthly earnings). These results suggest that there is no productivity effect of knowledge of Catalan among public sector workers, and that the positive economic value of language proficiency consists only in a higher chance of being selected into that sector. In contrast, Catalan proficiency seems to increase productivity for private sector workers (assuming the correspondence between earnings and productivity), given that we obtain a positive earnings premium even controlling for the double simultaneous selection. In more detail, we found that the return to language proficiency in the private sector is underestimated by OLS because of the presence of negative selection effects; first, the propensity to be proficient in Catalan and the likelihood of working in the private sector are clearly negatively correlated. Second, the private sector workers who are more likely to be selected in the public sector perform worse than a random private sector worker. Third, those individuals who are more likely to be proficient in Catalan (keeping the observable determinants of language knowledge fixed) tend to earn less than a random individual. Especially with respect to the third point, it is quite likely that this negative selection of proficient workers may operate through occupational choices — i.e. private workers who are more likely to be proficient in Catalan because of their unobserved language determinants are also more likely to be selected into low-paid occupations than others.

Indeed, potential caveat of this work is that it neglects the role played by the type of occupation and its interrelation with language proficiency, in the spirit of Aldashev et al. (2009). However, we consider that occupation-type selection is a less relevant issue for estimating the return to knowledge of Catalan, because we believe and assume that education, and not Catalan proficiency, is the main channel for entering high-skill occupations. In other words, highly educated individuals may manage to enter highly-paid occupations with or without being fluent in Catalan (e.g. in multinational firms where English is the main language spoken). In contrast, low-educated individuals are precluded from entering highly-paid and high-skill occupations, regardless of their functional knowledge of Catalan. Even so, occupational components may account for some part of the estimated productivity effect for private sector workers; therefore, extending the selection process to a potential occupational selection for private sector workers would be an interesting issue for future research into the economic value of knowledge of Catalan in the labour market.

In any case, the global results show the existence of a positive economic value of the knowledge of Catalan. Even though we still need to clarify whether the positive estimated

value for private sector workers corresponds to a productivity effect or to an occupational effect, it is clear that knowledge of Catalan only represents a selection effect in the public sector — i.e. Catalan proficiency does not increase productivity of public sector workers. Definitely, this result questions the strict regulation and the high requirements of knowledge of Catalan in the Catalan public sector. On the one hand, it seems that after accounting for self-selection linguistic proficiency is not associated with higher productivity of public sector workers, and merely represents a requirement for being hired in that sector. On the other hand, the results also indicate that the probability of being proficient in Catalan is strongly related to individual characteristics, which are also related to labour market success (e.g. age, origins, education, etc.). This means that, in all likelihood, many disadvantaged individuals are prevented from being able to speak and write Catalan because of the same characteristics that tend to penalize them in the labour market. As a consequence, the strict regulations on language requirements for entering the public sector represent a clear barrier to them, and may be responsible for some discrimination in the labour market. In terms of policy implications, it is quite possible that lowering the linguistic requirements for working in the public sector may generate a positive effect, at least in terms of equity; this is especially true if we consider the historical role of public sector occupation in Mediterranean countries as a social safety net.

References

Adamchik, V. A. & Bedi, S.A. (2000). "Wage Differentials between the Public and the Private Sectors: Evidence from an Economy in Transition," Labour Economics, vol. 7, pp. 203–24.

Alarcón, A. & Solé, C. (2001). "*Llengua i Economia a Catalunya*", Institut d'Estudis Catalans.

Aldashev, A., Gernandt, J. & Thomsen, S.L., (2009). "Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection," Labour Economics, vol. 16, pp. 330-341.

Bender, K. (2003). "Examining equality between public – and private sector wage distributions", Economic Inquiry, vol. 41, pp. 62-79.

Berman, E., Lang, K. & Siniver, E. (2003). "Language-skill complementarity: returns to immigrant language acquisition," Labour Economics, vol. 10, pp. 265-290.

Chiswick, B. (2008). "The Economics of Language: An Introduction and Overview," IZA Discussion Papers 3568, Institute for the Study of Labor (IZA).

