Departament d'Economia Aplicada

Language knowledge and earnings in Catalonia

Antonio Di Paolo,
Josep Lluís Raymond-Bara

DOCUMENT DE TREBALL

10.01

Universitat Autònoma de Barcelona

Facultat de Ciències Econòmiques i Empresarials

LANGUAGE KNOWLEDGE AND EARNINGS IN CATALONIA

Antonio Di Paolo[a], Josep Lluís Raymond-Bara[b]

[a,b] *Universitat Autònoma de Barcelona, Institut d'Economia de Barcelona*

Abstract

In this paper we are aimed to investigate the relationship between Catalan knowledge and individual earnings in Catalonia. Using data from 2006, we find a positive earning return to Catalan proficiency; however, when accounting for self-selection into Catalan knowledge, we find a higher language return (20% of extra earnings), suggesting that individuals who are more prone to know Catalan are also less remunerated than others (negative selection effect). Moreover, we also find important complementarities between language knowledge and completed education, which means that only more educated individuals benefit from Catalan knowledge.

Keywords: Language, Earnings, Self-Selection, Skill-Complementarity, Catalonia

JEL Classifications: J79, J24, J61, C31

[a]Corresponding author: 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. [b] josep.raymond@uab.cat.

1.   Introduction

This paper aims to investigate the relationship between Catalan knowledge and labour market success in Catalonia. Specifically, we want to check whether there exists an expected earning increase associated with Catalan proficiency, accounting for the determinants of self-selection into Catalan knowledge. Moreover, we are also moved to understand whether Catalan knowledge has an economic value in terms of earnings if it is only accompanied by a certain level of skills; namely, we analyze whether language and education are complementary inputs for generating earnings in Catalonia.

Since the late 70's ―with the end of the fascist dictatorial regime (1939-1975)―, Catalan is a co-official language in Spain, together with Castilian (more commonly known as Spanish). The several public linguistic policies implemented by the Catalan Autonomous Government (the so called *Generalitat de Catalunya*), aimed to stimulate the recovery of the local language and culture, may have also contributed to the increase of the economic value of being fluent in Catalan. The first important attempt to promote the use of Catalan was the "Linguistic Normalization Policy" of 1983 (*Llei de Normalització Lingüística*). This law established not only that Catalan was to be the

official language of the Catalan Government and of the Local Public Administrations, but also the main language used in primary and secondary education[1].

Nevertheless, what at the present time is still more relevant for the economic value of Catalan knowledge is the subsequent "Linguistic Policy Act" of 1998 (*Llei de Política Lingüística*), which attempted to further regain the presence of Catalan (versus Castilian), especially in the socioeconomic and cultural[2] domains, and reinforce its use in the public sphere. Indeed, the potential economic value of Catalan proficiency substantially increased with the implementation of this policy, as the Catalan Government a) increased fluency requirements for public sector employees, and b) introduced important incentives, and also requirements, for the improvement of Catalan knowledge in the private sector (Solé and Alarcón 2001). In general, if Catalan knowledge has currently some economic value, it would be (also) because these linguistic policies may have had important socioeconomic implications. Therefore, individuals with a high proficiency in Catalan could have some advantage in the labour market (considering that everyone is proficient in Castilian[3]).

As a general framework for analyzing the value of Catalan knowledge in the labour market, we refer to research on the earning returns to language knowledge, with the pioneer studies of Chiswick (1992), and Chiswick and Miller (1995, 1999). These authors consider that language constitutes another form of human capital, as well as education and labour market experience. This is because language knowledge is likely to be embodied in the person, it is productive in the labour market and it is obtained with individual effort and out-of pocket time and economic resources. But what is still more important is that Chiswick and Miller explicitly consider the potential endogeneity between earnings and language proficiency: specifically, they claim that the (positive) correlation between language fluency and earnings (among immigrants) may only indicate that more able workers are more likely to earn more and speak better their host-country language. They try to correct for this potential endogeneity using

---

[1] In fact, Spanish (or Castilian) was taught as a second language in pre-university education. However, the language used in university courses was not determined by the law, and it is discretionally established by the professor.

[2] Moreover, the Catalan Government tried to stimulate the diffusion of the language through the mass media by introducing incentives in the use of Catalan on radio, TV and also in the newspapers or general written publications.

[3] That is, we can claim that natives and national immigrants are fully proficient in Castilian. Moreover, non-national immigrants tend to consider learning Castilian as more important than learning Catalan (or simply a priority), probably due to the international value of the former. Therefore, we can also assume that those foreigners who are fluent in Catalan are also fluent in Spanish (in every case, we take into account Castilian knowledge for non-national immigrants).

instrumental variables and sample selection correction methods, finding positive (but unstable) effects of language proficiency among immigrants.

The same framework can be applied to the Catalan case. However, Catalan proficiency may not only have an economic return for national and non-national immigrants, but also for the native population[4]. In fact, during the 60's, Catalonia received high flows of (permanent) national immigrants coming from other Spanish regions. Given the subsequent coexistence of Castilian and Catalan, many national immigrants may have not achieved a high level of fluency in Catalan (due to a lack of incentives), when this language was recognized as a co-official one. But what is still more important is that a relevant proportion of native-born are national immigrants of second generation. Thus, their mother-tongue is Castilian and their knowledge of Catalan may be limited (particularly in the case of those who have been schooled before the linguistic normalization of 1983). Finally, in the last years, Catalonia has experienced new immigration flows, proceeding from other countries in the world, especially from less-developed countries (Fernández-Huerta and Ferrer-i-Carbonell 2007). Even if many of them are coming from Spanish-speaking countries (from Latin America), Catalan is a new language for immigrants and Catalan fluency is often considered as a possible source of cultural and economic assimilation for them.

The first study focusing on the case of Catalonia is the paper by Rendon (2007), which analyzes the potential premium of having Catalan knowledge on the probability of being employed. Using Census data of 1991 and 1996, he found that being proficient in Catalan (both speaking and writing) has a positive effect on the probability of being employed among native and national immigrants. The analysis takes into account the presence of self-selection into Catalan knowledge, using as instruments externality effects of the place of residence, origin variables, and a variable indicating whether the individual was affected by the Linguistic Normalization Act during childhood. Unfortunately, as the author recognizes, this study is limited to language return on employment probabilities, since Census data does not contain information about earnings. Moreover, it is possible that the modest impact that he found (between 2 and 5% in the case of speaking, and between 3 and 6% in the case of writing), is related to

---

[4] Chiswick and Miller (2007), among others, consider the issue of limited language knowledge of immigrants of the second generation (children of immigrant parents). Other studies (see Lindley 2002, for instance) consider the effect of bilingualism (with respect to local or disglossic language), which seems to be having negative effects on earnings. Nevertheless, it is not clear whether (within the current institutional framework) Catalan might be still considered a disglossic language.

the fact that this data is limited to 1996, that is before the implementation of the Linguistic Policy Act (1998). Thus, posterior data may reflect that this act has potentially increased the economic value of Catalan knowledge.

Many other recent studies are focused on obtaining an (unbiased) estimate of the effect of language proficiency on earnings. As commented before, there is a general concern that an OLS estimate of the effect of language knowledge on earnings could be biased, due to the presence of unobserved individual heterogeneity (which may affect both language fluency and economic outcomes). Moreover, several authors suggest that misclassification error represents another potential source of bias in the earning equation, given that (as in this case) language knowledge measures are usually self-reported on a categorical scale.

For example, in order to account for misclassification error and unobserved heterogeneity, Dustmann and Van Soest (2001) exploited a panel structure of the GSOEP, whereas (Dustmann and Van Soest 2002), combine a panel estimation with IV techniques[5]. In a subsequent study, Dustmann and Fabbri (2003) combine the use of the interview language as an instrument for the measurement error in language proficiency, and the propensity score matching for correcting the potential individual heterogeneity bias. They confirm that the two source of bias operate in opposite directions, but their final results are inconclusive.

An interesting contribution is the paper by Bleakley and Chin (2004): they propose an interesting IV approach, based on the idea that younger immigrants easily learn their host-country language (English), but immigrants for English-speaking countries do not need to learn any new language. They suggest that the OLS estimates are slightly upward biased by the endogeneity between earnings and language knowledge, and downward biased by measurement error. Moreover, they argue that an important component of language proficiency effects on earnings is mediated by education; specifically, young immigrants who are more likely to be proficient in English are also more prone to achieve more education, thus increasing their earning opportunities.

---

[5] Specifically, Dustmann and Van Soest (2002) exploit lagged self-reported language fluency (for time-independent measurement error), partner and household characteristics (for unobserved heterogeneity) and parental education (for time-persistent measurement error). These authors suggest that language knowledge has a positive earning effect among immigrants (in Germany), and also that the negative bias of measurement error overcomes the positive bias of unobserved heterogeneity.

