Three Essays on Economic Development

A dissertation presented by Paula Luciana Méndez Errico to
The Department of Applied Economics
in partial fullfiment of the requirements for the degree of
Doctor of Philosophy in the subject of Economics
Universidad Autónoma de Barcelona

Director
Xavier Ramos Morilla

Director
Roxana Gutiérrez Romero

October, 2014

Aknowledgements

This thesis would not be possible without the immense support received during these years. First, I want to thank to my advisors Xavier Ramos and Roxana Gutiérrez, who kindly accepted me as a PhD student, and for sharing with me their knowledge and experience, for their valuable suggestions and encouragement during this research.

In a personal dimension, my profound gratitude goes to Matías, with whom we shared this adventure and for always encouraging me during this period. Also, to my parents for their unconditional love, and for taking care of all those issues I could not afford because of the distance. This thesis in part reflects their transmitted sense of what should be fair and their concern on equity. I also want to mention to my brothers Sebastián and Federico for their love and support, and to Fiorella for her great advices. To my grandmother and my niece Sofia, my motives for always want to come back home. To my grandfather, Negra and Beba for taking care of me wherever they are. Also, I am extremely grateful to Gabriela and Luis for being unconditionally close to me.

From this side of the ocean, to my cousin Damiana and my niece Olivia, my family and friends in Barcelona, my strongest support in good and not so good times. To my family in Italy, who warmly received me any time I visited them making me feeling at home.
During these six years I met great friends who became my 'chosen' family in Barcelona. Special thanks go to Roberto and Ana, Dolores, Paula, Adriana, Lina, Cristina, Jorge, Orlando and Areli for all the great moments shared. Also, I want to specially thank to Mónica, with whom we share not only similar research interests and rich discussions, but also a great friendship. I am also grateful to Camilo for teaching me facing bad times with a great sense of humor. To Mamen and Marina for making me feeling at home, and to Fausto for the saturday's vermouths in Gràcia. To Macarena, Natalia, Victoria, Vicky Prieto, Paula, Gonzalo and Martín, my Uruguayan friends in Barcelona, for the mates and with whom we shared our homesickness. I would also thank to Roger, Miquel and Sonia for sharing their culture and uses with me. To Rodrigo and Nery, for the lovely times we spent together in Barcelona and for kindly receiving me in Mexico.

I also would like to thank to Pilar, Monste, Imma and Miquel from the secretary of the Applied Economics Department of the UAB who always helped me with the bureaucracy. Special thanks go to Rosella Nicollini, for always taking care of the PhD students at the Department. Also, to Cristina López Mayan who was my evaluator during my PhD, for being always available to evacuate my doubts, for her insightful comments and advices during all this period.

Finally, special thanks go to the Spanish Ministry of Science and Innovation (MICINN) with reference project ECO2010-21668-C03-02 and reference scholarship BES-2011-048083 for the financial support. Specific acknowledgements are shown at the end of each essay.

Index

Introduction

Over the last decades there has been a resurgence of interest in economic research on economic development. The great income differences observed between as well as within countries has turned economic science's attention to explain why countries differ in their economic growth, and why within countries some people may be entrapped in poverty.

The recognition that income inequality and economic status perpetuates over time not only in poorer countries but also in wealthier societies, and the associated costs in different aspects of individual's and social well-being, such as happiness, health, education, crime, violence, corruption, among others (Wilkinson and Picket, 2011), lead to the development of new and insightful theories in economics. This literature on the effect of income inequality and economic growth suggests alternative mechanisms that could cause poverty to persist, addressing both the question of how whole economies may fail to develop, and how population subgroups within rich economies may fail to share in overall prosperity.

I broadly identify three set of theories that explain dispersion in income across individuals and social groups and divergence in economic growth across countries; such as those based on (i) individual characteristics, (ii) institutional factors, and (iii) social interactions. Although individual, social interactions and institutional factors are interdependent, alternative explanations of poverty have different implications, both in terms of understanding the sources of poverty and inequality as well as in terms of the design of public policies (Durlauf, 2006).