Chiswick, B. & Miller, P.W. (2007). "*The Economics of the Language: International Analyses*"; Routledge editors.

Chiswick, B. & Miller, P.W. (2010). "Occupational language requirements and the value of English in the US labor market". Journal of Population Economics, vol. 23, pp. 353-372.

Christofides, L. N. & Pashardes, P. (2002). "Self/paid-employment, public/private sector selection, and wage differentials," Labour Economics, vol. 9, pp. 737-762.

Di Paolo, A. & Raymond, J.L. (2010). "Language knowledge and earnings in Catalonia," Working Papers XREAP2010-7, Xarxa de Referència en Economia Aplicada (XREAP), revised Jul 2010.

Dustmann, C. & van Soest, A. (1998). "Public and private sector wages of male workers in Germany," European Economic Review, vol. 42, pp. 1417-1441.

Dustmann, C. & van Soest, A. (2001). "Language Fluency And Earnings: Estimation With Misclassified Language Indicators"; The Review of Economics and Statistics, 83, pp. 663-674.

Dustmann, C. & van Soest, A. (2002). "Language and the earnings of immigrants," Industrial and Labour Relations Review, vol. 55, pp. 473-492.

Fishe, R., Robert, P.H., Trost, P. & Philip M. Lurie (1981). "Labour force earnings and college choice of young women: An examination of selectivity bias and comparative advantage", Economics of Education Review, vol. 1, pp.169-191.

Ham, J.C. (1982). "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment". The Review of Economic Studies, Vol. 49, pp. 335-354

Hartog, J., & Oosterbeek, H. (1993). "Public and private-sector wages in the Netherlands", European Economic Review, vol. 37, pp. 97-114.

Heckman, J. (1979). "Sample Selection Bias as a Specification Error"; Econometrica, vol. 47, pp.153-162.

Heitmueller, A. (2006). "Public-private sector pay differentials in a devolved Scotland," Journal of Applied Economics, Universidad del CEMA, vol. 0, pp. 295-323.

Lang, K. & Siniver, E. (2009). "The Return to English in a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel," The B.E. Journal of Economic Analysis & Policy, vol. 9, 50.

Lee, L., (1978). "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." International Economic Review, vol.19, pp. 415-433.

Quella, N. & Rendon, S. (2008). "Occupational Selection in Multilingual Labor Markets," Department of Economics Working Papers 09-02, Stony Brook University, Department of Economics.

Rendon, S. (2007). "The Catalan premium: language and employment in Catalonia," Journal of Population Economics, vol. 20, pp. 669-686.

Tunali, I. (1986). "A general structure for models of douple-selection and an application to a joint migration/earnings process with remigration", Research in Labor Economics, vol. 8 B, pp. 235-283.

van der Gaag, J., & Vijverberg, W. (1988). "A switching regression model for wage determinants in the public and private sector of a developing country", Review of Economics and Statistics vol. 70, pp. 244-252.

APPENDIX

Table 1A: Monthly Earnings in Brackets (in Euros) — Overall Sample

EARNING INTERVALS (in Euros)	FREQ.	% SAMPLE
Less than 450	111	4.30
Between 451 and 600	168	6.51
Between 601 and 900	672	26.03
Between 901 and 1050	372	14.41
Between 1051 and 1200	417	16.15
Between 1201 and 1500	373	14.35
Between 1501 and 1800	212	8.21
Between 1801 and 3000	221	8.56
More than 3000	36	1.49

Definition of constructed variables:

Age = *mid points of the original age variable collected in intervals (16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65).*

YSM = *2005 minus year of arrival in Catalonia; year of arrival in Catalonia collected in intervals (before 1940, between 1941 and 1945, 1946-1950, 1951-1955, 1956-1960, 1961-1965, 1966-1970, 1971-1975, 1975-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005).*

Arrived at 10 or younger* = *1 if the individual arrived in Catalonia at age 10 or younger, 0 otherwise.*

Complete Normalization* = *1 if the individual was born after 1977, and arrived at age 6 or younger if born outside Catalonia, 0 otherwise.*