In this paper we use data from the 2006 wave of the "Survey on Living Conditions and Habits of the Catalan Population" (*Enquesta de Condicions de Vida i Hàbits de la Població*, ECVHP06), carried out by the Statistical Institute of Catalonia (IDESCAT)[6]. This database is very appealing for our purposes, because it contains socioeconomic information (including individual monthly earnings) and it includes information related to language knowledge and language usage. However, the cross-sectional structure of the ECVHP06 survey limits the possibilities of accounting for misclassification into Catalan knowledge (which is self-reported). In order to limit the potential bias, we adopt a stringent definition of language proficiency, considering as proficient only those individuals who are able to speak and write Catalan. We only focus on individuals born outside Catalonia and on native-born with non-Catalan parents (immigrants of second generation), as we consider that natives with Catalan origins are fully proficient in Catalan simply because this language is transmitted within the family[7]. Moreover, we opt for estimating the earning return to Catalan proficiency using the technique originally proposed by Heckman (1979), which allows us controlling for the self-selection process behind language knowledge.

We consider that this methodology is particularly appropriate for estimating the language return in Catalonia. In general, many of the contributions previously reviewed are focused on correcting for unobserved heterogeneity caused by individual innate ability. The argument of individual unobserved ability makes sense in these studies, given that the primary aim consists in analyzing whether those immigrants, who are proficient in their host-country language, earn more than those who are not proficient. In such cases, learning the destination country's language is like a "need" for immigrants; therefore, unexplained differences in language fluency among them may be principally related to an unobserved ability.

On the contrary, we believe that the endogeneity between Catalan proficiency and earnings could be caused by a more complex self-selection process behind the fact of knowing Catalan[8], where innate ability might only play a secondary role. Given the co-

---

[6] We are very grateful to IDESCAT for providing the data; in any case, we disclaim IDESCAT any mistake or opinion expressed in this paper, which are sole responsibility of the authors.
[7] On the contrary, for individuals with non-Catalan parents achieving proficiency in Catalan it represents an investment in language skills. To some extent, our objective consists in quantifying the earnings return to such investment.
[8] As we explain below, the individuals' propensity to being proficient in Catalan depends on many elements of the daily life. Even if in a "reduced form", we estimate this decision process in the first stage of the Heckman's two-step procedure, we think that in this case modelling the process behind Catalan knowledge has a special intrinsic interest.

officiality with Castilian, and the absence of a legal requirement[9], knowing Catalan is closer to a choice than to a requirement (which is true not only for natives, but also for the immigrant population). In fact, Catalan in Catalonia does not play the same function than English in the US, for instance, where an immigrant from Latin America in the US might learn English in order to find a job, but also for developing a normal daily life. On the contrary, an individual coming from Latin American can speak Castilian in Catalonia without any problem, given the co-officiality of these two languages.

Nevertheless, we suspect significant complementarities exist between Catalan fluency and individual skills. For example, more educated individuals may face lower costs of achieving proficiency. Moreover, given the Catalan institutional setting, only more educated individuals may obtain such occupations where Catalan language is really important. Therefore, more skilled workers may also have higher benefits, which imply higher returns to language fluency. Several studies confirm this intuition in other contexts; Carliner (1996) and Mora and Davila (1998) suggested the idea of non-constant language returns across skill levels, finding that the wage penalty of limited language proficiency is higher for more educated workers.

Berman et al. (2003) ruled out unobserved heterogeneity bias from the language augmented earning regression through differentiation with panel data. They suggest that individual fixed-effects (interpreted as unobserved individual ability) have a small effect on language returns in high skilled occupations. However, the small positive cross-sectional return for low-skilled workers turns out to be insignificant when introducing individual fixed-effects (that is, the estimated language effect for low-skilled workers is entirely accounted by the ability bias). Moreover, in the recent contribution of Lang and Siniver (2009), the authors confirm the previous findings of complementarity between language and skills, obtaining a higher return to language knowledge among high-educated workers (and an insignificant return for low-educated workers), in both cross-sectional and longitudinal estimates.

According to this recent literature, in this paper we also check for potential complementarities between Catalan knowledge and completed education. Particularly,

---

[9] Apart from those individuals who have been educated in Catalan, after the implementation of the Linguistic Normalization Act of 1983. Even so, there are many young individuals who were schooled in Catalan, but are more used to speak in Castilian in their daily life. However, as established by the above mentioned language acts, Catalan proficiency is required for working in the Local Public Administration.

we repeat the estimations dividing the sample between low-educated (individuals with no education or with primary education), and medium-high educated (individuals with secondary or tertiary education). Additionally, in order to adapt the self-selection correction term for Catalan knowledge, we also re-estimate the Catalan proficiency equation of the first step separating low-educated and medium-high educated workers.

Keeping these purposes in mind, we proceed in the rest of this paper as follows: in Section 2, we describe the empirical methodology for accounting for self-selection into Catalan knowledge. Section 3 is dedicated to a brief illustration of the Catalan labour market in 2005-2006, as well as describing the data with some descriptive statistics. Section 4 contains the estimation results from the two stages (determinant of Catalan fluency and the earning regressions), including some sensitivity analysis. In Section 5, we check for potential language-skills complementarities. In Section 6, we discuss the results, providing some policy implications and, finally, Section 7 includes our conclusions.

2. Empirical Methodology

Our goal is to estimate the effect of being proficient in Catalan on individual monthly earnings. The standard approach consists in using the Mincerian equation, which regresses the natural logarithm of earnings (our data does not contains wages) on a set of explanatory variables, such as completed education, labour market experience, and other socio-demographic variables (gender, marital status, etc.). Moreover, when the purpose consists in estimating the language return, the standard earning equation is extended by including an indicator of language proficiency. The equation to be estimated takes the following form:

$$\ln(Y_i) = \alpha + \beta' X_i + \delta CAT_i + \phi' control + \varepsilon_i,$$

where Y_i represents individual earnings, X_i is the vector of individual explanatory variables, CAT_i is an indicator which takes the value of 1 if the individual is proficient in Catalan, and ε_i is an error term.

Nevertheless, a cross-sectional estimate of the δ parameter could not be interpreted as the true effect of language proficiency on earnings, because of the potential

endogeneity of the language indicator in the earning equation. If individuals make optimal choices with respect to language proficiency on the basis of their (observable and unobservable) characteristics, the OLS estimate of δ would also contain the effect of those characteristics that make some individuals more prone to be proficient in Catalan. Specifically, consider that the achievement of Catalan proficiency depends on a set of observable language determinants W_i (which may share some common elements with vector X_i), and on an error term u_i (with $u_i \sim N(0,1)$), such as

$$CAT_i = 1\left(\psi'W_i + u_i \geq 0\right).$$

The endogeneity problem arise if, conditional to W_i, the unobservable determinants of Catalan fluency (u_i) are correlated with the unobservable earnings determinants (ε_i). If this happens, the OLS estimates of the language return (δ) are biased by this correlation.

A solution to this problem consists in controlling for the part of the error term in the earning equation that is correlated with the language proficiency indicator, using an adaptation of the sample selection correction model originally proposed by Heckman (1979). This method consists in constructing a control function, to be included in the earning equation; this control function represents the part of the earning regression error term (ε_i), which is correlated with the Catalan proficiency (CAT_i) (that is $E[\varepsilon_i/CAT_i, W_i]$). When including this estimate of the conditional expectation of the unobservable in the earning equation, the effect of Catalan proficiency on earnings (δ) is consistently estimated. This happens since, by construction, what remains of the error term ε_i is not correlated with language fluency. Assuming joint normality for the error terms in the two equations (earnings and language proficiency), the correction term is known as the inverse of the Mills Ratio, and takes the following form:

$$E\left[\varepsilon_i \mid CAT_i = 1, W_i\right] = E\left[\varepsilon_i \mid u_i \geq \psi'W_i\right] = \rho_{\varepsilon u}\sigma_\varepsilon \lambda_{i,1} = \rho_{\varepsilon u}\sigma_\varepsilon \frac{\phi(\psi'W)}{\Phi(\psi'W)} \quad \text{if } CAT_i = 1$$

$$E\left[\varepsilon_i \mid CAT_i = 0, W_i\right] = E\left[\varepsilon_i \mid u_i < \psi'W_i\right] = \rho_{\varepsilon u}\sigma_\varepsilon \lambda_{i,0} = \rho_{\varepsilon u}\sigma_\varepsilon \frac{\phi(-\psi'W)}{1-\Phi(\psi'W)} \quad \text{if } CAT_i = 0$$

where φ and Φ are the normal density and the normal distribution function, respectively, and $\rho_{\varepsilon u}$ represents the correlation between the error term in the earning equation (ε_i) and the error term in language proficiency (u_i).

