doi.org/10.26398/IJAS.210

THE TRANSITION FROM UPPER SECONDARY TO HIGHER EDUCATION: SURVEY INSIGHTS FROM ITALY

Michele Lalla
Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy. ORCID: 0000-0002-1639-7300

Patrizio Frederic[1]
Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy. ORCID: 0000-0001-9073-2878

Abstract. *Two data sets for 2009 were used to compare Italians and immigrants: the European Union Statistics on Income and Living Conditions (EU-SILC) and the Italian Survey on Income and Living Conditions of Families with Immigrants (IM-SILC). A sub-sample of subjects between 20 and 25 years of age was set up, containing individual, family, and contextual variables. Their effects on the choice of tertiary education (yes/no) were assessed using a Lasso method to determine the significant explanatory set of variables through a Bayesian approach also aimed at identifying interaction terms. The transition from high school to higher education showed a complex pattern involving many variables: young women continued with their education more than young men; the educational level of the parents and many components of income entered the model in a parabolic form. Significant contextual factors included the degree of urbanisation and household tenure status. New elements of this study include the sample, the Lasso method in this field, and some empirical results.*

Keywords: *Transition to university, Educational inequality, Educational territorial patterns, Lasso method, Bayesian logistic model.*

1. INTRODUCTION

Tertiary education is not compulsory in almost all educational systems and enrolment decisions constitute a difficult step for students because they are making decisions about their future without knowing much about themselves and/or the likely evolution and needs of society. These decisions may be affected

[1] Corresponding author: Patrizio Frederic, email: patrizio.frederic@unimore.it

by differences in individual characteristics and/or the socio-economic conditions of families, as well as social and contextual conditions in the area where they reside. Additionally, such decisions are likely to impact opportunities for future employment and upward mobility, while individual difficulties and critical family situations may also lead to dramatically lower grades and dropping out of school (Grove et al., 2006; Wintre et al., 2011; Armstrong and Biktimirov, 2013). All these aspects may differ among young immigrants and non-immigrants and, in the case of the former, tertiary education plays an important role not only in terms of investment in human capital, the cultural formation process, and social integration, but also as an instrument for social mobility and transformation, individual development through attuned interactions and collective healing through cooperation (Entwisle and Alexander, 1993; Ichou, 2014; Paba and Bertozzi, 2017; De Clercq et al., 2017).

The first objective of this paper is to point out the differences with respect to citizenship, a binary variable distinguishing between immigrants and non-immigrants (hereinafter also referred to as Italians), and the decision to continue with tertiary education or to discontinue their studies after finishing their upper secondary schooling, using a sufficiently large sample of immigrants in comparison to non-immigrants.

The second objective is to identify the determinants of this transition by using the Lasso method through the Bayesian approach, selecting automatically the interactions between explanatory variables, while also accounting for the marginal effects of individual characteristics, family, and social background. The data were extracted from two surveys (the reference year being 2009) carried out by the Italian National Institute of Statistics (Istat): one is the European Union Statistics on Income and Living Conditions (EU-SILC) restricted to Italy only (IT-SILC) – an annual survey conducted since 2004 coordinated by Eurostat (Istat, 2008; Eurostat, 2009) – and the other being the Italian Survey on Income and Living Conditions of families with Immigrants (IM-SILC),[2] which is a single

[2] Note that the letter S in the acronym EU-SILC is often assumed to mean "Survey", rather than "Statistics". The same has been done here to provide correspondence with the acronym for the Italian Survey on Income and Living Conditions of families with immigrants, where the term "immigrants" refers to individuals without Italian citizenship. The term adopted here for this group is "Immigrant Survey on Income and Living Conditions" (IM-SILC) to obtain a similar structure of the acronyms for the two surveys.

cross-sectional survey (Istat, 2009a) that involved families with at least one immigrant component resident in Italy.

Multinomial choices may be applied to the transition from upper secondary education to employment or to attend post-secondary non-tertiary or tertiary education (Nguyen and Taylor, 2003) involving several alternatives (employed, unemployed, inactive or out of the labour force, and/or distinguishing between various university degree levels). However, this paper focuses mainly on the decision to attend a degree programme rather than the other options. The binary nature of the dependent variable, hereinafter referred to as the "tertiary" (dependent) variable, implies that it is equal to 1 when an individual is attending a tertiary education level, and equal to zero otherwise, i.e., when the student has achieved an upper secondary level of education. It directly involves some specific techniques, such as ordinary logistic regression in the classical approach or a Bayesian approach, both of which were applied here. In the latter case, the set of independent variables was identified with the Lasso method, which simultaneously allows for the selection of the explanatory variables, the interaction terms, and the estimation of the model coefficients.

The paper is organised as follows. Section 2 provides an overview of the theoretical background, and Section 3 illustrates the sample, the data and some descriptive results concerning the main variables used in the subsequent analyses. Section 4 briefly explains the ordinary logistic model and the Bayesian model combined with the Lasso techniques. Section 5 describes the model obtained through the peculiar Lasso techniques for selection of the independent variables and a Bayesian approach for the estimation of parameters. Finally, Section 6 concludes with some comments and remarks.

2. BACKGROUND

Educational decisions that young people face are made at a particular stage in their lives when influences inside and outside the home are strongly felt. In this sense, such decisions strongly depend on both individual and family characteristics, as well as the environment, but also on psychological and school-related factors (Parker et al., 2004; Wilson and Gillies, 2005).

Several researchers have investigated the factors that influence the choices of young people, mainly from a socio-economic point of view, allowing for status inequalities, i.e., how individual, parental, and family characteristics affect and

interact with human capital accumulation (among many others, Brunello and Checchi, 2007; Dustmann, 2008; Van de Werfhorst and Mijs, 2010). The modelling of the decision to attend a tertiary education programme is often based on human capital theory, dating back to Becker (1964), as school students are faced with two alternatives: to invest in education or to enter the labour market (Nguyen and Taylor, 2003). With respect to the explanatory variables, three sets are generally considered in these studies: personal, parental, and family/environmental characteristics.

First, individuals possessing greater ability than others may benefit from investment in further education, implying that educational achievement is an indicator of this ability. Past analyses have shown that other personal characteristics such as gender and ethnicity are significant factors (Perreira et al., 2006; Bubritzki et al., 2018). Here ethnicity was discarded because the focus was on the immigrant/Italian dichotomy. Health conditions, rarely taken into consideration proved to be associated with the choice of continuing in education and training (Lalla and Pirani, 2014; Ichou and Wallace, 2019).

Second, educational decisions reflect and originate from the context of the family, as human capital theory suggests. The effect of family background on assimilation and expectations has been thoroughly analysed and different factors have been identified as relevant in these processes: household size and family composition, educational level of the parents, socioeconomic status, language and expectations of parents, parental support and involvement, cultural background, and income (among many others, for Italy see Luciano et al., 2009; Buonomo et al., 2018). Extensive comparisons of groups of individuals at various stages of their careers have been carried out and many explanations have been given for employment and income inequalities (Glick and Hohmann-Marriott, 2007; Algan et al., 2010; Luthra and Flashman, 2017; Zwsyen and Longhi, 2018). In the absence of (reliable) income data, studies have taken the employment status of parents as proxies, while in the present study various reliable income variables and several occupational variables were included in the models.

Lastly, the social context of the community and the area of residence has also been found to be relevant (Bond Huie and Frisbie, 2000; Perreira et al., 2006; Sleutjes et al., 2018). Schooling has been analysed as a source of inequality between immigrants and natives and/or among different groups of immigrants as well. The social context includes attending kindergarten, previous experiences of success and failure, advice of teachers and peers, and the availability of schools

in the area (Bertolini and Lalla, 2012; Contini, 2013). The school environment can provide strong stimuli for integration in the community as a source of potential comparison with others, and induce motivation for all to improve their knowledge and education. The context of the community of residence may refer to social characteristics of the neighbourhood (Woodraw-Lafield, 2001; Pong and Hao, 2007) and its economic characteristics. Social characteristics have often been represented considering crime level, characteristics of peers, companionship and so on, while the economic factors may refer to the employment/ unemployment rate in the area of residence, the local gross domestic product, and the value added by sector (Bertolini et al., 2015; Zwysen and Longhi, 2018). The local area may provide an important indicator summarising many effects such as segregation and favourable or unfavourable economic conditions, thus affecting decisions on whether to continue in education. Here, macro-regions and the degree of urbanisation were considered as useful indicators of immigrant concentrations, sometimes as a result of settlement preferences, as some regions and towns attract more immigrants than others. Moreover, some indicators of housing conditions, personal and family possessions were introduced into the models.

In conclusion, participation in education is a highly complex phenomenon, offering countless avenues for investigation and analysis. Consequently, it has been widely studied across the globe, particularly during and after the COVID-19 pandemic (2020-2023). In Italy too, an extensive literature has also emerged in recent years. The transition from upper secondary school to tertiary education has been investigated at the national level using time series (Minerva et al., 2022) or through macro-socioeconomic indicators at the provincial level (Bertolini et al., 2015; Paba and Bertozzi, 2017), as well as via local/ regional surveys (Vettori et al., 2020; Rondinelli et al., 2024). Other studies have focused on students' choices regarding geographic mobility (Usala et al., 2023; Vittorietti et al., 2023), which can be seen as a form of internal migration of students. Given the vast number of articles on the topic, for the sake of brevity, it is worth noting at least that several studies have utilised Program for International Student Assessment (PISA) scores as a dependent variable, analysing student performance by comparing immigrants to Italians. After controlling for the relevant variables in the PISA data set, these studies found a negative performance gap for immigrant students compared to Italians and Europeans, largely due to the immigrants'

limited access to economic resources and educational materials (Murat, 2012; Murat and Frederic, 2015; Schnell and Azzolini, 2015).

3. DATA SOURCES AND PRELIMINARY EVIDENCE

The data were extracted from two surveys carried out by the Italian National Institute of Statistics (Istat) with 2009 as the reference year.