Partial Normalization* = *1 if the individual was born between 1969 and 1977 and arrived younger than age 16 if born outside Catalonia, 0 otherwise.*

Years of Schooling = *4 if the individual has no education, or if he/she has completed only primary education (grouped in the original database); 7 for lower-secondary uncompleted; 8 for lower-secondary completed; 12 for vocational education; 13 for general upper-secondary education; 18 for completed tertiary education.*

Job Tenure (in months) = *mid point of the original variable collected in intervals (< 2 years, 2 - 5 years, 5 - 10 years, 10 - 15 years, 15 - 20 years, > 20 years).*

Previous Experience = *potential work experience, previous to the current work (age-schooling-(job-tenure (in months))/12 -6).*

Hours of work (per month) = *mid point of the original variable (hours per week) collected in intervals times 4 (0-20, 20-30, 30-35, 35-40, 40-45, 45-55, 55-100).*

*Variables constructed by IDESCAT staff, from the original registers of the survey (maximum desegregation).

Table 2A: Descriptive Statistics

VARIABLE	PUBLIC SECTOR		PRIVATE SECTOR	
	MEAN	S.D.	MEAN	S.D.
SOCIO-DEMOGRAPHIC CHARACTERISTICS				
Age	40.92	10.42	38.72	11.12
Female (0-1)	0.565	0.496	0.404	0.491
Married (0-1)	0.590	0.492	0.579	0.494
Number of children	1.201	1.113	1.146	1.159
INDICATORS FOR COMPLETED EDUCATION				
No Education – Primary Education	0.058	0.234	0.138	0.345
Lower-Secondary Education	0.103	0.304	0.261	0.439
Vocational Secondary Education	0.186	0.389	0.255	0.436
General Upper-Secondary Education	0.134	0.341	0.173	0.378
Tertiary Education	0.520	0.500	0.173	0.379
University: Humanities	0.302	0.459	0.054	0.226
University: Exact Sciences	0.083	0.276	0.049	0.216
University: Social Sciences	0.128	0.334	0.069	0.254
Years of Education	14.52	3.991	11.41	3.805
INDICATORS FOR INDIVIDUAL ORIGINS				
Native-(Catalan Origins)	0.579	0.494	0.467	0.499
Native-(2nd gen immigrant)	0.189	0.392	0.213	0.409
Spain	0.177	0.382	0.168	0.374
Eastern Spain	0.017	0.129	0.008	0.087
Southern Spain	0.060	0.238	0.088	0.284
Central Spain	0.053	0.224	0.046	0.210
North-WesternSpain	0.008	0.092	0.012	0.108
North-Eastern Spain	0.039	0.193	0.014	0.117
Europe	0.010	0.098	0.023	0.151
Africa	0.017	0.129	0.037	0.189
Latin America	0.016	0.124	0.066	0.248
Asia and other countries	0.012	0.109	0.026	0.160
OTHER VARIABLES				
YSM (= 0 for native-born individuals)	7.572	18.34	8.313	19.07
Arrived Younger than 10 (0-1)	0.080	0.271	0.080	0.272
Job Tenure (in months)	158.2	136	103.5	114.5
(Potential) Years of Previous Experience	8.36	9.097	13.00	10.63
Hours of work (per month)	136.3	39.45	159.5	52.44
Union membership (0-1)	0.811	0.392	0.360	0.480
#Workers>500 (0-1)	0.663	0.473	0.201	0.401
Self-Employed (0-1)	—	—	0.179	0.384
Speak Catalan with Children (0-1)	0.448	0.498	0.320	0.466
Complete Normalization (0-1)	0.139	0.346	0.183	0.387
Partial Normalization (0-1)	0.197	0.398	0.225	0.417
Couple Employer (0-1)	0.035	0.184	0.040	0.196
Father Skilled White-Collar (0-1)	0.115	0.319	0.064	0.246
Father Skilled Blue-Collar (0-1)	0.104	0.305	0.049	0.215
Non-labour income (0-1)	0.332	0.471	0.270	0.444
Living in Barcelona (0-1)	0.258	0.438	0.255	0.436