If the estimated coefficient associated with the correction term $\lambda_i = \lambda\left(\psi'W_i\right)$ in the earning equation is not statistically significant, it means that the selection process into language proficiency is exogenous, and that language return can be estimated by the simple OLS. However, a significant coefficient indicates that the two error terms are correlated; in fact, the coefficient associated with a self-selection correction term (λ_i) consists in $\rho_{\varepsilon u}\sigma_\varepsilon$. Therefore, self-selection into Catalan knowledge is important; in this case, the OLS estimates of δ would be biased by the selection process behind Catalan proficiency. Nevertheless, formal identification of the language effect requires that at least one variable included in the vector W of language proficiency determinants might be excluded from the earning equation. That is, at least one exclusion restriction is needed, which is represented by one or more variable (or instruments) that are correlated with language proficiency, but not directly with earnings[10].

3.   Data and Descriptive Statistics

The empirical analysis is carried out with the data from the "Survey of Living Conditions and Habits of the Catalan Population (ECVHP06)" of 2006, carried out by the Statistical Institute of Catalonia (IDESCAT). The original sample consists in 10,397 observations of individuals aged 16 or more and residing in Catalonia. The survey is directed on obtaining socioeconomic and demographic information of the overall population, concerned with the individual and its family. The data have been collected between the fourth trimester of 2005 and the third trimester of 2006; therefore, the information about individual labour market status and monthly earnings reflects the situation of 2005-2006.

Analyzing this period is very appealing for our purposes since, as described in Table 1, the unemployment rate was exceptionally low $(6,6\%)$[11] back then. This allows us to only focus on the employed population, as neglecting for potential self-selection

---

[10] As explained in Cameron and Trivedi (2005), the model is also identified without exclusion restriction because of the non-linearity of the correction term (λ). However, the exclusion restriction reinforces the full identification of the true language effect on earnings (we will return to this point later).

[11] In order to find similar values to the unemployment rate in Catalonia, we need to go back to 1978. Please note that the value reported is the mean unemployment rate between the fourth trimester of 2005 and the third trimester of 2006. The information contained in Table 1 is taken from the EPA (*Encuesta de la Población Activa*, Active Population Survey (INE)) for Catalonia. We also report the same information computed from the survey used in the (ECVHP06). Please note that the results are very similar, and even if both samples underrepresent non-national immigrants, this underestimation is more pronounced in the ECVHP06 sample.

into employment might not be problematic during this period[12]. Moreover, the female activity rate was sufficiently high in order to avoid excluding them from the analysis. Finally, the structure of occupation reveals that, as mentioned before, a significant proportion of workers are not born in Catalonia (34,4%): indeed, this part of the occupied labour force is more or less equally distributed among national immigrants (17,6%, mainly arrived between the 60s and the 70s), and non-national or foreigner immigrants (17,8%, the new migration).

[TABLE 1 ABOUT HERE]

The final sample used consists in 2,599 observations of the whole of individuals aged 16 to 65 who are regularly working, with valid information on earnings (collected in brackets). Also, we drop the observation of individuals with at least one parent born in Catalonia (see Table 1A in the Appendix for details). An interesting feature of the ECVHP06 sample is that it contains detailed information regarding Catalan knowledge and many other variables related to the individual's use of languages (Catalan and Castilian). The information about Catalan knowledge is reported in four categories; namely, individuals may claim they "do not understand", "understand but unable to speak", "able to speak but not to write", and "able to speak and write" Catalan. Table 1 contains the relative frequencies of each category of Catalan knowledge.

[TABLE 2 ABOUT HERE]

As expected, individuals born in Catalonia exhibit higher levels of Catalan knowledge, even if they are immigrants of the second generation. However, the proportion of foreign-born individuals who made an investment in Catalan knowledge is substantial, given that 24% of the foreign-born population claims to be able to speak and write in Catalan. Nevertheless, as it is also suggested by Rendon (2007), the self-declared language knowledge may be affected by a misclassification error, due to the potential overreporting of Catalan competences. Therefore, as mentioned before, we adopt a stringent definition of Catalan proficiency, considering as proficient only those

---

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

individuals who declare to speak and write in Catalan. Table 3 contains mean monthly earnings (in Euros)[13] by language proficiency.

[TABLE 3 ABOUT HERE]

These simple descriptive evidences indicate that being proficient in Catalan is associated with higher monthly earnings. Moreover, it seems that the earning penalty associated with a limited Catalan knowledge is higher for immigrants. However, these results may only reflect that individuals with higher chances to earn more have also a higher propensity to be proficient in Catalan. In order to obtain a correct estimate of the language effect on earnings, in the next section we estimate the earning regression accounting for the possible self-selection into Catalan proficiency. Before doing so, we briefly describe the variable used in the first-step estimation of the Catalan knowledge equation, and the specific variables included in the earning regression.

For the specification of the Catalan proficiency equation, we follow the general approach proposed by Chiswick and Miller (1995, 2001). Using the information contained in the ECVHP06 database, we consider that the observable determinants of Catalan knowledge are represented by the exposure to the language, the intensity per unit of exposure, and the efficiency in converting the exposure to the language into proficiency (apart from individual controls); in other words, $W_i = W$ *(exposure, intensity, efficiency, + control)*. Moreover, we also believe that monetary incentives, as the expected benefits from Catalan knowledge, might also represent important determinants of the propensity of being fluent in Catalan; however, such elements are not observed in the data[14]. Table 4 contains the descriptive statistics of the selected variables included in the first stage regression (the exact definition of constructed variables is reported in the Appendix).

First of all, the more direct indicators of exposure to Catalan are the duration in Catalonia (years since migration) for immigrants, and the individual age for natives; an increase of time spent in Catalonia should increase the propensity of being proficient in Catalan. Moreover, we include an indicator which takes the value of one if the

---

[13] Given that the information about monthly earnings is collected in brackets (see Table 2A in the Appendix for details), mean earnings are computed using an interval regression with a constant and the indicator for Catalan proficiency (=1 if the individual is able to write).

[14] Therefore, monetary incentives (expected benefits in the labour market plus the expected duration in Catalonia for immigrants) are contained in the error term of the Catalan knowledge equation (u).

individual normally spoke Catalan with at least one parent. This variable captures the possibility of being born in areas where Catalan is also used, such as the Balearic Islands, Valencia, la Franja (an area in the Aragon region, on the border with Catalonia), Alghero (Sardinia) and the south of France[15]. In addition, we consider that the exposure to the language is more intense for those individuals who speak Catalan at home and/or with their children (controlling also for the number of children and for those individuals who are living alone)[16]. Another element of the exposure to the language, never considered in previous studies (due to the lack of information), consists in the effect of the Media. We include two indicators for capturing the effect of Catalan Media: namely, whether the individual normally watches the news on the Catalan television channel (only broadcasted in Catalan), or whether they normally read Catalan newspapers.

[TABLE 4 ABOUT HERE]

Following Rendon (2007), we include other three elements which might capture the intensity of exposure to Catalan per unit. First, we consider that those individuals who arrived in Catalonia when they were younger than 10 have been more intensively exposed to the language. Second, we introduce a dummy for being affected by the Linguistic Normalization at a young age, which takes the value of one for all the individuals who have been potentially schooled in Catalan. Third, we include the proportion of individuals who are able to speak and write in Catalan in the region of residence (*Comarca*). This variable is included in order to capture territorial neighbourhood externality effects on the likelihood of being fluent in Catalan; in other words, this variable would represent the effect of Catalan-linguistic community concentration in the neighbourhood, on the propensity to being proficient in this language.

Apart from the exposure to Catalan and its intensity, the efficiency in transforming exposure into language learning is also a fundamental determinant of the probability of being proficient. For capturing differences in Catalan proficiency related to the

---

[15] Unfortunately, we do not have the exact place of birth, as in the ECVHP06 the information regarding the place of birth is grouped as follows: a) born in Catalonia, b) born in Spain, c) born in Europe, d) born in America, e) born in Africa, d) born in Asia. The available information about the place of birth is included as control, as well as a dummy for individual sex, an indicator for individuals whose mother-tongue is different from Spanish and an indicator for (bad) Spanish knowledge.
[16] Many individuals who are not Catalan mother-tongue could be used to speak Catalan to their children, because the latter are schooled in Catalan and may be used to speak this language in the daily life.

efficiency, we include completed education indicators, as well as the number of books and the frequency of reading (also representing the cultural environment of the individual). Moreover, given that the easiness for learning a language decreases with age, we also include a quadratic specification of the age effect.

Table 4 also contains summary statistics of the variables included in the vector X_i (earning determinants). Specifically, apart from gender, completed education and years since migration (which are also directly included in the first-stage equation), the specific regressors included in the earning equation are an indicator for marital-status, job-tenure and potential previous experience. Moreover, given that we dispose of information about earnings instead of wages, we include some indicator for job characteristics, in order to control for earning differences irrespective of human capital and language knowledge. Namely, we consider whether the individual is unionized, whether he/she is a part-time worker, whether the firm has more than 500 workers, and whether the individual is a self-employed worker. In the next section, we report the estimation results; we briefly illustrate the estimations of the determinants of Catalan proficiency, which are used for correcting the earning equations for the potential self-selection into Catalan knowledge. Subsequently, we analyze the results from the earning equation estimates, and the earning return to Catalan proficiency.