The first one was the EU-SILC, an annual survey aimed at gathering information based on nationally representative random samples of private households in each European country concerning individual socio-demographic characteristics, micro-level data on income, poverty, social exclusion and living conditions using a unique sampling design and identical definitions of the concepts currently used for these purposes (Eurostat, 2009). As a result, the target population refers to all private households and all persons aged 16 and over. The nation considered was Italy, IT-SILC, and the selected reference year, 2009, was a necessary choice because the IM-SILC (see below) was carried out by Istat only in that year. The IT-SILC target information is distributed over four different groups or data sets, each one grouping different variables: (D) Household Register, (H) Household Data, (R) Personal Register, and (P) Personal Data. The four files were matched to obtain a complete file with information at different levels. In the resulting matched file, the total number of cases was equal to the number in the personal register file (R): 51,196. However, the number of useful and manageable records remained the same as the number of personal records, each one corresponding to an interviewed individual, for a total of 43,636. Obviously, the 0–14-year age class was empty because no individuals under the age of 16 were interviewed.

The IM-SILC was funded by the Ministry of Labour and Social Policies and conducted by Istat in 2009 only, i.e., it was a one-shot survey design using a national probability sample of the population, greater than or equal to 16 years old, residing in private households in Italy. The IM-SILC design was similar to the IT-SILC project. The specificity of the reference population in the IM-SILC, compared to the IT-SILC, involved numerous expedients to improve the representativeness of the sample. (1) The sample design at the origin of the IM-SILC was based on the extraction of municipalities as primary sampling units, subdivided on the basis of the degree of urbanisation [densely-, moderately-, and thinly-populated areas (see Eurostat, 2009)] and taking into account the

distribution of the main groups of foreign nationals in Italy, reducing the risk of excluding some groups of foreign nationals who could be particularly concentrated in some areas. (2) Non-respondent families were replaced with other families of the same nationality, minimizing self-selection of the most collaborative nationalities and consequent bias. (3) The questionnaires were translated into the ten most common languages among foreigners residing in Italy, to support the interviewers and facilitate the interviewees' understanding of the questions. (4) The sample was post-stratified at a geographical distribution level, taking into account, in addition to the usual constraints on the known total population, the number of families with immigrants and the foreign population classified into the 13 main nationalities residing in Italy, for better calibration with respect to the reference population (Istat, 2009b). In the resulting matched file, the total number of cases was equal to the number of cases in the Personal Register file (R): 15,036. However, the number of useful and manageable records remained the same as the number of personal records, which was 11,611, each one corresponding to an interviewed individual. Again, the 0–14-year age class was empty.

The target sample was obtained by first selecting individuals in the age range of 20 to 25, obtaining a sample of 3,166 cases. Then, in this sample, the eligible cases were only those individuals whose highest attained International Standard Classification of Education (ISCED) level (UNESCO, 2012) was equal to 3 (upper secondary education) or 4 (post-secondary non-tertiary education). The final target sample consisted of 2,874 individuals (Table 1). Further details about these two data sets can be found in Eurostat (2009), as IM-SILC is largely similar to IT-SILC. The variables introduced in the models are described in the Appendix (§7.1) and Lalla and Frederic (2020).

Table 1 shows that the sample of young immigrants aged 20 to 25 (inclusive) represents approximately 4.5% of the total. This low percentage, i.e. the small sample size, results in certain limitations. For instance, in the model estimation, many categorical variables are unable to discriminate properly between immigrants and Italians because these variables do not present observations for certain modalities among immigrants. Additionally, relationships that are statistically significant in the real world may not be significant in this sample due to the small number of immigrants in the survey. For brevity, this is sufficient to justify the use of these two data sets, even if they are not updated, because they increase the number of immigrants to 22.4%. Moreover, the data set adopted

offers unique advantages not found in local surveys, such as a national perspective, detailed information on the health conditions of individuals and their parents, living conditions, and individual and family incomes collected with precision and accuracy, while distinguishing between their sources.

Table 1: Absolute frequencies and row percentages of the sample by the type of survey (TOS) and age

TOS\ Age	**20**	**21**	**22**	**23**	**24**	**25**	**Total**
IT-SILC Italians	388	416	382	360	373	311	2230
%	17.4	18.7	17.1	16.1	16.7	14.0	100
IT-SILC Foreigners	18	28	14	19	13	13	105
%	17.1	26.7	13.3	18.1	12.4	12.4	100
IM-SILC	62	86	93	99	88	111	539
%	11.5	16.0	17.3	18.4	16.3	20.6	100
Total	468	530	489	478	474	435	**2874**
%	16.3	18.4	17.0	16.6	16.5	15.1	100

3.1 BIVARIATE AND TRIVARIATE ANALYSIS OF THE MAIN VARIABLES

The relationship between the tertiary (binary) dependent variable and the ISCED Level Currently Attended (ILCA) showed that 55.3% of individuals, with an ISCED level equal to 3 or 4, were not enrolled in tertiary education courses (termed "not-attending"), while 44.7% were currently attending tertiary education (Table 2).

The ILCA was examined with respect to several qualitative variables and revealed many significant relationships. With respect to gender, $CS(2)= 30.15$ ($p<0.000$), where CS(g) stands for "Chi-Square with g degrees of freedom": women tended to be attending more than men (49.5% versus 39.4%), with the exception of the post-secondary non-tertiary category, in which the percentage of men (1.5%) was unexpectedly equal to that of women (1.5%). The percentage of women not in education was lower than that of men: 49.0% versus 59.2%. The ILCA showed a significant relationship with respect to citizenship, $CS(2)= 115.33$ ($p<0.000$): fewer immigrants attended tertiary education than Italian citizens (26.6% versus 50.0%), while the percentage of immigrants not in education was higher than that of Italians (72.4% versus 48.4%), supporting well-known empirical evidence of difficulties relating to the integration process for

immigrants who are also conditioned by scarce economic resources to be allocated to education.

Table 2: Absolute frequencies and row percentages of the tertiary (binary) education (EDU) dependent variable by the ISCED level currently attended (ILCA)

Tertiary\ ILCA	Not-attending	Post-Secondary EDU	Tertiary EDU	Total
Tertiary = 1			1285	1285
%			100.0	100
Tertiary = 0	1546	43		1589
%	97.3	2.7		100
Total	1546	43	1285	**2874**
%	53.8	1.5	44.7	100

A significant relationship emerged between the ILCA and self-perceived health, CS(2)= 10.87 ($p<0.004$), implying that individuals perceiving fair or bad or very poor health tended to discontinue their studies (66.0%) with respect to those perceiving good or very good health (53.1%), see Ichou and Wallace (2019). The ILCA was related: (*i*) to the degree of urbanisation, CS(4)= 26.26 ($p<0.000$) – as the density of the area increased, the ILCA increased, and (*ii*) to the Italian macro-regions, CS(8) = 24.27 ($p<0.002$) – as industrialisation and the possibility of finding employment increased, the percentage of individuals continuing their education decreased. This could be a possible effect. Industrialisation has a positive and indirect impact on primary education, but contributes less to increasing participation rates at higher levels of education and/or to the development of human capital through schooling (Montalbo, 2020). These effects may vary across countries due to cultural, political, and social factors (among many others, Le Brun et al., 2011; Federman and Levine, 2005) or over time (Minerva et al., 2022). In Italy, a weak regressive effect was observed in the transition to higher education in areas with abundant employment opportunities, where individuals may choose to forgo further education. The costs of tertiary education, coupled with the relatively low expected returns from a university degree, are likely to prompt some individuals to enter the labour market rather than continue their studies (Paba and Bertozzi, 2017).

The ILCA was related to the maximum ISCED level attained by the parents, CS(12)= 198.80 ($p<0.000$), but it was also related to the father's and mother's level of educational attainment. The ILCA yielded significant relationships also

with several variables describing the working conditions of both parents, although the correlation was often weak.

The ILCA was analysed with respect to the main quantitative variables.

The age of fathers, with respect to the ILCA and citizenship, showed that the fathers of immigrants were younger than the fathers of Italians by about 12 years. Similarly, the mothers of immigrants were younger than the mothers of Italians by about 12 years.

Disposable Family Income (DFI) per capita (in thousands of euros) is reported in Table 3 by the ILCA and citizenship. On average, the DFI per capita for immigrants was reported to be significantly lower than that of Italians by about 4,000 euros: about 35.7%.

Table 3: Absolute frequencies (n), means, and standard deviations (SD) of the disposable family income per capita (in thousands of euros) by citizenship and by the ISCED level currently attended (ILCA) by their children (E=Education)

Citizenship\ ILCA		Not-attending	Post-Secondary E	Tertiary E	Total
Italian citizen:	*n*	1080	36	1114	2230
	Means	*11.389*	*11.508*	*12.543*	*11.967*
	SD	6.999	6.432	9.147	8.153
Foreign citizen:	*n*	466	7	171	644
	Means	*7.777*	*5.868*	*7.563*	*7.699*
	SD	5.315	3.489	5.877	5.452
Total:	***n***	1546	43	1285	**2874**
	Means	*10.300*	*10.590*	*11.880*	*11.011*
	SD	6.742	6.376	8.942	7.835

The size of immigrant families proved to be smaller than those of Italians, although non-significantly for the marginal effects of citizenship with $F(1;2868)= 1.16$ ($p=0.282$), but it was statistically significant for the ILCA ($p<0.001$) and for their interaction ($p<0.001$). Given that the total fertility rate of immigrant women is generally higher than that of Italian women, one might expect that the size of immigrant families would be larger than that of Italians. However, many immigrants come to work in Italy without their families, and this presumably accounts for the decrease in the size of immigrant families.

Citizenship was examined with respect to some other variables, even if it was not a target dependent variable. Its relationship with the maximum ISCED level attained by parents was statistically significant, $CS(6)= 217.01$ ($p<0.000$).