The main objective of this dissertation is to study some of the mechanisms suggested by the literature as factors that could prevent individuals from attaining certain domains of well-being.

Specifically, this thesis is divided in three independent essays providing new evidence on three issues within the field of economic development: the effect of social networks on immigrants' labor market outcomes (first essay), the long-lasting impact of income inequality on entrepreneurial success and job creation (second essay), and the importance of multiple abilities, parental educational background and race in explaining educational gaps (third essay). Also, different cases of study are provided: immigration issues in a developed country such as Spain, initial conditions for a broad set of

countries with different levels of economic development, and education in a middle-income country such as Uruguay. Finally, different databases and econometric techniques are properly selected to address each case of study. I explain in further detail the goal and findings of these three essays next.

The **first essay** "*The impact of social networks on immigrants' employment prospects: the Spanish case 1997-2007"* analyzes the factors that could prevent or foster immigrants' social and economic integration in the host country. Specifically, this essay contributes to the empirical literature on immigration and social networks by studying the extent to which social networks affect labor market outcomes -job match and wages- for immigrants living in Spain. To this end, I first study the impact of social networks on the job matching process by studying the probability of keeping the first job in Spain relative to not keeping it; namely, changing jobs, being unemployed or inactive. Secondly, for those immigrants actually employed in the same job since arrival, we analyze the effect of social networks on wage.

Labor market participation and conditions in terms of employment and wage, is one of the main immigrant's integration channel to the host country, and also an important source of immigrant's income. In turn, social networks have been recognized in the literature as an important channel through which information is transmitted, especially relevant for immigrants in the host country as it provides -among others- information on labor market institutions and job opportunities (Calvó-Armengol and Jackson, 2004 and 2005). But also, social networks could prevent immigrants' to integrate in the host country, since widespread reliance on social networks in the labor market can lead to social stratification by limiting individuals' opportunities to those that their peer group can provide (Mouw, 2009). The persistent segregation of immigrants in the labor market may affect future prospects of their offspring, leading to the extreme case of economic immobility in which immigrants are entrapped into poverty.

Despite the growing literature on social networks and immigrants' labor market outcomes, no conclusive effects of social networks on immigrants' workers have been found yet (Ioannides and Loury, 2004). By focusing on the effects of social networks on immigrants' labor market outcomes, this study contributes to the empirical literature by addressing a less explored channel through which immigrants' social and economic integration could be affected.

To empirically analyze the effect of social networks on job match and wages, I use data from the National Immigrant Survey conducted in 2007. In this study two measures of social networks are considered: the strength of the network (close and weak ties); and the size of the network proxied by the proportion of immigrants from the same country of origin living in the same region (Autonomous Community) on the total immigrant population in the region of destination. It is also considered the alternative mechanisms of job access: relatives or friends (network jobs) and formal methods (such as public and private employment agencies, newspaper advertisements, among others).

Endogeneity issues are likely to emerge in this study, because a selection process of immigrants in labor market statuses may take place, and because social network formation is likely to take place among individuals with particular traits. Therefore, a two-step procedure is applied, first for analyzing job match, and then for wage quantile regression estimations.

Also, as individuals are more likely to socially interact if they share some individual traits as being sociable and responsible, education or occupation, an extensive set of exogenous variables like occupation and sector of activity in the country of origin is included.

The findings suggest that social networks are likely to help immigrants to find a job in the short-run, but may limit opportunities to fully integrate in the longer term. In this sense, these findings shed light on the importance of social networks preventing immigrants' integration, as well as help to orientate the design of integration policies for immigrants living in Spain.

The **second essay**, entitled "*The Long-Term Effect of Inequality on Entrepreneurship and Job Creation"* studies the extent to which initial conditions understood as income inequality in 1700s and 1800s, and credit market institutions, can condition entrepreneurship and job creation to flourish over time.