4. Estimation Results

4.1 The Determinants of Catalan Proficiency

The results from the Probit model for the determinants of Catalan proficiency are reported in Table 5; the high pseudo-R^2 indicates that the estimation performs very well, confirming that the variables included in the model have a high explanatory power for the probability of being proficient in Catalan (namely, being able to speak and write). As expected, those individuals who are born outside Catalonia have less chances of being proficient in Catalan than natives (even if the latter are children of immigrants). Moreover, the penalization is lower for immigrants who are coming from the rest of Spain, given the proximity between Castilian and Catalan. The results also indicate that females are slightly more likely to be proficient in Catalan than males.

The estimates of the age effect suggest that the relationship between the likelihood of having a good knowledge of Catalan and age is U-shaped. This means that the probability of being proficient in Catalan decreases with age, but older individuals are more likely to be able to speak and write Catalan (that is, a consistent picture of the historical situation of Catalonia). On the contrary, the positive coefficient of years since migration indicates that a longer stay in Catalonia (for immigrants) increases the probability of knowing Catalan[17]. In addition, those individuals who speak Catalan with their parents are advantaged in being proficient in this language. There is also a positive and significant effect of speaking Catalan at home and/or with the children, indicating that the intensity of the exposure related to family environment matters for language knowledge. Surprisingly, however, individuals who are living alone are more prone to be proficient in Catalan, *ceteris paribus*. Finally, the expected positive effect of local media seems to be limited, given that only the coefficient associated with reading Catalan newspapers is slightly significant.

As also suggested by Rendon (2007), for those individuals who arrived in Catalonia when they where younger than 10, the linguistic assimilation is more likely; the same happens for those individuals who, keeping fixed the other variables, have been affected by the Linguistic Normalization of 1983 when young. Moreover, our results also indicate that living in areas where there is a higher percentage of individuals who are fluent in Catalan increases the chances of being proficient in this language. This evidence means that the externality effects within the neighbourhood of residence are very important for Catalan knowledge. The coefficients of the dummies for completed education are highly significant and have the expected sign, indicating that (comparing to individuals with no education) more educated individuals have higher chances of being proficient in Catalan. Finally, the effect of cultural resources (the number of books at home and the frequency of reading) on the probability of being able to speak and write in Catalan is positive and statistically significant.

---

[17] However, the negative sign of the interaction Spanish✗YSM indicates that national immigrants are less likely to be proficient in Catalan with the increase of the duration in Catalonia; therefore, the advantage of individuals who were born in Spain with respect to foreigners decreases with the duration in Catalonia. This represents that individuals who came from the rest of Spain in the past may have had fewer incentives to learn Catalan since the use of this language was still restricted to oral communication.

4.2 Catalan Proficiency and Earnings

In this section we analyze the results from the earning regressions[18] and the effect of Catalan proficiency, which are reported in Table 6. Specifically, the first column of the table contains the estimates from a standard Mincer-type equation, augmented by years since migration (capturing the immigrants' economic assimilation) and several indicators of job-characteristics (in order to control for horizontal earning differences, irrespective of human capital and language knowledge[19]). In addition, the second column contains the same earning equation augmented by our indicator of language proficiency (being able to speak and write Catalan). In general, the model performs very well, and the coefficient estimates have the expected sign.

The first interesting evidence is that foreigners' earning penalty (with respect to native-born) is strongly reduced when controlling for Catalan knowledge; consistently with the evidences from previous studies (Amuedo-Dorantes and De La Rica 2007, Izquierdo et al. 2009), African immigrants seem the most disadvantaged, whereas European immigrants do not earn significantly less than native-born. Moreover, when accounting for language proficiency, national immigrants earn more than natives (children of immigrants). The coefficient for years since migration has a negative sign, and it is statistically significant when controlling for language knowledge. This evidence at first sight seems contradictory; however, it could be explained by the fact that individuals who arrived in Catalonia many years ago were less skilled than more recent immigrants. Females earn significantly less than males, and this difference is not modified when accounting for Catalan proficiency. Moreover, married individuals tend to have higher earnings than those who are not married in both equations.

[TABLE 6 ABOUT HERE]

As expected, an increase in the completed level of education is associated with higher monthly earnings, and the return to education is slightly reduced when the regression includes the language proficiency indicator. Additionally, increasing current

---

[18] Given that the dependent variable (monthly earnings) is coded in intervals, we use the interval regression method ("intreg" with STATA). Using OLS or ML methods with the mid-points of each interval yields the same estimate results; however, the coefficients are somewhat more precisely estimated with the interval regression method.

[19] One may claim that human capital and/or language proficiency might be also related with these variables (namely, being unionized, working in a large firm, working part-time and being a self-employed worker). However, excluding such variables does not modify the general results. We decided to keep these variables in the model, in order to reduce the variability of the residuals (relevant in the sample selection corrected equation).

job-tenure (linearly) raises individual earnings, while the positive effect of previous potential labour market experience seems to be convex (given the negative and significant coefficient of the squared experience). However, the return to these elements of human capital decreases when we take into account Catalan knowledge.

Similarly in both equations, the coefficients associated with job-characteristic control variables show, unsurprisingly, that unionized workers and/or those who work in a large firm earn more than the mean, whereas the contrary happens for part-time workers. The effect of being a self-employed worker is slightly positive, but lowly significant. Finally, those individuals who reside in the metropolitan area of Barcelona earn somewhat more than those who reside in other zones of Catalonia. Nevertheless, the main result showed in the second column of Table 6 is that the "OLS[20]" estimate of the effect of Catalan proficiency is positive and statistically significant, suggesting that knowing Catalan is associated with earning about 7% more than the mean (that is, $\exp(\hat{\delta}) - 1 = \exp(0.066) - 1 \cong 0.068$).

However, as explained above, this simple estimate of the return to language proficiency could be biased, because it may also contain the effect of those elements which help being more prone to knowing Catalan, which are also potentially correlated with individual earnings[21]. In order to obtain an estimate of language knowledge return which takes into account the self-selection process behind Catalan knowledge, we augment the earning equation with the estimated self-selection correction term ($\hat{\lambda}_i = \lambda\left(\hat{\psi}'W_i\right)$) obtained from the first-stage equation. The third column of Table 6 contains the earning equation corrected for selection into Catalan proficiency. The most relevant differences with respect to the previous estimates consist in *i*) the further reduction of earning penalties for foreign workers when accounting for both language knowledge and self-selection. Consistently, when controlling for selection into Catalan knowledge, individuals who are born in other Spanish regions earn even more than natives (immigrants of the second generation). Moreover, when we take into account

---

[20] Remember that the estimated coefficients are obtained through the interval regression, given that the dependent variable (monthly earnings) is reported in brackets in the database used. However, the coefficients from an interval regression are practically identical to OLS coefficients (especially in terms of interpretation); moreover, the results obtained with OLS over the mid-points of earning intervals are almost the same than those obtained by applying the interval regression method.

[21] Moreover, measurement error in Catalan knowledge may also represent an additional source of bias. However, due to data limitation, we are not able to control for measurement error; therefore, the estimate that we obtain represents a lower bound of the true value of Catalan proficiency (under the standard hypothesis of the measurement error model).

those elements which improve the likelihood of being proficient in Catalan, *ii*) there is an additional reduction of the returns to education, and *iii*) an increase of the returns to current and previous experience.

Finally, *iv*) the estimated earnings return to Catalan knowledge is significantly higher when controlling for self-selection, showing that those individuals who are able to speak and write Catalan earn approximately 20% more than the rest of individuals ($\exp(0.176) - 1 \cong 0.19$). Moreover, notice that the coefficient associated with the Inverse of the Mills Ratio (λ) is statistically significant with a negative sign. The χ^2 test for the statistical significance of the correlation coefficient between the error terms of the earning regression and of the first-stage equation ($\hat{\rho}_{\varepsilon u}$) is 8.56 (P-value 0.0034), rejecting the null hypothesis of independence of the two equations. This evidence indicates that a) in order to obtain an unbiased estimate of the effect of Catalan proficiency it is important to correct for self-selection, given that the earning equation and the Catalan knowledge equation are not independent. In addition, b) the negative coefficient of the correction term indicates that the unobservable elements that make one prone to speaking and writing in Catalan (keeping the observable determinants fixed), are negatively correlated with the unobservable determinants of earnings. This means that there is a negative selection effect on earnings, given that individuals with a higher propensity to knowing Catalan are more likely to earn less than the mean (we will return to this evidence in the discussion of the results).