The two-sample Kolmogorov-Smirnov test (K-S) of the equality of distribution functions showed that they were statistically different (combined K-S= 0.265, p<0.000). Immigrant parents had attained upper secondary education levels more frequently than Italians (73.1% versus 42.4%) as expected because other empirical findings have revealed this tendency (Bertolini and Lalla, 2012; Bertolini et al., 2015). In fact, this was also the case with vocational qualifications achieved through post-secondary non-tertiary education (1.6% versus 0.6%). On the contrary, this behaviour was not evident for post-tertiary education, as immigrant parents tended to avoid this type of education (0.9% versus 2.9%), seeking employment immediately after a degree because of scarce economic resources compared to non-immigrants (Forster and van de Werfhorst, 2020).

Citizenship was significantly related to the degree of urbanisation, CS(2)= 19.18 (p<0.000), which was confirmed by the two-sample K-S test of the equality of distribution functions (combined K-S= 0.078, p<0.004). Immigrants tended to settle in densely populated areas more than Italians (36.2% versus 35.3%) or in intermediate areas (46.6% versus 39.6%). As expected, the reverse was true for thinly populated areas (17.2% versus 25.1%). Interaction of foreign nationals with other foreign nationals is facilitated in densely populated areas, but at the same time, integration measures for immigrants may be more efficient in highly populated cities than in other areas.

Citizenship showed a significant relationship with the Italian macro-regions, i.e., the geographical subdivision of Italy into five zones (the North-West, North-East, Centre, South, and the Islands): CS(4)= 50.58 (p<0.000). The immigrants tended to establish themselves in the North-East (24.4% versus 20.0% of Italians), in the North-West (19.9% versus 16.6%), in the Centre (25.6% versus 23.4%), where Rome attracts many immigrants, and the Islands (14.8% versus 10.9%), prefiguring a sort of embryonal segregation (Andersson et al., 2018). If the North-South contrast framework is applied, then the data show that immigrants tend to settle more frequently in the North than in the South. However, the Islands exhibit percentage differences similar to those in the North, particularly Sicily, as they are primary points of entry. Immigrants often need time and favourable conditions to continue the journey toward northern Italy and other European countries.

Citizenship yielded a significant relationship with the index summarising the total self-perceived health of parents, CS(3)= 134.99 (p<0.000) and the K-S test (combined K-S= 0.245, p<0.000), implying that when the number of health

problems increased, the percentages of Italians decreased, but they were always higher than that of immigrants, although in slightly nonlinear way. For example, the percentage of immigrants with parents without health problems was greater than that of Italians: 84.0% versus 59.5%.

Citizenship proved to be associated with many variables describing working conditions. The relationship between citizenship and the parents' activity status was statistically significant: CS(4)= 105.20 (p<0.000). Immigrants presented lower percentages than those of Italians for the category "both parents employed" and the category "at least one parent is retired": 21.9% and 1.4% versus 34.1% and 9.8%, respectively. Immigrants presented higher percentages than those of Italians for "employment of father only" and for "employment of mother only": 41.6% and 19.3% versus 31.9% and 13.8%, respectively. Citizenship revealed a significant relationship with the maximum position of parents on the job, CS(4)= 134.03 (p<0.000) and K-S= 0.489 (p<0.000), implying that with higher positions (i.e., when one of the parents has a high position), the percentage of Italians increases, although in a slightly nonlinear way. For example, there was a lower percentage of immigrants in managerial positions with respect to Italians: 0.9% versus 4.9%. The difference concerning the position of executive director was 1.1% versus 6.7%. Citizenship yielded a significant association also with the working conditions of parents, CS(5)= 147.14 (p<0.000).

4. MODEL BY BAYESIAN LASSO SELECTION OF REGRESSORS

Let Y be the binary variable denoting for the i-th individual, the dichotomised choice with respect to attending a tertiary level of education (y=1) versus not attending (y=0). Let $\mathbf{x}_i$ be a vector of regressors. Let π_i be the probability that Y=1 given $\mathbf{x}_i$. Let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)$ be the parameters vector of the model. The logit model is

$$\pi_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}) / \left|1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})\right| \qquad (1)$$

The Lasso method (Tibshirani, 1996) was applied to carry out the estimation and model selection. In fact, it is a procedure involving an additional penalisation term, L_1, summed up to the negative log-likelihood of the model that depends on an additional parameter named λ, $\lambda \geq 0$. More precisely, let $\Phi(\cdot)$ be the objective function of the logit model, hence

$$\Phi(\boldsymbol{\beta},\lambda;\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}\left[y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)\right] + \lambda \sum_{j=0}^{K} |\beta_j| \qquad (2)$$

where $\mathbf{x} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$, $\mathbf{y} = (y_1,\ldots,y_n)$, and $\pi_i = \pi_i(\mathbf{x}_i,\boldsymbol{\beta})$. Finally, $\Phi(\cdot)$ is minimised for different values of parameter λ. It should be noted that when λ=0, then $\Phi(\cdot)$ is the negative log-likelihood of the logit model. On the other hand, larger values of λ yield many β's exactly equal to zero.

In many penalised methods, $\Phi(\cdot)$ can be interpreted as the negative logarithm of a posterior distribution in a purely Bayesian fashion. Let $p(y_i|\mathbf{x}_i,\boldsymbol{\beta}) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$ be the usual logit model in the usual Bayesian notation, and let $p(\boldsymbol{\beta}|\lambda) \propto \exp(-\lambda \sum_{j=0}^{K}|\beta_j|)$ be the Laplace prior distribution on coefficients $\boldsymbol{\beta}$; then the posterior distribution is

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{x},\mathbf{y},\lambda) \propto & \quad p(\mathbf{y}|\mathbf{x},\boldsymbol{\beta})\ p(\boldsymbol{\beta}|\lambda) \\ \propto & \quad \prod_{i=1}^{n} p(y_i|\mathbf{x}_i,\boldsymbol{\beta})\ p(\boldsymbol{\beta}|\lambda) \\ = & \quad \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}\ \exp\left(-\lambda \sum_{j=0}^{K}|\beta_j|\right). \end{aligned} \qquad (3)$$

Note that $\Phi(\boldsymbol{\beta},\lambda;\mathbf{x},\mathbf{y}) = -\log\left[p(\boldsymbol{\beta}|\mathbf{x},\mathbf{y},\lambda)\right]$. Hence the Lasso method can be interpreted as a maximum posterior Bayesian estimation method, where the prior distribution on β's is Laplace and λ plays the role of the hyper-parameter. Let $\hat{\boldsymbol{\beta}}_\lambda$ be the minimizing of $\Phi(\cdot)$, then $\hat{\boldsymbol{\beta}}_\lambda$ is the maximum posterior estimation of $\boldsymbol{\beta}$ conditioned to the data and λ:

$$\hat{\boldsymbol{\beta}}_\lambda = \arg\min_{\boldsymbol{\beta}} \Phi\left(\boldsymbol{\beta},\lambda;\mathbf{x},\mathbf{y}\right) = \arg\max_{\boldsymbol{\beta}} p\left(\boldsymbol{\beta}|\mathbf{x},\mathbf{y},\lambda\right). \qquad (4)$$

The choice of parameter λ plays a crucial role in the estimation procedure. Many different studies have focused on this issue; see Zou, Hastie, and Tibshirani (2007) for an extensive review. In addition to the classic AIC and BIC criteria, a *k-fold Cross Validation* (CV) procedure and the *One Standard Error Rule* (1SE) have been proposed. The CV procedure consists of randomly partitioning the original sample into *k* equal-sized subsamples (usually *k*= 5 or *k*= 10). Of the *k*

subsamples, a single subsample is retained as validation data for testing the model and the remaining (k–1) subsamples are used as training data. The process is repeated k times, and each of the k subsamples is used exactly once as validation data. The CV for a given λ is the average of binomial deviance in each step. The optimal value of λ is

$$\lambda_{CV} = \arg\min_{\lambda} \mathrm{CV}(\lambda). \tag{5}$$

In order to achieve greater regularisation, the 1SE rule consists in choosing $\lambda_{1\mathrm{SE}} > \lambda_{1\mathrm{CV}}$ such that $\mathrm{CV}(\lambda_{1\mathrm{SE}}) = \mathrm{CV}(\lambda_{\mathrm{CV}}) + \mathrm{SE}[\mathrm{CV}(\lambda_{\mathrm{CV}})]$, where $\mathrm{SE}[\mathrm{CV}(\lambda_{\mathrm{CV}})]$ is the standard error estimated in the k steps.

It is well known (Hastie et al., 2015) that CV estimates prediction error at any fixed value of the tuning parameter, and thus by using it, it is implicitly assumed that achieving the minimal prediction error is the goal, which is not the case here. The 1SE rule is the best candidate for achieving the goal of recovering the true model. Actually, 1SE adds more regularisation than CV. As a result, the 1SE rule was used for selecting the variables.

The model was estimated using the glmnet (Friedman et al., 2010) package in R (R Core Team, 2019). The glmnet package, like many other penalised likelihood packages, provides point estimation for coefficients $\boldsymbol{\beta}$ and statistics for evaluating the CV, but it does not provide confidence intervals for the parameters or standard errors. However, it is possible to draw samples from the posterior distribution $p(\boldsymbol{\beta}|\mathbf{x}, \mathbf{y}, \lambda_{1\mathrm{SE}})$ and then to perform a full Bayesian analysis.

5. OUTCOMES OF THE LOGISTIC MODEL

The interpretation of coefficients is not easy, and the odds ratios (OR) reported in Table 4 and displayed in Figure 1 facilitate data interpretation only at first sight because calculations are necessary to quantify probabilities (Gould, 2000; Kleinbaum, 1994). In fact, OR is equal to odds (if the examined variable is incremented by 1)/ odds (if that variable is not incremented) or, more formally,

$$\mathrm{OR}=\left[P(y=1\,|\,x+1)\big/\big(1-P(y=1\,|\,x+1)\big)\right]\Big/\left[P(y=1\,|\,x)\big/\big(1-P(y=1\,|\,x)\big)\right]. \tag{6}$$

Moreover, Table 4 only presents interaction terms of the first order because the analysis of interaction orders was limited to the first order to simplify interpretation. The interactions are indicated by the symbol ×, which is read as "by".