This essay adds to the literature on the long-lasting effects of income inequality on economic development by empirically testing the predictions of the model by Banerjee and Newman (1993). This model predicts that initial conditions understood as historical distribution of wealth, can have a long and persistent effect on development. Specifically the model assumes that people can become either entrepreneurs or workers. Since entrepreneurial activities require an up-front investment, they are available only to wealthy people and to those individuals who can provide collateral to access credit.

Poorer and credit constrained individuals can only choose to work for a wage or to be self-employed. Then, occupational choice will in turn give rise to a new distribution of income by determining the returns and allocation of occupations, affecting the process of economic development through, for instance, its effects on saving, investment, risk bearing, and the composition of demand and production. Therefore, countries with initially low income inequality would grow over time aided by a strong entrepreneurial sector. A contrasting equilibrium could be reached if a country starts with a high ratio of poor to wealthy people. In this case development runs out of steam.

Two hypotheses are derived from the model: 1) countries that have a historical high ratio of wealthy to poor people have a lower probability of firms being created, surviving, and of these creating jobs over time, and 2) countries that currently have more efficient credit markets have a higher probability of people being involved in entrepreneurship and of higher job creation.

To test the predictions of this model, a pseudo-panel of entrepreneurs across 48 countries over 2001-2009 is built using the Global Entrepreneurship Survey, and is complemented with historical indicators of income distribution prevailing in 1700 and 1800 and current business environment, conditions that can affect the probability of firms being created, surviving and creating jobs over time.

The methodology combines pseudo-panel techniques with instrumental variables, given that current business environment could be affected by the proportion of people involved in entrepreneurial activities, for instance by lobbying for certain laws.

The findings of this essay give empirical support to the predictions of the model, showing that historical income inequality and current credit market imperfections prevent firms to be created and surviving over time, at the time that affect job creation over time.

To the best of our knowledge, this article is the first one that tests the long-term effects of inequality on occupational choice, thus giving empirical evidence on a less studied channel through which income inequality can affect long-term development.

The **third essay**, entitled *"Schooling progression in Uruguay: why some children are left behind?"* studies the impact of parental traits on children's educational attainment in Uruguay. Specifically, I analyze whether long-term parental background, crystallized by parental educational background, race, cognitive and non-cognitive abilities, and

short-term family income measured by the opportunity cost of education, affect child' schooling progression, and at what stage of the educational path they take on their importance.

This study is motivated by the recent literature stressing the effects of multiple abilities on persistent economic status and education inequality developed by Bowles and Gintis (2001, 2002) and by Heckman and co-authors (Heckman et al., 2011; Heckman and Mosso, 2014; Heckman et al., 2006). In addition, the scarcity of this type of analysis found for less developed countries and the particularities of the Uruguayan educational system encourages choosing Uruguay as an interesting case of study.

The empirical methodology considers a sequential probability model proposed by Cameron and Heckman (1998, 2001), in which education attainment is the outcome of the individual's previous schooling decisions. Two main advantages are found in this methodology. First, it recognizes the selection taking place across schooling, in which more able and motivated individuals and those with better parental educational backgrounds are more likely to attain higher levels of education. Second, it allows identifying a direct effect of the key variables of the study on each schooling stage, and also an indirect effect of these variables by affecting previous schooling decisions. This analysis requires valid exclusion restrictions, thus I considered labor market conditions at the time schooling decisions are made.

The dataset used in this study is the National Youth Survey which enables me to construct individual's educational path and performance, and to exploit information on motivation and risky behavior to proxy socio-emotional endowments, as recognized by earlier studies (Gullone and Moore, 2000; Heckman et al.,2006; Heckman e al., 2014).