4.3 Sensitivity Analysis

In the last section we have showed that Catalan proficiency is associated with higher monthly earnings. Moreover, when taking into account the elements which make one more prone to know Catalan (that is, being able to speak and write in Catalan), the earning return to language proficiency is notably higher than the simple statistical association obtained through interval regression. Indeed, the negative (and significant) correlation coefficient between the error-terms of the first and the second stage equations indicates that individuals with a higher propensity of knowing Catalan (keeping the observed determinants fixed) are expected to earn less than the mean. This means that the simple association between Catalan proficiency and earnings is capturing the earning penalty of those individuals who are highly endowed with

attributes that are favourable for language assimilation (*ceteris paribus*); in other words, the negative selection effect.

However, before discussing which feature of the Catalan economy could explain this evidence, and the consequent implications for public policies, we briefly check how sensitive the results are to the sample selection and to the exclusion restrictions in the first-stage equation. In Table 7, we report the earning return to Catalan proficiency for different sub-samples and with different specifications of the Catalan knowledge[22]. First of all, we separately compute the return for males and females, allowing also for a gender-specific self-selection process; in fact, males and females may face different costs, but also different returns to Catalan knowledge.

The results indicate that the positive effect of Catalan proficiency is higher for females than for males (as found by Rendon 2007); however, given that the obtained estimates are not statistically different, we do not reject the joint estimation for both sexes. Moreover, we also try to exclude non-national immigrants from the sample, as among immigrants, knowing Catalan might represent additional advantages in the labour market (for example, a better adaptation to the local institutions)[23]. Consistently, the estimated return to Catalan proficiency excluding foreigners is lower than the return obtained from the estimation with the overall (selected) sample.

[TABLE 7 ABOUT HERE]

The other important issue related to the sensitivity of the results is the definition of the exclusion restrictions. As commented in Section 3, in order to fully identify the language effect on earnings without any effect of self-selection into Catalan proficiency, one exclusion restriction is needed at least. In other words, at least one variable appearing in the first-stage equation might be excludable from the earning equation of the second-stage (because it is correlated with language proficiency but not directly with earnings). If there are no variables which satisfy such requirements, the model is only identified through the non-linearity of the inverse of the Mills Ratio (the self-selection correction term). Among the available information, there are several candidates to be (jointly) used for generating some exogenous variation into Catalan

---

[22] Complete estimations are not showed here, but are available upon request to the authors.
[23] We are not able to carry out the estimation for the foreigners sub-sample, given the reduced number of observations.

knowledge (that is, as instruments). Namely, the three language-use variables (speaking Catalan with the parents, at home, and with the children); the local-media variables (watching Catalan TV[24], or reading a Catalan newspaper); having arrived at a younger age than 10, and being affected by the Linguistic Normalization; the percentage of individuals who speak and write Catalan in the area of residence; the number of books at home and the frequency of reading[25].

We may provide different arguments for sustaining the validity of these variables for identifying the true language effect, but there is probably the same number of arguments for claiming the contrary. Consequently, we check the sensitivity of the Catalan proficiency effect to the exclusion restrictions used. In Table 7, we report the earnings return estimated by progressively excluding each of the instruments used in the first-stage. Excluding the indicator for speaking Catalan with the parents reduces the estimated return to Catalan knowledge, whereas excluding the other two language-use variables (speaking Catalan at home and/or with the children) increases the premium. Excluding "News Catalan TV" has almost no effect on the premium, but excluding "Catalan newspapers" reduces it. The return to Catalan knowledge is slightly lower when estimating the first-stage without the indicator for "arrived at a younger age than 10" and for "being affected by Normalization". However, when we omit the percentage of individuals who speak and write in Catalan in the area of residence the return is notably higher. Important changes are also observed when excluding the "number of books at home" (reduction) and the frequency of reading (increase). Therefore, we might claim that the estimated earning return to language knowledge is sensitive to the exclusion restriction implemented (especially with respect to the last three variables); this evidence suggests that the identification of the premium is not obtained through the non-linearity of the selection correction term, but through the instrument used[26].

---

[24] Note that the "News Catalan TV" coefficient is not statistically different from zero; therefore, this variable could not serve as an instrument. However, leaving it into the model does not modify the general results (see Table 7).
[25] The rest of variables included in the Catalan knowledge equation are included in the earning equation directly or through linear combination (apart from those variables that are a simple control, such as the number of children, living alone, having a foreign language as mother-tongue or having a limited knowledge of Spanish).
[26] However, the estimated return may also depend on the joint-normality distributional assumption for the error terms of the first and second stage equations; in order to prove this eventuality we need to estimate the model non-parametrically, or semi-parametrically, but this is out of the purposes of the present work.

6. Language-Skills Complementarity: the Role of Education

According to the evidence from the recent literature, in this section we check the existence of complementarity between language knowledge and individual skills in Catalonia. More specifically, we focus on the role of completed education[27], as we suspect that being proficient in Catalan might have an economic reward only for medium/high-skilled workers. With more detail, we believe that Catalan knowledge represents a valuable asset only in occupations that require higher levels of formal education. Therefore, we estimate the return to Catalan knowledge separately for individuals with low education (that is, individuals who have no education or have completed only primary education) and for individuals with high education (individuals with completed secondary or tertiary education)[28].

[TABLE 8 ABOUT HERE]

The results indicate a strong complementarity between Catalan knowledge and completed education. The return estimated with the simple interval regression for the full sample is almost completely identified by the return to Catalan proficiency for more educated individuals; indeed, the estimated return to language knowledge for less educated individuals is not statistically different from zero. Moreover, we obtain similar results when accounting for self-selection effects. In fact, there exists a positive and statistically significant return for high educated individuals (somewhat higher than the estimate for the full sample), but for low educated individuals the return is almost zero. Finally, we repeat the analysis excluding non-national immigrants and again we find that the economic reward of knowing Catalan is highly positive only for high-educated individuals (and slightly lower than the previous estimate), and statistically null for low-educated individuals. Definitively, it seems that being proficient in Catalan increases expected monthly earnings, but only for individuals who have achieved a certain level of skills through the educational system[29]. The evidences that we have

---

[27] Unfortunately, we do not dispose of information about individual occupation. Therefore, we are limited in considering the complementarity between Catalan proficiency and completed education. Checking the potential complementarities between language and occupational skills will be the objective of future research.

[28] In order to capture different selection effects, we also estimate separately the first-stage equation — the results are not shown here. Complete results from the earning equations (with and without self-selection correction) are reported in Table 3A in the Appendix.

[29] However, we must take into account the results obtained by González (2005). Estimating non-parametric bounds to language returns, she suggests that the higher returns to language for high-educated workers are only due to a strong selection ability bias in higher education levels. We therefore might bear in mind that our results, which are obtained by only separating according to completed education, may be only (or in part) capturing the process of self-selection into higher levels of education produced by individual unobserved ability.

obtained raise several questions with regards to the planning of the linguistic and socioeconomic policy agenda in Catalonia.

7. Discussing the Results

In light of the results from this study we can claim that, in general, there exists a positive economic value of Catalan knowledge in Catalonia. Our estimations indicate that those individuals who are able to speak and write in Catalan earn more than those who are not. Nevertheless, the simple statistical association between Catalan proficiency and monthly earnings is quite modest, since it is confounded by the potential earning penalty of those individuals who are more prone to be proficient in Catalan. In other words, the estimated return to Catalan proficiency (without controlling for the determinants of language knowledge) consists in 7% of extra earnings. However, when the estimation of the earning return to language proficiency is corrected for the effect of self-selection into Catalan knowledge, the additional earnings attributed to this special form of human capital are close to 20%.

The negative and statistically significant selection effect indicates that, in order to obtain a correct estimate of the earning return to Catalan proficiency, it is important to control for the self-selection process into language knowledge. Moreover, the elements that make one more prone to know Catalan tend to be negatively correlated with (unobservable) individual skills for generating earnings. This evidence points out the possibility that in Catalonia individuals with higher language-skills are likely to be remunerated less than the others (*ceteris paribus*)[30]. In addition, further results concerning the potential complementarities between language and individual skills suggest that only more educated individuals can beneficiate from knowing Catalan (in terms of earnings). A potential explanation of this evidence is that low-skilled individuals can only obtain occupations where Catalan knowledge is not required, and/or is not considered as an economic asset.

---

[30] A potential explanation could be that individuals with higher propensity of being proficient in Catalan are also more prone to work in the Public Administration were, in general, salaries tend to be lower than in the private sector. In addition, the public sector also imposes more restrictions to (remunerated) extra-hours, which may contribute to explain the negative selection effect on monthly earnings.

Probably, additional information about an individual's occupation could help on clarifying these preliminary evidences concerning the effect of Catalan proficiency on earnings. Indeed, this information (which unfortunately is not available in our data) may allow understanding what produces the negative selection effect. Conditioning to occupation, we will be able to check if the negative selection effect is produced by the lower remuneration in those occupations which attract individuals with a higher propensity to know Catalan (or to other unobservable determinants of Catalan knowledge, negative correlated with unobservable individual skills). Moreover, disposing of information on occupation might also add further and more detailed evidences about the potential language-skills complementarity. In fact, knowing the individuals' occupation we would be able to understand whether language knowledge has some economic reward only in given types of occupations (a job-specific skills complementarity) or, more in general, for more educated people only.