Let $\mathbf{x}_b$ be the binary variables. Let $\mathbf{x}_c = \boldsymbol{\mu}$ be the mean values of the continuous regressors, limited to the ages of individuals, which can never be zero in practice. Note that: (1) the product of two binary variables is again a binary variable, (2) the percentage of variation of the reference probability, $\pi_{i|\mathbf{x}_b=\mathbf{0}\wedge\mathbf{x}_c=\boldsymbol{\mu}}$, is given by [100*(OR–1)] and is reported below in parentheses. The probability of having y=1 (i.e., of continuing one's education) was equal to $\pi_{i|\mathbf{x}_b=\mathbf{0}\wedge\mathbf{x}_c=\boldsymbol{\mu}} = \mathbf{0.120}$, calculated at the mean values of the continuous regressors $(\mathbf{x}_c = \boldsymbol{\mu})$ and the binary variables equal to 0 $(\mathbf{x}_b)$. A binary variable having an OR greater than 1 implied that the group represented by the binary variable equal to 1 had a higher probability of having y=1 than the group identified by the binary variable equal to 0; for example, for women with an OR=1.777, the probability of continuing their education was +77.7% greater than that of men. In other words, $\pi_{w|\square}$= 1.777×0.120= 0.213, which was +77.7% greater than the probability of men: 0.120. Note that the dot in the index means keeping all other variables fixed, i.e., the binary and the continuous variables other than age being equal to zero. The successive binary variable having an OR>1 in Table 4 was "PES (Parents' Employment Status) is inactive" (x_1) × "Family living in a densely populated area" (x_2), denoted by x_{12}, which showed an OR=1.697, meaning that the odds of the event y=1, when x_{12} =1 (both x_1 and x_2 are equal to 1), were +69.7% greater than the odds of the event y=1, when x_{12} =0. Similarly, highly significant probabilities of continuing in education were observed for other interaction terms: "Father with permanent contract" × "Only mother employed" (+95.7%), "Father with permanent contract" × "Parents are managers or executives" (+132.1%), "Mother with permanent contract" × "Father is limited by health" (+64.7%), "Father with term contract" × "Mother is limited by health" (+266.5%, which is an unbelievable outcome), "TSH (Tenure Status of Household): Subtenant" × "Family living in a moderately populated area" (+46.6%), and "TSH: Free" × "Assets reduction for needs" (+173.3%).

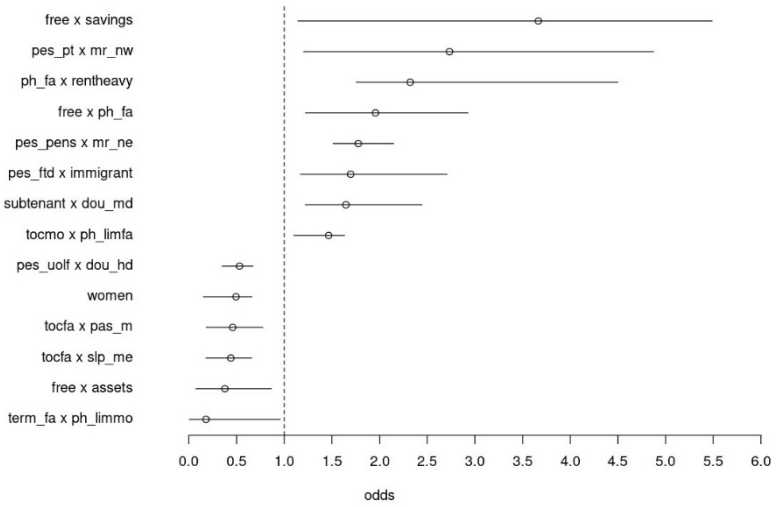

Figure 1: Odds ratios of dichotomous variables in the Bayesian model

In short, gender, favourable and stable parents' working conditions, and good actual and self-perceived health conditions strongly affected the probability of continuing from upper secondary to tertiary education, although this occurred in interaction with other factors.

The binary variables having an OR lower than 1 implied that the group represented had a lower probability of having y=1 with respect to the complementary group. In Table 4 there are six (interaction) binary variables with an OR lower than 1. For example, "Father perceives poor health" × "Rent is burdensome" had an OR=0.440 and hence its complement to one, expressed as a percentage, was equal to [100*(0.440–1)] = –56.0%. As a result, the probability of continuing in education amounted to −56.0% of the probability of the complementary group whose fathers did not perceive poor health and a burdensome rent. In other words, the group with x_{12}=1 had a probability equal to $\pi_{x_{12}=1|\square}$= 0.440×0.120= 0.053. The other five interactions were: "PES= pensioners" × "North-Est" (−50.6%), "PES: part-time" × "North-West" (−62.2%), "PES: full-time employee" × "immigrant" (−46.9%), "TSH= Free" × "Father: poor health" (−54.1%), "TSH= Free" × "Savings" (−82.1%). It is worth noting that the effect of "PES= full-time employee" × "immigrant" (–46.9%) may seem counterintuitive, as full-time parental employment would typically

increase the likelihood of continuing in education. However, despite this, immigrants were still less likely than Italians to go on to tertiary education.

Table 4: Logistic regression with Lasso method and Bayesian approach: Estimated odds ratio (OR), standard errors (SE), p-values (p), and means (M)

B=Binary/ C=Continuous regressor	OR	SE	*p*	M
B- Women	1.777	0.263	0.000	0.530
C- [(Individual's age)/10]^2	0.714	0.044	0.000	5.064
C- (Father's age)/10	1.175	0.073	0.003	4.973
C- (Mother's age)/10	1.548	0.094	0.000	4.727
C- (Education Level of Father: years)^2	1.003	0.001	0.000	1.552
C- FDPI= (Father's DPI)/ 10000	1.452	0.070	0.000	2.372
C- MDPI= (Mother's DPI)/ 10000	1.285	0.062	0.000	1.248
C- FTIPC= (Family's total income per capita)/ 10000	0.314	0.046	0.000	1.101
Interactions of first order				
B- (Father: poor health) × (Burdensome rent)	0.440	0.156	0.011	0.023
B- (PES= Parents' Employment Status: pensioners) × (North-Est)	0.494	0.188	0.031	0.017
B- (PES: inactive) × (Densely populated area)	1.697	0.428	0.048	0.043
B- (PES: part-time) × (North-West)	0.378	0.237	0.042	0.008
B- (PES: full-time employee) × immigrant	0.531	0.092	0.000	0.121
B- (Father: PC= permanent contract) × (Only mother employed)	1.957	0.544	0.013	0.038
B- (Father: PC) × (Parents: manager/ executive)	2.321	0.737	0.013	0.048
B- (Mother: PC) × (Father: limited by health)	1.647	0.322	0.010	0.065
B- (Father: Term C.) × (Mother: limited by health)	3.665	1.776	0.011	0.010
B- (TSH^+: Subtenant) × (Moderately populated area)	1.466	0.175	0.001	0.246
B- (TSH^+: Free) × (Father: poor health)	0.459	0.186	0.025	0.016
B- (TSH^+: Free) × (Assets reduction for needs)	2.733	1.041	0.010	0.016
B- (TSH^+ [Tenure Status of House.]: Free) × Savings	0.179	0.220	0.023	0.003
Intercept	0.043	0.029	0.000	
Bayesian Pseudo-R square	0.227	*n* =	2874	

In short, unstable and unfavourable working conditions of parents, poor actual and self-perceived health conditions of parents, and critical and costly tenure status of the household negatively affected the probability of making the transition from upper secondary school to tertiary education, although this emerged through the interaction terms.

The continuous variables. The individual's age (range 20-25), expressed in decades, showed a parabolic and negative impact on education paths, while the ages of both parents revealed a linear positive impact on the probability of making the transition to higher education. The other continuous single variables (which may be conceptually and concretely equal to 0) entering the model showed significant effects on going on to higher education. With the increase in the parents' educational level, the probability of continuing in education increased quadratically. The father's and mother's disposable personal income (FDPI and MDPI) indicated a linear positive effect (Ochsen, 2011; Krause et al., 2015), whereas the family's total income per capita (FTIPC) yielded an unexpected negative effect, but perhaps the father's income balanced out the effect of the mother's income. In fact, FTIPC included both FDPI and MDPI. However, the algebraic sum of their impacts remained positive, implying the importance of welfare programmes to help families experiencing economic (and physical) difficulties, with the specific aim of reducing the number of students not continuing with their education. The trends of $\pi_i = P(Y = 1)$ for FDPI and FTIPC are illustrated in Figure 2.

The main fault of the Lasso method in selecting significant explanatory variables concerns the possibility of selecting a theoretically unjustifiable variable, such as "Father with term contract" × "Mother is limited by health" (+266.5%) or of neglecting some important variables in the model.

The same model was estimated with the ordinary logistic procedure and the obtained odds ratios were approximately equal to those reported in Table 4, except for "Father with term contract" × "Mother is limited by health" involving an amount equal to +142.4%. Moreover, only the regressor "TSH= Free" × "Savings" was not significantly different from zero (p=0.125). The Hosmer and Lemeshow goodness-of-fit test for this classical estimation of the Bayesian model, not reported here, indicated a well-fitting model, with a p-value of 0.223, suggesting that the model's estimates fit the data at an acceptable level.

Another model was obtained using classical logistic regression with marginal effects, starting from the complete set of 65 regressors and applying backward selection. The aim was to verify the differences between the Bayesian approach, the focus of this study, with the classical method. The resulting estimates are presented in Table 5. The model also proved to be sufficiently robust to changes in the reference/base category of qualitative variables. The number of final regressors increased: 24 in the classical model versus 21 in the

Lasso Bayesian one. The components of income played a complex effect revealing eight regressors in the classical model versus three regressors in the Lasso Bayesian one.