The results show that parental educational background, cognitive and non-cognitive abilities have effects of diverse magnitude across stages of the educational path. Long-term parental background has increasing effect over the children's schooling progression in comparison to short-term parental income as it decreases its significance when students progress to higher schooling stages. Specifically, cognitive ability has increasing effects on the students' likelihood of dropping out across the educational path. Motivation and risky behavior measuring non-cognitive ability also influence children's schooling completion at early stages of education. This article finds that despite the great supply of public education, children are being left out. The reasons, we found, are initial conditions, understood as family background. Thus, with important policy recommendations.

References

Banerjee, A., and A.F. Newman (1993). "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101 (2): 363-394.

Bowles, S. and Gintis, H. (2001) "Schooling in Capitalist America Revisited", *Sociology of Education* 75(1):1-18.

Bowles, S. and Gintis, H. (2002) "The Inheritance of Inequality", *Journal of Economic Perspectives* 16 (3):3-30.

Calvó-Armengol, A., and Jackson. M. (2004). "The effects of social networks on employment and inequality", *American Economic Review* 94(3): 426-454.

Calvó-Armengol, A., and Jackson. M. (2005). "Job matching and word-of-mouth communication", *Journal of Urban Economics* 57: 500-522.

Cameron, S., and Heckman, J. (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy* 106 (2):262-333.

Cameron, S., and Heckman, J. (2001) "The dynamics of educational attainment for black, Hyspanic and white males", *Journal of Political Economy* 109 (3), 455-99. *American Economic Review*, 92(4): 727–744.

Durlauf, S. (2006) "Groups, social influences and inequality", (in) *"Poverty Traps",* (ed) Bowles, S., Durlauf, S., and Hoff, K., Princeton University Press

Gullone, E., and Moore, S. (2000) "Adolescent risky-taking and the five-factor model of personality", *Journal of Adolescence* 23:393-407.

Heckman, J. (1979). "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1): 153-161.

Heckman, J., Humphries, J., Veramendi, G; and Urzúa, S. (2011) "The Effects of Educational Choices on Labor Market, Health and Social Outcomes", University of Chicago WP No. 2011-002.

Heckman, J., Humphries, J., Veramendi, G; and Urzúa, S. (2014) "Education, Health and Wages", IZA DP No. 8027.

Heckman, J., and Mosso, S. (2014) "The Economics of Human Development and Social Mobility", IZA DP No. 8000.

Heckman, J.; Stixrud, J.; and Urzúa, S. (2006) "The Effects of Cognitive and Noncognitive abilities on Labor Market Outcomes and Social Behaviour", NBER WP No. 12006.

Ioannides, Y., and Loury, L. (2004). “Job Information Networks, Neighborhood Effects and Inequality”, *Journal of Economic Literature*, 42(4): 1056-1093.

Mouw, T. (2009). “*The Use of Social Networks Among Hispanic Workers: An Indirect Test of the Effects of Social Capital*”, University of North Carolina Press, Chapel Hill.

Wilkinson, R., and Pickett, K. (2011) “*The Spirit level: why greater equality makes society stronger”*, Bloomsbury Press.

Essay 1
The impact of social networks on immigrants' employment prospects: the Spanish case 1997-2007

The Impact of Social Networks on Immigrants' Employment Prospects: The Spanish Case 1997-2007*

Abstract

This paper studies the extent to which social networks influence the employment stability and wages of immigrants in Spain. By doing so, we consider an aspect that has not been previously addressed in the empirical literature, namely the connection between immigrants' social networks and labor market outcomes in Spain. For this purpose, we use micro-data from the National Immigrant Survey carried out in 2007. The analysis is conducted in two stages. First, the impact of social networks on the probability of keeping the first job obtained in Spain is studied through a multinomial logit regression. Second, quantile regressions are used to estimate a wage equation. The empirical results suggest that once the endogeneity problem has been accounted for, immigrants' social networks influence their labor market outcomes. On arrival, immigrants experience a mismatch in the labor market. In addition, different effects of social networks on wages by gender and wage distribution are found.