Moreover, we must also consider that our analysis neglects of more complicated selection mechanisms, related to Catalan proficiency and individual occupation, which may confound the earning/productivity effect of language knowledge. In fact, as suggested by Aldashev et al. (2009) for the German case, when accounting for the positive language self-selection effects into occupations and economic sectors (and allowing for correlation between these two choices), there is no direct productivity effect of language fluency among low-skilled workers. Furthermore, for high-skilled workers it is not clear whether such effect exists. The authors affirm that the most relevant effect of language knowledge (among immigrants) is on employment probabilities, on finding a "better occupation", or obtaining a job in a "better economic sector". In fact, the positive effects of language on earnings reported in many previous studies (as in this one), may be partially reflecting these other effects of language proficiency (explaining also our negative selection effect).

Note that a recent working paper by Quella and Rendon (2009), which focused on the Catalan situation, shows that national immigrants with a high level of Catalan fluency are more likely to be selected into white-collar and communication-intensive occupations (among other things). They also find a significant correlation between occupational selection and language selection. Consequently, we expect that the earnings return to Catalan knowledge that we have estimated (without considering the role of language for self-selection into occupations), may be capturing part of these

positive effects of knowing Catalan. In other words, we are not able to prove the existence of some productivity effect of Catalan proficiency in a separate fashion from these potential effects on employment probabilities and occupation types. Indeed, an objective of our future research will be modelling a simultaneous self-selection process into language knowledge and occupation, which would allow appreciating the global effects of language knowledge on labour market success. Therefore, we will wait for more detailed data, in order to produce further evidences regarding the economic value of Catalan knowledge in Catalonia.

8. Conclusions

This paper represents the first attempt of analyzing the effect of Catalan knowledge on individual monthly earnings in Catalonia. Using the data from the 2006 wave of the "Survey of Living Conditions and Habits of the Catalan Population (ECVHP06)", we find that for national immigrants, foreign immigrants and native children, whose parents are not born in Catalonia (immigrants of the second generation), Catalan knowledge has a positive reward in terms of earnings. In fact, those individuals who can speak and write in Catalan earn more than those who cannot. However, taking into account the elements which make one more prone to know Catalan (that is, taking into account self-selection into language knowledge), the extra earnings which can be exclusively attributed to Catalan proficiency are significantly higher. This evidence indicates a negative selection effect, which means that individuals who are better endowed of (unmeasured) elements which facilitate language assimilation are less remunerated than the mean (keeping fixed the other earning determinants). However, considering the role of individual occupation and its potential relationships with language knowledge would be extremely useful for obtaining more complete evidences about the economic value of Catalan knowledge.

These results must be taken into account in the future planning of linguistic policies in Catalonia, given that stimulating the use of Catalan through public policies, beyond the explicit cultural and linguistic purposes, has important socioeconomic implications as well. Moreover, what is still more important for the design of linguistic policies in Catalonia in the medium-long term is the strong complementarity between Catalan knowledge and individual schooling. Specifically, our results suggest that only

individuals with secondary or tertiary education receive an earning return to language knowledge. This means that for low-educated individuals, Catalan proficiency might scarcely improve labour market outcomes (especially if defined with respect to earnings). In fact, the target of linguistic policies (as of other socioeconomic policies) is often the foreign population, which currently includes those who are more likely to be low-educated. Therefore, policy-makers might consider that in order to achieve a complete economic and social assimilation (maybe a step ahead from cultural integration), fostering Catalan knowledge among immigrants is not sufficient. An effective public policy could consist in combining Catalan learning with an improvement of human capital and other skills of the most disadvantaged part of the Catalan population.

References

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(3), pp. 330-341.

Amuedo-Dorantes, C. & de la Rica, S. (2007). "Labour Market Assimilation of Recent Immigrants in Spain," British Journal of Industrial Relations, vol. 45(2), pp. 257-284, 06.

Berman, E., Lang, K. & Siniver, E. (2003). "Language-skill complementarity: returns to immigrant language acquisition," Labour Economics, vol. 10(3), pp. 265-290.

Bleakley, H. & Chin, A. (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants," The Review of Economics and Statistics, vol. 86(2), pp. 481-496, 05.

Cameron, A. C., & Trivedi, P. K., (2005). "*Microeconometrics: Methods and Applications*", Cambridge University Press.

Carliner G. (1996). "The wages and language skills of US immigrants". NBER Working Paper W5763.

Chiswick, B. (1991). "Speaking, Reading and Earnings Among Low-Skilled Immigrats"; Journal of Economics, 9 (2), pp. 149-170.

Chiswick, B. & Miller, P.W. (1995). "The Endogeneity Between Language and Earnings: International Analyses"; Journal of Labor Economics, 13(2), pp. 245-287.

Chiswick, B. & Miller, P.W. (1999). "Language skills and earnings among legalized aliens"; Journal of Population Economics, vol. 12(1), pp. 63-89.

Chiswick, B. & Miller, P.W. (2001). "A Model of Destination Language Acquisition: Application to Male Immigrats in Canada"; Demography, 38(3), pp. 391-409.

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

Dustman, C. & van Soest, A. (2001). "Language Fluency And Earnings: Estimation With Misclassified Language Indicators"; The Review of Economics and Statistics, 83(4), pp. 663-674.

Dustman, C. & van Soest, A. (2002). "Language and the earnings of immigrants," Industrial and Labour Relations Review, vol. 55(3), pp. 473-492, April.

Dustmann, C. & Fabbri, F. (2003). "Language proficiency and labour market performance of immigrats in the UK"; Economic Journal, 113(489), pp. 695-717, 07.

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

Fernández-Huertas Moraga, J & Ferrer-i-Carbonell, A. (2008). Immigration in Catalonia. El Centre d'Estudis de Temes Contemporanis (CETC).

González, L. (2005). "Nonparametric bounds on the returns to language skills," Journal of Applied Econometrics, vol. 20(6), pp. 771-795.

Institut Estadístic de Catalunya (IDESCAT), (2006) "*Enquesta de Condicions de Vida i Hàbits de la Població de Catalunya*" (Survey on Living Conditions and Habits of the Catalonian Population).

Izquierdo, M., Lacuesta, A. & Vegas, R. (2009). "Assimilation of Immigrants in Spain: a Longitudinal Analysis" Labour Economics, 16(6) pp. 669-678.

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(1), 50.

Lindley, J. (2002). "The English Language Fluency and Earnings of Ethnic Minorities in Britain," Scottish Journal of Political Economy, vol. 49(4), pp. 467-87.

Mora M.T., & Davila, A. (1998). "Gender, Earnings, and the English Skill Acquisition of Hispanic Workers in the U.S." Economic Inquiry, 26, pp. 631-644.

Quella, N. & Rendon, S. (2009). "Occupational Selection in Multilingual Labor Markets: The Case of Catalonia," 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(3), pp. 669-686, July.

TABLES

Table 1: Catalan Labour Market in 2005-2006

	EPA (2005-IV/2006-III)	ECVHP06
ACTIVITY RATE		
Males	72.2	69.7
Females	52.3	51.9
Total	62.0	60.7
UNEMPLOYMENT RATE		
Males	5.2	5.6
Females	8.4	5.8
Total	6.6	6.8
% OF WORKERS BY PLACE OF BIRTH		
Catalonia	65.9	69.8
Other Spanish Regions	17.6	17.3
Other Countries	16.8	12.9

Source: EPA and ECVHP06.

Table 2: Catalan Knowledge

PERCENTAGES	*OVERALL SAMPLE*	*NATIVES (SEC. GEN.)*	*IMMIGRANTS*
Do not Understand	4.6	0.38	7.68
Understand but unable to Speak	31.56	12.46	45.47
Able to Speak but Not to Write	20.08	16.04	23.01
Able to Speak and to Write	43.76	71.12	23.85
TOTAL (#)	2,599	1,048	1,551

Source ECVHP06.

Table 3: Mean Monthly Earnings by Catalan Proficiency

	MEAN MONTHLY EARNINGS		
	OVERALL SAMPLE	*NATIVES (SEC. GEN.)*	*IMMIGRANTS*
Proficient	1205.819	1146.294	1336.015
Not Proficient	1085.923	1123.777	1075.285

Source ECVHP06.