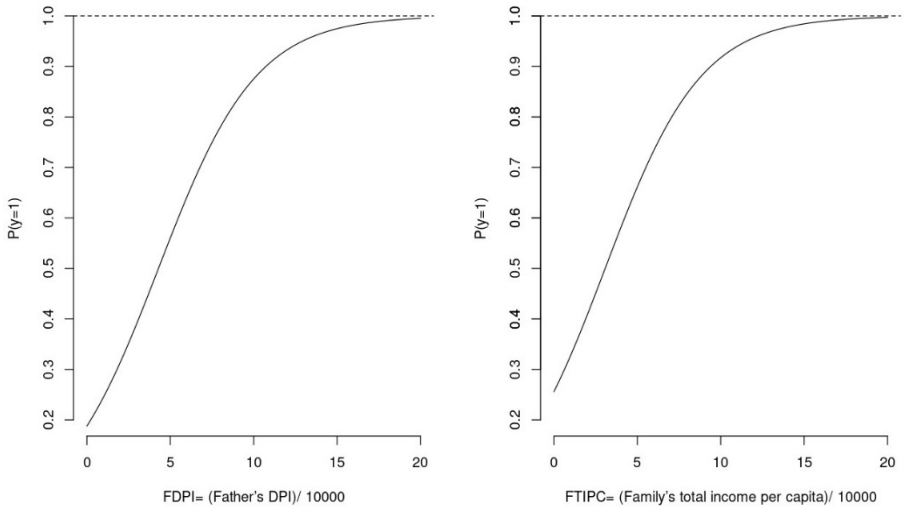

Figure 2: The probability of attending tertiary education in function of the father's disposable personal income (FDPI) and the family's total income per capita (FTIPC)

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are reported in Table 5, making it possible to compare them with those in Table 7 in the Appendix. AIC and BIC do not make much sense in comparing Bayesian and classical models, so they are not reported in Table 4. Moreover, the Hosmer and Lemeshow goodness-of-fit test indicated an unsatisfactory fitting model as its p-value was slightly lower than 0.05, implying that the model's estimates did not fit the data at an acceptable level. Since this model had only an illustrative function, it was not explored in depth. However, more information on how the regressors act on the dependent variable is provided in the Appendix (§7.2).

Table 5: Logistic regression with marginal effects: Estimated odds ratio (OR), standard errors (SE), p-values (p), and means (M)

Regressor	OR	SE	*p*	M
Woman	1.661	0.156	0.000	0.530
DPI= (Disposable Personal Income)/ 10000	0.204	0.023	0.000	0.521
DPI2 = DPI^2	1.226	0.034	0.000	0.851
SPH: Self-Perceived Health	0.408	0.093	0.000	0.055
SPH: limitation in activities	1.934	0.440	0.004	0.048
Macro-region: South	1.299	0.136	0.013	0.260
Degree of urbanisation: high density	1.286	0.156	0.038	0.355
Degree of urbanisation: average density	1.337	0.157	0.014	0.412
$[(\text{Father's age})/10]^2$	1.015	0.007	0.020	26.04
(Mother's age)/10	2.785	0.656	0.000	4.727
$[(\text{Mother's age})/10]^2$	0.934	0.023	0.005	23.55
ELF2 = $(\text{Education Level of Father: years})^2$	1.002	0.001	0.002	133.6
ELM = Education Level of Mother: years	1.078	0.018	0.000	11.01
FDPI= (Father's DPI)/ 10,000	1.198	0.061	0.000	2.372
MDPI= (Mother's DPI)/ 10,000	1.325	0.080	0.000	1.248
MDPI2 = $MDPI^2$	0.985	0.005	0.003	4.205
FTI= [(Family's Total Income)/ 10,000]	0.844	0.044	0.001	3.969
FTI2= FTI^2	1.008	0.003	0.010	24.70
$\text{FTIPC2= [(Family's total income per capita)/10,000]}^2$	0.935	0.023	0.006	1.826
SLP (Skill Level of Parents): manager or executive	1.574	0.289	0.014	0.095
SLP: employee parent	1.577	0.252	0.004	0.104
PES: parents' unemployed or inactive	1.809	0.273	0.000	0.106
Number of optional facilities in home	1.219	0.039	0.000	4.383
Repayments of loans to banks	0.752	0.080	0.007	0.254
Intercept	0.002	0.001	0.000	1.000
Pseudo-R square	0.251	*n* =	2874	

Note: Log-Lik= -1480.1, Akaike inf criterion= 3010.1, Bayesian inf criterion= 3159.2

For the sake of brevity, no further observations on the differences will be made here, but it should be noted that the traditional fit measures are less applicable in Bayesian models. The Bayesian pseudo-R^2 is reported in Table 4 because it provides a more comprehensive assessment of model fit. Moreover, the classification table presented below indicates how well the model predicts the actual outcomes. Metrics such as false positive and false negative rates are particularly informative for understanding how the model performs in predicting whether an individual will continue or decide not to continue their education.

The performance in the correct classification seemed better in the classical model than in the Lasso Bayesian one: the ordinary post-estimation statistics are reported in Table 6, where it is possible to consider the variations.

Finally, the classification of error rates was computed for $\hat{\boldsymbol{\beta}}_{\lambda_{1SE}}$ by assigning $\hat{y}_i = 0$ if $\hat{\pi}_i = \exp(x'_i \boldsymbol{\beta}_{\lambda_{1SE}}) \Big/ \left[1 + \exp(x'_i \boldsymbol{\beta}_{\lambda_{1SE}})\right] < 0.5$, and $\hat{y}_i = 1$, if $\hat{\pi}_i \geq 0.5$. The misclassified number of $\hat{y}_i$ equal to zero was 475 out of 2874, which is a false negative (minus) error rate equal to 37.0% and the misclassified number of $\hat{y}_i$ equal to one was 370 out of 2874, which is a false positive (plus) error rate equal to 23.3% (Table 6). The performance of the logistic model seemed to be slightly better than the Lasso model: the false negative rate was 28.4%, while the false positive rate was 22.8%. In the logistic model, the overall misclassification error rate was equal to 25.3% versus 29.4% in the Lasso model.

Table 6: Performance classification of the logistic and Lasso models (T=Tertiary)

	Lasso Model			**Logistic model**		
Classified\ T	**T=1**	**T=0**	**Total**	**T=1**	**T=0**	**Total**
Positive +	810	370	1180	920	362	1282
Negative –	475	1219	1694	365	1227	1592
Total	**1285**	**1589**	**2874**	**1285**	**1589**	**2874**
False ± rate	37.0	23.3		28.4	22.8	
for true T=0/1	P(– | T=1)	P(+ | T=0)		P(– | T=1)	P(+ | T=0)	
False ± rate	28.0	31.4		22.9	28.2	
for classified ±	P(– | $\hat{y}$ = –)	P(+ | $\hat{y}$ =+)		P(– | $\hat{y}$ = –)	P(+ | $\hat{y}$ =+)	

6. CONCLUSIONS

The key empirical evidence may be summarised as follows. In general, more women tended to continue in education than men. More women than men attended tertiary education (49.5% versus 39.4%), whereas the percentage of women not in school was lower than that of men: 49.0% versus 59.2%. Fewer young immigrants enrolled on education programmes than young Italians: 26.6% versus 50.0%. As a result, the percentage of immigrants not enrolled in education was higher than that of Italians (72.4% versus 48.4%).

In the model, self-perception of health was associated with enrolment in education: individuals with fathers perceiving poor health and a burdensome rent were 56.0% more likely not to continue their education. Immigrant status (i.e.,

citizenship) was not preserved as marginal effects in the explanatory variables set determined by the automatic model selection procedure of the Lasso method combined with Bayesian logistic model, but it demonstrated a combined effect in the interaction term "PES= full-time employee" × "immigrant" (–46.9%), confirming a disadvantage for immigrants compared to Italians in university enrolment. Similar findings have been observed in other countries, such as the United States (Barsha et al., 2024), France (Ichou and Wallace, 2019), Spain (Pantzer et al., 2006), among others.

The age of the parents of immigrants was significantly lower than that of the parents of Italians, showing on average a difference equal to about 12 years. The parents' level of education had a significant impact on the probability of young people continuing in education. Analogous findings have been reported in Italy (Cantalini et al., 2020) and other countries (Wilder, 2013; Kantova, 2024). The employment status of immigrant parents was significantly lower than that of Italian parents. The same was true for disposable personal incomes and for the total income of families, some of which were well represented by a parabolic form in the model.

The empirical results are coherent with those reported in the literature and suggest that an "immigration" gradient is present in educational decisions also in Italy. Differences in educational enrolment/ attainment at the tertiary level among immigrants and Italians were explained by the socio-economic status of parents, i.e., their level of education, employment status, and occupational position. These results highlight the need for integrated policies in educational programmes, directed both at sustaining young people and helping their families, in order to stimulate and promote the enrolment of young immigrants in education programmes and to foster a complete integration process.

The outcomes obtained, when compared with those of the bibliographical references cited above, confirmed the findings of previous research while also presenting some novel insights. Two key results are highlighted here. First, the automatic selection of interactions provided interesting and interpretable outcomes, even if this strategy can lead to the selection of highly significant interaction terms, albeit difficult to interpret, and to the elimination of important variables that come into play indirectly through these interactions. This situation may be viewed in terms of the degree of urbanisation and the North-East and North-West microregions in the model in Table 4. Second, the positive impact of individual and family/ parental incomes (recorded to the highest level of

precision) on university enrolment was confirmed through a national survey and with a large sample of immigrants. In contrast, previous studies had largely identified this effect through municipal or district-level surveys, or indirectly, using provincial macroeconomic data.

The affordances provided by the two cross-sectional surveys, IT-SILC and IM-SILC (reference year 2009), such as a more consistent sample size of immigrants and a national perspective, also represent limitations of this study. First, the use of IT-SILC data from 2009, which is now admittedly dated, was necessary because that year marked the last survey of its kind on immigrants across Italy. Second, individual educational performance data are absent in surveys like IT-SILC and IM-SILC.