* This essay has been co-written with Xavier Ramos (Departament d'Economia Aplicada – Universitat Autònima de Barcelona).

1.1 Introduction

The immigrant population in Spain has largely increased over the past decade, from 2.3% of the total population in 2000 to 10% in 2007. This large immigration inflow has turned Spain into the second largest recipient of immigrants after Germany in the European context (OECD, 2010). The social relevance of this new phenomenon has turned the immigration process into a key subject of social and economic research. Different studies have focused on the assimilation process and occupational mobility of immigrants in Spain (Izquierdo et al., 2009; Alcobendas and Rodríguez-Planas, 2009, Simón et al., 2011; among others). However, less attention has been paid to the role of social networks on immigrants' labor market outcomes.

Empirical and theoretical studies point out the influence of social networks in various areas of social and individual behavior, such as labor market performance, education attainment, and crime among others (Jackson, 2008; Wahba and Zenou, 2005). For immigrant workers, social networks may accelerate the job finding process. For instance, employers within an enclave may prefer to hire workers from their own country (Borjas, 2000). However, belonging to an enclave, may in turn affect the quality of the job offers an immigrant receive, as it influences the speed at which the immigrant learn the skills of the host country (such as language). Therefore, strong dependence on the social network may isolate immigrants from the native population and from the organizations and institutions in the host country. In the long run, immigrants' enclaves may develop, reflecting social and economic disintegration.

In this paper, the focus is on the effects of social networks on the job quality an immigrant finds, mainly because social and economic integration largely depends on an immigrant's labor market outcomes. The objective of this paper is to analyze to what extent social networks affect immigrants' labor market outcomes in terms of employment stability and wages in Spain.

Theoretical literature agrees on the positive impacts of strong and weak ties on the rate at which jobseekers receive employment offers.[1] Moreover, the quality of the members of the network influence the quality of the job an individual can find (Calvó-Armengol and Jackson, 2004). Several empirical studies show that individuals' probability to find a job increases with the individual social networks. For instance,

[1] Close or strong ties refer to the strength of the network. Close ties include family and friends, while weak ties are expressed in terms of a lack of overlapping in personal networks between any two agents (e.g. professional acquaintances).

Munshi (2003) finds that Mexican migrants in the U.S who obtained a job through social networks improve their labor market outcomes. Wahba and Zenou (2005) show that, conditional on being employed, individuals' probability to find a job through social networks relative to formal search mechanisms, increases and it is concave with the size of the networks. In addition, they stress that this effect is bigger for the less educated workers. Patacchini and Zenou (2008) find that individuals' probability of being employed increases with the size of close and weak ties.

However, despite the growing empirical literature, no consensus of the impacts of social networks on job quality has yet been reached (Ioannides and Loury, 2004). Dustmann et al. (2010) show that through referrals, social networks reduce informational deficiencies in the labor market, leading to better quality matches between workers and firms. Some authors argue that immigrants with social resources obtain more advantageous occupational positions, as friends and relatives sort through jobs to reserve the better ones for their network's members (Aguilera and Massey, 2003; Nee and Sanders, 2001). Conversely, Bentolila et al. (2010) find that worker/job matches tend to be poorer for jobs found through the network. In a similar line, Ottaviano and Peri (2006) point out that job matches depend on the strength of the network. They argue that mismatch happens if social networks are based on close ties because relatives and friends are unrelated to the individual's previous experience or training. Instead, good matches can happen if job information is transmitted through professional affiliations.

This paper aims to contribute to the empirical literature on the impact of social networks on job quality, through studying the relationship between social networks and job match on one hand, and the effects of social networks on wages on the other. Little is known about the mechanisms through which social networks affect immigrants' labor market outcomes in Spain. We intend to provide empirical evidence of the mechanisms through which social networks affect immigrants' employment outcomes and thus, contribute to the vast empirical literature on the assimilation process of immigrants in Spain. Unlike previous studies, in this paper the focus is on the role of social networks on immigrants' employment outcomes, an issue not addressed before for the Spanish case.