Table 4: (Weighted) Descriptive Statistics

VARIABLE	DESCRIPTION	MEAN	S.D	MAX	MIN
Native-S.G.	*= 1 if born in Catalonia (parents born outside Catalonia), 0 otherwise*	0.421	0.494	0	1
Spanish	*= 1 if born in other Spanish regions, 0 otherwise*	0.334	0.472	0	1
European	*= 1 if born in Europe, 0 otherwise*	0.044	0.205	0	1
African	*= 1 if born in Africa, 0 otherwise*	0.067	0.25	0	1
American	*= 1 if born in America, 0 otherwise*	0.123	0.329	0	1
Asian	*= 1 if born in Asia, 0 otherwise*	0.01	0.1	0	1
Foreign Language	*= 1 if foreign mother-tongue*	0.107	0.309	0	1
Limited Castilian	*= 1 if limited knowledge (not able to speak Castilian), 0 otherwise*	0.8	0.4	0	1
Sex =1 if Female	*= 1 if the individual is Female, 0 if Male*	0.425	0.494	0	1
Age	*individual age in years*	38.494	10.652	17.5	62.5
YSM	*years since migration to Catalonia*	12.91	16.474	0	62.5
Speak Catalan with Parents	*= 1 if the individual speaks Catalan with at least one parent, 0 otherwise*	0.044	0.205	0	1
Speak Catalan at Home	*= 1 if the individual speaks Catalan at home, 0 otherwise*	0.17	0.375	0	1
Speak Catalan with Children	*= 1 if the individual speaks Catalan with his/her children, 0 otherwise*	0.2	0.4	0	1
# Children	*number of children*	1.272	1.208	0	6
Living alone	*= 1 if the individual lives alone, 0 otherwise*	0.042	0.2	0	1
News Catalan TV	*= 1 if the individual watches the newscast on the Catalan television channel, 0 otherwise*	0.327	0.469	0	1
Catalan Newspaper	*= 1 if the individual normally reads a Catalan newspaper, 0 otherwise*	0.245	0.43	0	1
Arrived Younger than 10	*= 1 if the individual was younger than 10, when he/she arrived to Catalonia, 0 otherwise*	0.125	0.331	0	1
Normalització	*= 1 if the individual was younger than 12 in 1984, 0 otherwise*	0.289	0.453	0	1
% Write Catalan (*Comarca*)	*percentage of individuals able to speak and write in Catalan in the zone of residence*	52.345	6.572	36.1	81
# Books at Home	*number of books at home*	163	215.97	0	1000
Read Frequently	*= 1 if the individual reads frequently, 0 otherwise*	0.378	0.485	0	1
No Education	*= 1 if the individual has no education, 0 otherwise*	0.041	0.198	0	1
Primary Education	*= 1 if the individual has completed primary education, 0 otherwise*	0.359	0.48	0	1
Secondary Education	*= 1 if the individual has completed secondary education, 0 otherwise*	0.42	0.494	0	1
Tertiary Education	*= 1 if the individual has completed tertiary education, 0 otherwise*	0.18	0.384	0	1
Married	*= 1 if the individual is married, 0 otherwise*	0.617	0.486	0	1
Job Tenure	*job tenure in years (current job)*	8.758	7.414	1	22.5
(Previous) Experience	*previous potential experience in years*	13.62	10.649	0	51
Unionized	*= 1 if the individual is affiliated to a labour union, 0 otherwise*	0.44	0.496	0	1
Part-time	*= 1 if the individual works part-time, 0 otherwise*	0.29	0.454	0	1
#Workers>500	*= 1 if the individual works in a firm with more than 500 employees, 0 otherwise*	0.101	0.301	0	1
Self-employed	*= 1 if the individual is a self-employed worker, 0 otherwise*	0.128	0.334	0	1
Living in Barcelona	*= 1 if the individual lives in the Barcelona Metropolitan Area, 0 otherwise*	0.755	0.43	0	1

Table 5: Probit Results for the Determinants of Catalan Proficiency

Dependent Variable: Being Proficient in Catalan (0-1)		
	Coefficients	*Robust S.E.*
Constant	-0.376	0.880
Native-S.G.	*ref. cat.*	—
Spanish	-1.065***	0.187
European	-2.448***	0.425
African	-2.407***	0.401
American	-2.872***	0.375
Asian	-2.803***	0.511
Foreing Language	-0.081	0.246
Limited Castilian	-0.086	0.320
Sex (=1 if Female)	0.141*	0.076
Age/10	-1.235***	0.301
Age^2/100	0.077**	0.035
YSM/10 (=0 for Natives)	0.659***	0.110
Spanish×YSM	-0.047***	0.011
Speak Catalan with Parents	0.498***	0.192
Speak Catalan at Home	0.663***	0.125
Speak Catalan with Children	0.289**	0.113
# Children	-0.011	0.043
Living alone	0.470***	0.171
News Catalan TV	0.095	0.088
Catalan Newspaper	0.182**	0.078
Arrived Younger than 10	0.274**	0.117
Normalització	0.297**	0.134
% Write Catalan (*Comarca*)	0.042***	0.006
No Education	*ref. cat.*	—
Primary Education	0.975**	0.410
Secondary Education	1.616***	0.410
Tertiary Education	2.205***	0.419
# Books at Home/100	0.048**	0.019
Read Frequently	0.378***	0.080
Number of Observations	2,599	
Log-Likelihood	-923.43	
Pseudo-R2	0.482	

*NOTE: *** significant at 0.01%; ** significant at 0.05%; * significant at 0.1%.*
Robust standard errors are in parenthesis. The estimation is weighted with the
populational weights provided in the ECVHP06 database.

Table 6: Earning Regressions with Catalan Proficiency

Dependent Variable: *Ln(Earnings)*	INTERVAL REGRESSION	INTERVAL REG. + CATALAN	INT. REG. + SELF SELECTION
Constant	6.557***	6.499***	6.404***
	(0.055)	(0.057)	(0.066)
Native-S.G.	*Ref. Cat.*	*Ref. Cat.*	*Ref. Cat.*
Spanish	0.045	0.076**	0.126***
	(0.035)	(0.036)	(0.040)
European	-0.023	0.015	0.078
	(0.042)	(0.043)	(0.050)
African	-0.145***	-0.108***	-0.046
	(0.030)	(0.032)	(0.040)
American	-0.085***	-0.040	0.035
	(0.025)	(0.028)	(0.036)
Asian	-0.114	-0.071	0.002
	(0.081)	(0.081)	(0.087)
YMS/10	-0.018	-0.024**	-0.034***
	(0.011)	(0.011)	(0.012)
Sex (=1 if Female)	-0.275***	-0.278***	-0.283***
	(0.016)	(0.015)	(0.016)
Married	0.075***	0.077***	0.079***
	(0.016)	(0.016)	(0.016)
No Education	*Ref. Cat.*	*Ref. Cat.*	*Ref. Cat.*
Primary education	0.109**	0.108**	0.106**
	(0.043)	(0.043)	(0.043)
Secondary Education	0.241***	0.231***	0.215***
	(0.046)	(0.046)	(0.045)
Tertiary Education	0.545***	0.527***	0.497***
	(0.050)	(0.049)	(0.050)
Job Tenure/10	0.123***	0.133***	0.147***
	(0.013)	(0.013)	(0.014)
(Previous) Experience/10	0.116***	0.126***	0.145***
	(0.023)	(0.023)	(0.024)
Experience^2/100	-0.024***	-0.024***	-0.025***
	(0.006)	(0.006)	(0.006)
Unionized	0.103***	0.100***	0.100***
	(0.017)	(0.017)	(0.017)
Part-time	-0.430***	-0.431***	-0.432***
	(0.032)	(0.031)	(0.032)
#Workers>500	0.172***	0.173***	0.171***
	(0.030)	(0.030)	(0.030)
Self-employed	0.035*	0.032*	0.030
	(0.018)	(0.018)	(0.019)
Living in Barcelona	0.032**	0.038**	0.046***
	(0.016)	(0.016)	(0.016)
Catalan Proficiency	—	0.066***	0.176***
		(0.019)	(0.041)
Self-Selection Correction (λ)		—	-0.079***
			(0.027)
$\hat{\sigma}_\varepsilon$	0.323(0.006)	0.322 (0.006)	0.322 (0.006)
$\hat{\rho}_{\varepsilon u}$	—	—	-0.247 (0.084)

NOTE: *** significant at 0.01%;** significant at 0.05%; * significant at 0.1%. Standard errors are in parenthesis; for the OLS regressions heteroskedasticity robust standard errors are used; for the earning regressions with self-selection correction the coefficient standard errors are obtained through bootstrapping (1000 replications), all the estimations are weighted with the populational weights provided in the ECVHP06 database.