Finally, few models with interactions exist in the literature. In fact, in the applications, the interactions should be supported by social, behavioural, psychological, and economic theories. Otherwise, they may be obtained automatically simply by using an adaptive procedure like the Lasso method and only as empirical findings. The interactions are likely to be easily found among binary or categorical variables, but this case is relatively interesting because they can be replaced with specific typologies. The same holds true for the interactions of a continuous variable with other explanatory binary variables, but the interaction between two continuous variables is difficult to grasp immediately. In general, it is useful to find a theoretical justification for the existence of the interactions, instead of blindly searching for interaction terms. However, it is highly plausible that almost all phenomena are outcomes of interactions among many variables, but the explanation of these results is likely to be complicated and challenging.

ACKNOWLEDGEMENTS

This research was carried out as part of the "Determinant and economic effects of international migration" project of the Marco Biagi Department of Economics and the Marco Biagi Foundation, with financial support from the University of Modena and Reggio Emilia, UNIMORE FAR 2017 Grant, which is gratefully acknowledged. The present paper is a major revision of "Factors affecting tertiary education decisions of immigrants and non-immigrants in Italy", presented at the Scientific Conference on "Data-Driven Decision Making", ASA (Associazione per la Statistica Applicata) and DISPI (Dipartimento di Scienze Politiche e

Internazionali), University of Genoa (12-14 September 2022). In addition, the authors wish to thank William Bromwich for his painstaking attention to copy-editing this paper. Last but not least, this study is dedicated to the memory of Lorenzo Bernardi for his teachings, honesty, concern, and eagerness in stimulating all who worked with him and for his generous engagement with academic and other public institutions. He has left an indelible mark in our hearts.

REFERENCES

Algan, Y., Dustmann, C., Glitz, A. and Manning, A. (2010). The economic situation of first and second-generation immigrants in France, Germany and the United Kingdom. *The Economic Journal*. 15(2): 353-385.

Andersson, E.K., Lyngstad, T.H. and Sleutjes, B. (2018). Comparing patterns of segregation in North-Western Europe: A multiscalar approach. *European Journal of Population*. 34(2): 151-168.

Armstrong, M.J. and Biktimirov, E.N. (2013). To repeat or not to repeat a course. *Journal of Education for Business*. 88(6): 339–344.

Barsha, R.A.A., Najand, B., Zare, H. and Assari, S. (2024). Immigration, educational attainment, and subjective health in the United States. *Journal of Mental Health & Clinical Psychology*. 8(1): 16-25.

Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, New York.

Bertolini, P. and Lalla, M. (2012). Immigrant inclusion and prospects through schooling in Italy: An analysis of emerging regional patterns. In C. E. Kubrin, M. S. Zatz, and R. Martínez Jr., editors, *Punishing Immigrants: Policy, Politics and Injustice*. New York University Press, New York: 178-206.

Bertolini, P., Lalla, M. and Pagliacci, F. (2015). School enrollment of first- and second-generation immigrant students in Italy: A geographical analysis. *Papers in Regional Science*. 94(1): 141-160.

Bond Huie, S.A. and Frisbie, W.P. (2000). The component of density and the dimensions of residential segregation. *Population Research and Policy Review*. 19(6): 505-524.

Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*. 22(52): 781-861.

Bubritzki, S., van Tubergen, F., Weesie, J. and Smith, S. (2018). Ethnic composition of the school class and interethnic attitudes: a multi-group perspective. *Journal of Ethnic and Migration Studies*. 44(3): 482-502.

Buonomo, A., Strozza, S. and Gabrielli, G. (2018). Immigrant youths: Between early leaving and continue their studies. In B. Merrill, M. T. Padilla-Carmona, and J. González-Monteagudo, editors, *Higher Education, Employability and Transitions*

to the Labour Market. EMPLOY Project & University of Seville, Seville (ES): 131-147.

Cantalini, S., Guetto, R. and Panichella, N. (2020). Parental age at childbirth and children's educational outcomes: Evidence from upper-secondary schools in Italy. *Genus*. 76(8): 1-24.

Contini, D. (2013). Immigrant background peer effects in Italian schools. *Social Science Research*. 42(4): 1122-1142.

De Clercq, M., Galand, B. and Frenay, M. (2017). Transition from high school to university: A person-centered approach to academic achievement. *European Journal of Psychology of Education*. 32(1): 39-59.

Dustmann, C. (2008). Return migration, investment in children, and intergenerational mobility. *The Journal of Human Resources*. XLIII(2): 299-324.

Entwisle, D.R. and Alexander, K.L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*. 19: 401-423.

Eurostat (2009). *Description of Target Variables: Cross–Section and Longitudinal*. EU–SILC 065 (2009 operation). Directorate F, Unit F-3. Eurostat, Luxembourg.

Federman, M. and Levine, D. I. (2005). The effects of industrialization on education and youth labor in Indonesia. *Contributions to Macroeconomics*. 5(1), 1: 1-34.

Forster, A.G. and van de Werfhorst, H.G. (2020). Navigating institutions: Parents' knowledge of the educational system and students' success in education. *European Sociological Review*. 36(1): 48-64. https://doi.org/10.1093/esr/jcz049.

Friedman, J.H., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. 33(1): 1-22.

Glick, J.E. and Hohmann-Marriott, B. (2007). Academic performance of young children in immigrant families: The significance of race, ethnicity and national origins. *International Migration Review*. 41(2): 371-402.

Gould, W.W. (2000). sg124: Interpreting logistic regression in all its forms. *Stata Technical Bulletin*. 53: 19–29. Reprinted in *Stata Technical Bulletin Reprints*. Vol. 9: 257–270. Stata Press, College Station, TX.

Grove, W.A., Wasserman, T. and Grodner, A. (2006). Choosing a proxy for academic aptitude. *Journal of Economic Education*. 37(2): 131-147.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC, London.

Ichou, M. (2014). Who they were there: Immigrants' educational selectivity and their children's educational attainment. *European Sociological Review*. 30(6): 750-765. https://doi.org/10.1093/esr/jcu071.

Ichou, M. and Wallace, M. (2019). The Healthy Immigrant Effect: The role of educational selectivity in the good health of immigrants. *Demographic Research*. 40(4): 61-94.

Istat (2008). Ceccarelli C., Di Marco M., and Rinaldelli C., editors, *L'indagine europea sui redditi e le condizioni di vita delle famiglie* (Eu-Silc). Metodi e Norme n. 37. Istat, Rome.

Istat (2009a). Reddito e condizioni di vita delle famiglie con stranieri [electronic resource]. Rome: Istat. https://www.istat.it/it/archivio/52405. Last access: 03/01/2020.

Istat (2009b). Nota informativa sull'utilizzo dell'UDB (User Data Base) CVS 2009. Released by Istat together with data sets. Istat, Rome.

Kantova, K. (2024). Parental involvement and education outcomes of their children. *Applied Economics*. 56(48): 5683-5698.

Kleinbaum, D.G. (1994). *Logistic Regression: A Self-Learning Text*. Springer-Verlag, New York.

Krause, A., Rinne, U. and Schüller, S. (2015). Kick it like Özil? Decomposing the native-migrant education gap. *International Migration Review*. 49(3): 757-789.

Lalla, M. and Pirani, E. (2014). The secondary education choices of immigrants and non-immigrants in Italy. *Rivista Italiana di Economia Demografia e Statistica*. LXVIII(3-4): 39-46.

Lalla, M. and Frederic, P. (2020). Tertiary education decisions of immigrants and non-immigrants in Italy: An empirical approach. *DEMB Working Papers Series*. N. 168: 1-39. University of Modena and Reggio Emilia, Marco Biagi Department of Economics.

Le Brun, A., Helper, S. R. and Levine, D. I. (2011). The effect of industrialization on children's education. The experience of Mexico. *Review of Economics and Institutions*. 2(2): 1-34.

Luciano, A., Demartini, M. and Ricucci, R. (2009). L'istruzione dopo la scuola dell'obbligo. Quali percorsi per gli alunni stranieri? In G. Zincone, editor, *Immigrazione: segnali di integrazione. Sanità, scuola e casa*. il Mulino, Bologna: 113-156.

Luthra, R.R. and Flashman, J. (2017). Who benefits most from a university degree?: A cross-national comparison of selection and wage returns in the US, UK, and Germany. *Research in Higher Education*. 58(8): 843-878.

Minerva, T., De Santis, A., Bellini, C. and Sannicandro, K. (2022). A time series analysis of students enrolled in Italian universities from 2000 to 2021. *Italian Journal of Educational Research*. Anno XV(29): 09-22.

Montalbo, A. (2020). Industrial activities and primary schooling in early nineteenth-century France. *Cliometrica*. 14(2): 325-365.

Murat, M. (2012). Do immigrant students succeed? Evidence from Italy and France. *Global Economic Journal.* 12(3), Article 8: 1-22.

Murat, M. and Frederic P. (2015). Institutions, culture and background: The school performance of immigrant students. *Education Economics*. 23(5): 612-630.

Nguyen, A.N. and Taylor, J. (2003). Post-high school choices: New evidence from a multinomial logit model. *Journal of Population Economics*. 16(2): 287-306.

Ochsen, C. (2011). Recommendation, class repeating, and children's ability: German school tracking experiences. *Applied Economics*. 43(27): 4127-33.

Paba, S. and Bertozzi, R. (2017). What happens to students with a migrant background in the transition to higher education? Evidence from Italy. *Rassegna Italiana di Sociologia*. 58(2): 315-351.

Pantzer, K., Rajmil, L., Tebé, C., Codina, F., Serra-Sutton, V., Ferrer, M., Ravens-Sieberer, U., Simeoni, M.-C., Alonso, J. (2006). Health related quality of life in immigrants and native school aged adolescents in Spain. *Journal of Epidemiology and Community Health*. 60(8): 694-698.

Parker, J.D.A., Summerfeldt, L.J., Hogan, M.J. and Majeski, S.A. (2004). Emotional intelligence and academic success: Examining the transition from high school to university. *Personality and Individual Differences*. 36: 163-172.

Perreira, K.M., Harris, K.M. and Lee, D. (2006). Making it in America: High school completion by immigrant and native youth. *Demography*. 43(3): 511-536.

Pong, S. and Hao, L. (2007). Neighborhood and school factors in the school performance of immigrants' children. *International Migration Review*. 41(1): 206-241.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (Austria). URL http://www.R-project.org/. Last access: 18/04/2020.

Rondinelli, R., Policastro, V. and Scolorato, C. (2024). How student characteristics affect mobility choices at the university level: Insights from two surveys in Campania region. *Statistica Applicata – Italian Journal of Applied Statistics*. 36(1): 7-39.

Schnell, P. and Azzolini, D. (2015). The academic achievements of immigrant youths in new destination countries: Evidence from southern Europe. *Migration Studies*. 3(2): 217- 240.

Sleutjes, B., de Valk, H.A. G. and Ooijevaar, J. (2018). The measurement of ethnic segregation in the Netherlands: Differences between administrative and individualized neighbourhoods. *European Journal of Population*. 34(2): 195-224.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*. B. 58(1): 267-288.

UNESCO Institute for Statistics (2012). *International Standard Classification of Education ISCED 2011*, Ref. UIS/2012/INS/10/REV, UNESCO–UIS, Montreal (Quebec - Canada).

Usala, C., Porcu, M. and Sulis, I. (2023). The high school effect on students' mobility choices. *Statistical Methods & Applications*. 32(4): 1259-1293.

Van de Werfhorst, H.G. and Mijs, J.J.B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology* 36: 407-428.

Vettori, G., Vezzani, C., Pinto, G. and Bigozzi, L. (2020). The predictive role of prior achievements and conceptions of learning in university success: evidence from a retrospective longitudinal study in the Italian context. *Higher Education Research & Development*. 40(7): 1564–1577.

Vittorietti, M., Giambalvo, O., Genova, V. G. and Aiello, F. (2023). A new measure for the attitude to mobility of Italian students and graduates: A topological data analysis approach. *Statistical Methods & Applications*. 32(2): 509-543.

Wilder, S. (2013). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*. 66(3): 377–397.

Wilson, G. and Gillies, R. (2005). Stress associated with the transition from high school to university: The effect of social support and self-Efficacy. *Australian Journal of Guidance and Counselling*. 15(1): 77-92. doi:10.1375/ajgc.15.1.77

Wintre, M.G., Dilouya, B., Pancer, S.M., Pratt, M.W., Birnie–Lefcovitch, S., Polivy, J. and Adams, G. (2011). Academic achievement in first–year university: Who maintains their high school average? *Higher Education*. 62(4): 467-481.

Woodraw-Lafield, K.A. (2001). Implication of immigration for apportionment. *Population Research and Policy Review*. 20(4): 267-289.

Zou, H., Hastie, T. and Tibshirani, R. (2007). On the degrees of freedom of the Lasso. *The Annals of Statistics*. 35(5): 2173-2192.

Zwysen, W. and Longhi, S. (2018). Employment and earning differences in the early career of ethnic minority British graduates: The importance of university career, parental background and area characteristics. *Journal of Ethnic and Migration Studies*. 44(1): 154-172.

7. APPENDIX

The initial set of variables used to estimate the models and an example of models obtained by the introduction of explanatory variables blocks step by step are illustrated in this Section.

7.1 LIST OF VARIABLES

The data set contained many variables describing different aspects of each individual, as stated above. A factor analysis might have aggregated them into a reduced set. In general, there are difficulties in understanding and interpreting these factors. As a result, only the original variables indicated below were included in the models, sometimes with modifications and/or adaptations. For further details the reference provided in the introduction of Section 3 may be useful.

Qualitative variables are listed separately from quantitative variables.

Gender was dichotomised as 0 (men) and 1 (women) and termed women.

Citizenship was dichotomised as non-immigrant (0) and immigrant (1) and termed immigrants.

Self-perceived health (SPH), measured on a Likert scale (1=very good, 2=good, 3=fair, 4=bad, 5=very bad), was dichotomised assuming the value of 1 when SPH was problematic (6.4%), i.e., when the answer was fair or bad or very bad, and the value of 0 otherwise.

Suffering from any chronic illness or condition was equal to 1 for "Yes" and equal to 0 otherwise.

Limitation in activities because of health problems was dichotomised assuming the value of 1 when SPH was problematic (5.1% for severely limited or limited) and the value of 0 otherwise.

The "unmet need for medical treatment or examination" (5.0%) and the "unmet need for dental examination or treatment" (8.1%) were not included in the models, in order to reduce the number of explanatory variables, but also because this information is likely to be captured by the income of the family.

This block of variables was repeated for each young individual, their father, and their mother.

Education level of the father (ELF) and mother (ELM) were transformed into years and considered continuous variables. ELF and ELM were introduced into the models through a second-degree polynomial form: see below. Their modalities were the following: (0=ILL) illiterate, (0=NENI) no education and not illiterate, (1=PE) primary education, (2=LSE) lower secondary education, (2=VS3Y) vocational school of 2-3 years, (3=USE) upper secondary education, (4=PS-NTE) post-secondary non-tertiary education, (5=SCTE) short-cycle tertiary education, (6=BACH) bachelor's or equivalent level, (7=MAST) master's or equivalent level after bachelor, (8=PhD) research doctorate, doctor of philosophy or equivalent.

The characteristics concerning the labour market situation of the parents involved several categorical variables, which were combined between father and mother to reduce their numerosity and the results were transformed into binary variables.

The *parents' activity status* (PAS) reported the combination of the father's and mother's conditions: (1) PAS-FM equal to 1 when the father and mother were both employed and 0 otherwise, (2) PAS-F equal to 1 when only the father was

employed and 0 otherwise, (3) PAS-M equal to 1 when only the mother was employed and 0 otherwise, (4) PAS-R equal to 1 when at least one of the parents was retired and 0 otherwise, (5) PAS-O equal 1 when both parents were classifiable under "other conditions" and 0 otherwise.

The *skill level of parents* (SLP) in the job was determined retaining the maximum between the positions of the father and the mother: (1) SLP-ME if at least one of the parents was a manager or executive director and the other in a lower position, (2) SLP-EMPL if at least one of the parents was an employee and the other in a lower position, (3) SLP-LAB if at least one of the parents was a labourer and the other was unemployed, (4) SLP-OTHER was the residual category containing any other situations not included above.

The *parents' employment status* (PES) was not entirely reliable, but it was constructed combining the conditions of the father and those of the mother: (1) PES-FTD if only one or both parent were full-time salaried workers, (2) PES-FTSE if only one or both parents were full-time self-employed workers, (3) PES-PT if only one or both parents were part-time salaried or self-employed workers, (4) PES-MIX if only one or both parents were full-/part-time salaried or self-employed workers but different from the previous modalities, (5) PES-PENS if at least one of the parents was retired and the other was employed part-time, unemployed or out of labour force, and (6) PES-UOLF if at least one of the parents was unemployed or out of the labour force. Note that PAS-R and PES-PENS coincide.

The *type of contract permanent* (PRM) was a binary variable equal to 1 when the father (PRM-F) or the mother (PRM-M) had a job/work contract of unlimited duration. The *type of contract temporary* (TMP) was a binary variable equal to 1 when the father (TMP-F) or the mother (TMP-M) had a job/work contract of limited duration.

The base/ reference category for these three variables is made up of those who are not in the labour market and, hence, the dichotomous variables obtained can all enter the model.

The *tenure status of the household* (TSH) presented four modalities: (1) tenant, (2) subtenant, (3) owner, (4) free accommodation.

Other five binary variables concerned household: the amount of rent was substantial, the amount of loan/mortgage was substantial, repayment of loans to banks, there was a saving in 2008, there was a reduction of disposable income for needs.

Three ordinal/counting variables summarised certain types of information: house-evaluation, optional, and needs. The *house-evaluation* (range 0-9) counted if the dwelling had a habitable kitchen, indoor flushing toilet, cellar and/or attic, terrace and/or balcony, garden, hot water, garage, roofs or ceilings or doors or floors damaged, moisture in the walls or ceilings or floors or foundations. The *optional* (range 0-7) counted if the family had a telephone (fixed landline or mobile), a dishwasher, a fridge, a VCR-DVD player, a camera, a satellite dish/antenna, and internet access. The *needs* (range 0-7) related to the lack of money in the family for necessary food, for necessary clothes, for illness to be treated, for school, for transport, for the payment of taxes, and added the request for help to purchase essential goods.

The local and geographical variables were limited to two variables.

The *macro-region* (MR) subdivision of Italy was provided by Istat. The North-West (NW) included Valle d'Aosta, Piedmont, Liguria, and Lombardy. The North-East (NE) included Trentino Alto Adige, Veneto, Friuli Venezia Giulia, and Emilia-Romagna. The Centre (C) included Tuscany, Umbria, Marche, and Latium: it was chosen as base/reference category. The South (S) included Abruzzo, Molise, Campania, Basilicata, Apulia, and Calabria. The Islands (I) included Sicily and Sardinia.

The *degree of urbanisation* (DOU) provided three modalities: densely-, moderately-, and thinly-populated area. The latter was chosen as base/reference category.

For the sake of brevity, the description of some other qualitative variable is omitted and some of the previous variables are illustrated in Figure 3.