Table 7: Sensitivity Analysis of the Return to Catalan Proficiency (% Monthly Earnings)

CHANGING SAMPLE SELECTION	$\exp\left(\hat{\delta}\right)-1$	STANDARD ERROR
All the Sample — Interval Regression	0.069	0.021
All the Sample — Int. Reg. + Self Selection	0.193	0.050
Only Males — Int. Reg. + Self Selection	0.139	0.062
Only Females — Int. Reg. + Self Selection	0.230	0.073
Excluding Foreigners — Int. Reg. + Self Sel.	0.143	0.059
CHANGING EXCLUSION RESTRICTIONS	$\exp\left(\hat{\delta}\right)-1$	STANDARD ERROR
Without "Speak Catalan with Parents"	0.181	0.050
Without "Speak Catalan at Home"	0.205	0.050
Without "Speak Catalan with Children"	0.202	0.050
Without "News Catalan TV"	0.195	0.049
Without "Catalan Newspapers"	0.180	0.047
Without "Arrived Younger than 10"	0.186	0.049
Without "Normalization"	0.190	0.049
Without "% Write Catalan (Comarca)"	0.239	0.050
Without "# Books at Home"	0.177	0.049
Without "Read Frequently"	0.213	0.053

Table 8: Language Return by Completed Education

	$\exp\left(\hat{\delta}\right)-1$	STANDARD ERROR
All the Sample — Interval Regression	0.069	0.021
All the Sample — Int. Reg. + Self Selection	0.193	0.050
High Education — Interval Regression	0.064	0.027
High Education — Int. Reg. + Self Selection	0.228	0.069
Low Education — Interval Regression	0.040	0.031
Low Education — Int. Reg. + Self Selection	0.017	0.065
Excluding Foreigners — High Education	0.217	0.073
Excluding Foreigners — Low Education	-0.020	0.076

Note: High Education = Individuals who Completed Secondary or Tertiary Education

Low Education = Individuals with No Education or who Completed Primary Education

APPENDIX

Table 1A: Sample Selection Criteria

Total sample	10,397
Only individuals with valid information for Catalan knowledge and earnings	-2,481
Only Individuals aged 16 to 65	-1,708
Excluding natives with at least one parent born in Catalonia	-2,960
Only individuals who are regularly working when interviewed	-649
Final sample	2,599

Table 2A: Monthly Earnings in Brackets (in Euros) — Selected Sample

EARNING INTERVALS (in Euros)	FREQ.	% SAMPLE
Less than 450	110	4.23
Between 451 and 600	169	6.5
Between 601 and 750	243	9.35
Between 751 and 900	435	16.74
Between 901 and 1050	372	14.31
Between 1051 and 1200	419	16.12
Between 1201 and 1500	373	14.35
Between 1501 and 1800	212	8.16
Between 1801 and 2400	155	5.96
Between 2401 and 3000	67	2.58
More than 3000	44	1.69

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 a Younger Age than 10 = *1 if year of arrival in Catalonia minus year of birth (2005 - age) is minor than 10, 0 otherwise.*

Normalització = *1 if the individual was younger than 12 years old in 1984 (and was living in Catalonia if not native), 0 otherwise.*

Of Books at Home = *mid points of the original age variable collected in intervals (0, 1-10, 11-25, 26-100, 101-500, 501-1000, 1000 or more).*

Job Tenure = *mid point of the original variable collected in intervals (less than 2 years, more than 2 and less than 5 years, more than 5 and less than 10 years, more than 10 and less than 15 years, more than 15 and less than 20 years, more than 20 years).*

Previous Experience = *potential work experience, previous to the current work (age-schooling-job-tenure-6); years of schooling are imputed from individual completed education (schooling equal to 2 if no completed education, 6 if completed primary education, 12 if completed secondary education, 17 if completed tertiary education).*

Table 3A: Earning Regressions by Completed Education

Dependent Variable: Ln(Earnings)	INT. REG. Low Education	INT. REG. +SEL. Low Education	INT. REG. High Education	INT. REG. +SEL. High Education
Constant	6.684***	6.704***	-0.228***	-0.238***
	(0.078)	(0.094)	(0.020)	(0.020)
Native-S.G.	*Ref. Cat.*	*Ref. Cat.*	*Ref. Cat.*	*Ref. Cat.*
Spanish	0.072	0.063	0.067	0.135***
	(0.048)	(0.056)	(0.048)	(0.052)
European	-0.048	-0.054	0.021	0.119*
	(0.074)	(0.078)	(0.052)	(0.063)
African	-0.120***	-0.130***	-0.165***	-0.062
	(0.040)	(0.049)	(0.053)	(0.068)
American	-0.030	-0.040	-0.042	0.067
	(0.043)	(0.050)	(0.035)	(0.053)
Asian	-0.072	-0.081	-0.073	0.034
	(0.065)	(0.081)	(0.111)	(0.120)
YMS/10	-0.029**	-0.027*	-0.010	-0.027*
	(0.014)	(0.015)	(0.015)	(0.016)
Sex (=1 if Female)	-0.364***	-0.364***	-0.228***	-0.238***
	(0.023)	(0.023)	(0.020)	(0.020)
Married	0.025	0.025	0.103***	0.104***
	(0.023)	(0.024)	(0.020)	(0.020)
No Education	*Ref. Cat.*	*Ref. Cat.*	—	—
Primary education	0.093**	0.094**	—	—
	(0.043)	(0.042)		
Secondary Education	—	—	*Ref. Cat.*	*Ref. Cat.*
Tertiary Education	—	—	0.288***	0.271***
			(0.024)	(0.024)
Job Tenure/10	0.073***	0.070***	0.172***	0.193***
	(0.018)	(0.019)	(0.018)	(0.019)
(Previous) Experience/10	0.108***	0.104**	0.123***	0.147***
	(0.041)	(0.043)	(0.035)	(0.037)
Experience^2/100	-0.022**	-0.021**	-0.023*	-0.024**
	(0.009)	(0.009)	(0.012)	(0.012)
Unionized	0.081***	0.080***	0.112***	0.108***
	(0.024)	(0.024)	(0.023)	(0.023)
Part-time	-0.465***	-0.464***	-0.393***	-0.396***
	(0.044)	(0.043)	(0.041)	(0.040)
#Workers>500	0.052*	0.053*	0.013	0.012
	(0.029)	(0.030)	(0.023)	(0.023)
Self-employed	0.194***	0.194***	0.157***	0.152***
	(0.043)	(0.044)	(0.040)	(0.039)
Living in Barcelona	0.037*	0.036	0.029	0.038*
	(0.022)	(0.022)	(0.023)	(0.023)
Catalan Proficiency	0.039	0.017	0.062**	0.205***
	(0.030)	(0.061)	(0.025)	(0.057)
Self-Selection Correction (λ)	—	0.016	—	-0.102***
		(0.037)		(0.037)
$\hat{\sigma}_\varepsilon$	0.298(0.010)	0.298(0.010)	0.329(0.008)	0.328(0.008)
$\hat{\rho}_{\varepsilon u}$	—	0.052(0.135)	—	-0.311(0.113)
Number of Observations	1,082	1,082	1,517	1,517

Note: weighted estimations; robust standard errors for simple interval regressions, bootstrapped standard errors for interval regressions with self-selection correction. High education = completed secondary or tertiary education, Low education = completed primary education or no education.

Últims documents de treball publicats

NUM	TÍTOL	AUTOR	DATA
10.01	Language knowledge and earnings in Catalonia	Antonio Di Paolo, Josep Lluís Raymond-Bara	Febrer 2010
09.12	Inflation dynamics and the New Keynesian Phillips curve in EU-4	Borek Vasicek	Desembre 2009
09.11	Venezuelan Economic Laboratory The Case of the Altruistic Economy of Felipe Pérez Martí	Alejandro Agafonow	Novembre 2009
09.10	Determinantes del crecimiento de las emisiones de gases de efecto invernadero en España (1990-2007)	Vicent Alcántara Escolano, Emilio Padilla Rosa	Novembre 2009
09.09	Heterogeneity across Immigrants in the Spanish Labour Market: Advantage and Disadvantage	Catia Nicodemo	Novembre 2009
09.08	A sensitivity analysis of poverty definitions	Nicholas T. Longford, Catia Nicodemo	Novembre 2009
09.07	Emissions distribution in postKyoto international negotiations: a policy perspective	Nicola Cantore, Emilio Padilla	Setembre 2009
09.06	Selection Bias and Unobservable Heterogeneity applied at the Wage Equation of European Married Women	Catia Nicodemo	Juliol 2009
09.05	La desigualdad en las intensidades energéticas y la composición de la producción. Un análisis para los países de la OCDE	Juan Antonio Duro Moreno, Vicent Alcantara Escolano, Emilio Padilla Rosa	Maig 2009
09.04	Measuring intergenerational earnings mobility in Spain: A selection-bias-free approach	María Cervini Pla	Maig 2009
09.03	The monetary policy rules and the inflation process in open emerging economies: evidence for 12 new EU members	Borek Vasicek	Maig 2009
09.02	Spanish Pension System: Population Aging and Immigration Policy	Javier Vázquez Grenno	Abril 2009
09.01	Sobre los subsistemas input-output en el análisis de emisiones contaminantes. Una aplicación a las emisiones de CH4 en Cataluña	Francisco M. Navarro Gálvez, Vicent Alcántara Escolano	Març 2009
08.10	The monetary policy rules in EU-15: before and after the euro	Borek Vasicek	Desembre 2008
08.09	Agglomeration and inequality across space: What can we learn from the European experience?	Rosella Nicolini	Desembre 2008