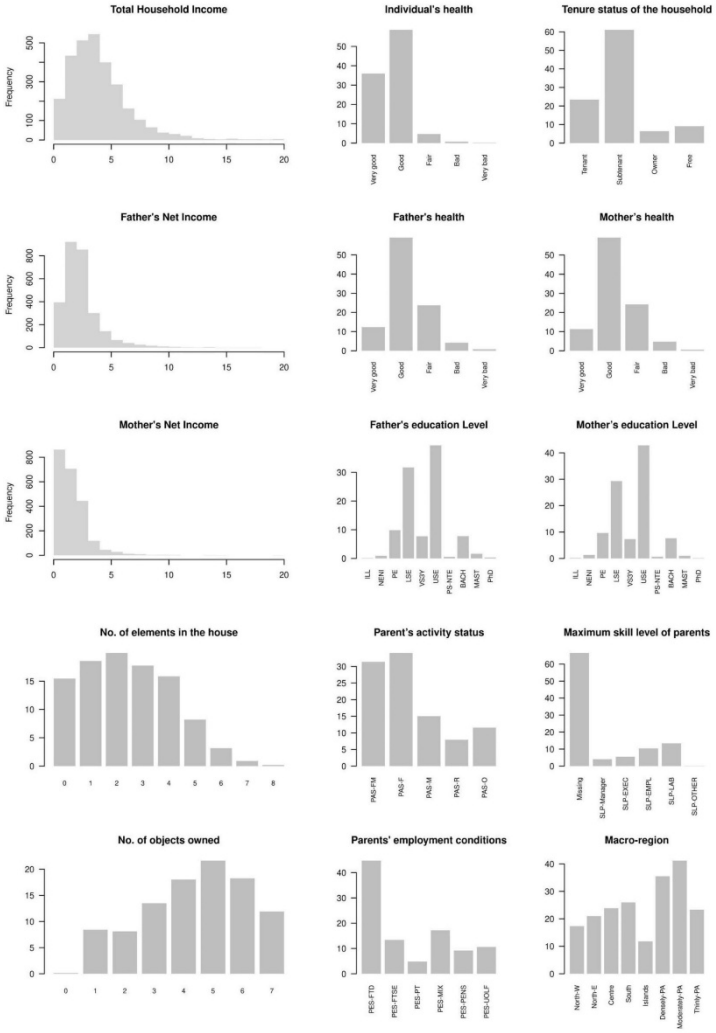

Figure 3: Descriptive plots of some qualitative and quantitative variables

Continuous variables were almost always introduced into the model through a second-degree polynomial form to capture some nonlinearities in the behaviours of individuals who took on different values in them. Only the list of these variables is reported here, for the purpose of brevity.

Age concerned the young individuals, the fathers, and the mothers: the ages were divided by 10 to have a range of values comparable with the binary variables. For the *education of parents* see above.

Income concerned many variables and components, which were all divided by 10,000. The net *disposable personal income* (DPI), as for age, concerned the young individuals, the father (FDPI), and the mother (MDPI). The net *disposable family income* (DFI) was available and *family income per capita* (FIPC) was calculated using the *number of family members*. The income variables were mutually correlated, and the correlation coefficients differed significantly from zero, but the values were surprisingly low, except for the coefficient between the total income of the family and the father's income ($r=0.786$, $p<0.000$). However, DPI should be used in the model with caution because its value was zero in the case of 1122 individuals (39.0%) and 723 of the latter (64.4%) had achieved or were currently attending tertiary education. Only 25 individuals (0.9%) reported negative income. Some continuous variables are illustrated in Figure 3.

7.2 LOGISTIC REGRESSION MODELS WITH VARIABLES BLOCKS

The original data set contained many variables describing different aspects of each individual, as stated above. A factor analysis might have aggregated them into a reduced set. However, in general, there are difficulties in understanding and interpreting these factors. As a result, some variables were omitted to reduce their number and only the original variables, most of them described above, were included in the models, sometimes with modifications and/or adaptation.

Table 7 shows the odds ratios for five different models, each one obtained adding a block of variables representing a dimension or a specific situation. Column (1) shows only the estimated odds ratios, without the standard errors for shortness, of the first block of variables referring to the young individuals (*Model 1*) constituting the sample cases. Furthermore, it also shows the estimated odds ratios of the models containing only the single next added block. The remaining fours columns concern respectively the addition of the father's data block (*Model 2*), then the addition of the mother's data block (*Model 3*), then the tenure status of the household data block (*Model 4*), and finally the addition of the block containing the Macro-Region (MR) and the degree of urbanisation.

Table 7: Logistic regressions with marginal effects and estimated odds ratio for different models: (1) single block of variables (M2) blocks 1+2, (M3) blocks 1+2+3, (M4) blocks 1+2+3+4, (M5) blocks 1+2+3+4+5

Regressor/ Block	(1)	(M2)	(M3)	(M4)	(M5)
1. YOUNG INDIV.					
Intercept	2.746***	0.246#	0.010***	0.009***	0.008***
Woman	1.320***	1.373***	1.607***	1.648***	1.651***
$(Age/10)^2$	0.906#	0.925	0.939	0.950	0.938
Immigrant	0.356***	0.507***	0.632***	0.909	0.946
DPI	0.177***	0.183***	0.182***	0.181***	0.186***
$DPI2 = DPI^2$	1.231***	1.227***	1.227***	1.225***	1.219***
SPH: Self-Perc Health	0.395***	0.440***	0.454***	0.447***	0.446***
SPH: chronic illness	1.743*	1.857**	1.907**	1.929**	1.940**
SPH: limitat. activ.	1.069	0.924	0.864	0.874	0.880
2. FATHER BLOCK					
Intercept	0.007***				
[(Father's age)/10]	3.489***	1.388	1.315	1.206	1.145
$[(Father's age)/10]^2$	0.919**	0.997	0.982	0.985	0.990
ELF (Educ. Level)	0.973	0.945	0.925	0.910	0.901
ELF^2	1.006*	1.007**	1.006#	1.007*	1.007*
FDPI= (Father's DPI)	1.109*	1.164***	1.070	1.022	1.026
$FDPI^2$	0.998	0.996	0.999	1.000	1.000
SPHF: SPH of father	0.793*	0.794#	0.814	0.838	0.868
SPHF: chronic illness	1.081	1.023	1.006	0.989	0.961
SPHF: limitat. activ.	1.168	1.191	1.259	1.315#	1.327#
PRM-F: permanent	1.081	0.959	0.953	0.934	0.950
TMP-F: temporary	0.830	0.878	0.932	0.990	1.042
No. of observations	2874	2874			
Log Likelihood	–1674.66+	–1577.61			
Akaike Inf Criterion	3367.31+	3195.22			
Bayesian Inf Criterion	3420.98+	3314.49			

Notes: + First block only. # p<0.1; * p<0.05; ** p<0.01; *** p<0.001 (*continue*)

Table 7 (continued from previous page): Logistic regressions with marginal effects and estimated odds ratio for different models: (1) single block of variables (M2) blocks 1+2, (M3) blocks 1+2+3, (M4) blocks 1+2+3+4, (M5) blocks 1+2+3+4+5

Regressor/ Block	(1)	(M2)	(M3)	(M4)	(M5)
3. MOTHER BLOCK					
Intercept	0.002***				
(Mother's age)/10	3.970***		2.366***	2.303***	
[(Mother's age)/10]^2	0.915***		0.954#	0.955#	
ELM (Educ. Level)	1.074		1.126	1.107	
ELM^2	1.002		0.999	0.999	
MDPI= (Moth. DPI)	1.081#		1.160**	1.145**	
MDPI^2	0.998		0.993**	0.994*	
SPHM: SPH of mother	0.844		0.938	0.962	
SPHM: chronic illness	1.115		1.084	1.104	
SPHM: limitat. activ.	1.090		0.930	0.970	
PRM-M: permanent	1.363***		1.138	1.135	
TMP-M: temporary	1.115		0.997	1.048	
4. TENURE STATUS					
Intercept	0.211***				
TSH: Subtenant	1.473**			1.091	
TSH: Owner	1.155			1.009	
TSH: Free	1.061			1.105	
Rent is substantial	0.957			0.963	
Loan is substantial	0.745*			0.751#	
Repayment to bank	0.757**			0.741**	
House-evaluation	0.975			0.976	
No. of optional	1.325***			1.183***	
No. of observations	2874	2874	2874	2874	
Log Likelihood	–1674.66$^+$	–1577.61	–1520.33	–1499.51	
Akaike Inf Criterion	3367.31$^+$	3195.22	3102.65	3077.02	
Bayesian Inf Criterion	3420.98$^+$	3314.49	3287.52	3309.59	

Notes: $^+$ First block only. # p<0.1; * p<0.05; ** p<0.01; *** p<0.001 (*continue*)

Table 7 (continued from previous page): Logistic regressions with marginal effects and estimated odds ratio for different models: (1) single block of variables (M2) blocks 1+2, (M3) blocks 1+2+3, (M4) blocks 1+2+3+4, (M5) blocks 1+2+3+4+5

Regressor/ Block	(1)	(M2)	(M3)	(M4)	(M5)
5. MR AND DOU					
Intercept	0.629***				
North-West	0.847				1.044
North-Est	0.808#				1.081
South	1.286*				1.349*
Islands	0.836				0.923
Densely-pop area	1.622***				1.352*
Moderately-pop area	1.293**				1.354*
No. of observations	2874	2874	2874	2874	2874
Log Likelihood	–1674.66+	–1577.61	–1520.33	–1499.51	-1491.71
Akaike Inf Criterion	3367.31+	3195.22	3102.65	3077.02	3073.42
Bayesian Inf Criterion	3420.98+	3314.49	3287.52	3309.59	3341.78

Notes: + First block only. # p<0.1; * p<0.05; ** p<0.01; *** p<0.001

The blocks concerning the working condition of parents (13 binary variables) and the family data (five binary variables and 10 continuous variables) were not shown to reduce the length of the Table. According to the Bayesian information criterion (BIC), the *Model 3*, including the father's and mother's data, showed the best fitting model in Table 7. Overall, *Model 3* turned out to be the best model (with the lowest BIC) also adding the other two omitted blocks. According to the Akaike information criterion (AIC) the best model resulted the *Model 6*, not reported here, given by Model 5 plus the block of working conditions of parents.