In contrast to other studies, we do not rely on the identification assumption that individuals within a given group (such as ethnic group, neighborhood or firm) actually know each other and are members of the same network. Most empirical studies of the

effect of social networks on immigrants' labor market outcomes focus on indirect measures of social interactions such as the number of other immigrant's own country (Munshi,2003); geographical proximity or group affiliation (e.g. Topa, 2001;Weinberg et al., 2004; Bayer et al., 2008; Dustmann et al., 2010). The dataset used in this study, the National Immigrant Survey (ENI, its Spanish acronym) allows us to use direct information on social interactions provided by the immigrant such as having relatives and friends on arrival to Spain, social participation in organizations and the job access mechanisms used to obtain the first job in Spain.[2] In addition, the richness of the ENI, with retrospective information on individuals' labor market characteristics and histories, enables us to address the potential unobserved endogeneity problem controlling for labor status and last occupation in the country of origin.

First, we study the impact of social networks on the job matching process through studying the probability of keeping the first job relative to not keeping it; namely, changing jobs, being unemployed or inactive. As the individuals considered in this analysis are those with some labor experience in Spain, we estimate the multinomial regression controlling for sample selection. Then, the effects of social networks on wages are estimated for immigrants who keep their first jobs. We estimate a wage equation, separately for women and men, through ordinary least squares (OLS) and quantile regressions (QRs). We exploit a novel methodology for the study of social network effects on wages through QRs controlling for sample selection bias. These effects are estimated in a semi-parametric fashion using a two-step procedure similar to that suggested by Heckman (1979).

Our results show that social networks have significant effects on the job matching process for immigrant workers and wages. A job mismatch is observed for immigrants upon arrival, they prefer to quickly accept a job offered through the social network, even if it is not the most suitable given their human capital endowments. In addition, we find positive effects of network size on job match, possibly reflecting the existence of ethnic niches in the labor market. Finally, social networks differently impact the wage distribution for women and men. The strength of the network (close or weak ties) only affects men' wages but does not affect women's wage. Wage penalties

[2] Cappelari and Tatsiramos (2010) and and Goel and Lang (2011 and 2012) also uses direct information on social interactions in their studies of the effect of social networks on employment outcomes. Cappellari and Tatsiramos (2010) construct a measure of the quality of the worker network based on each respondent's three best friends and their characteristics using the British Household Panel Survey. Goel and Lang (2011 and 2012) use immigrants' contacts at arrival obtained from the Longitudinal Survey of Immigrants to Canada (LSIC).

are observed for both women and men who obtained the job through social networks. This effect varies across the wage distribution between women and men. The network size also penalizes both women's and men's wages.

The remainder of this paper is organized as follows. The next section describes the data and provides summary statistics for the key variables of interest. Section 3 introduces the empirical strategy. Section 4 presents the results of the analysis. Finally, the last section concludes.

1.2 Data and descriptive analysis

This study uses the National Immigrants' Survey (ENI, its Spanish acronym), a single and unique cross-sectional national representative survey on immigration conducted so far only for 2007 by the National Statistics Institute (*Instituto Nacional de Estadística*). The sample is based on the Municipal Register. In total, the original survey comprises 15,441 individuals.[3] The ENI offers information on socio-demographic characteristics, migration experience, social networks, and labor market experience.[4] In particular, it features detailed information on activity condition before migration and at the moment of the survey and retrospective information on employment (e.g. occupation and activity sector) at three times: in the country of origin, first job on arrival, and current job in Spain. In addition, information on the finding methods used for the first job (social networks or formal methods) and personal income (net amount of money employees receive), among others, is provided.

The original sample is restricted to immigrants that arrived in Spain after 1996. This constraint prevents selection bias in the analysis for different reasons. As Borjas (1985, 1995) states, cross-sectional estimates of immigrant performance in the host country could induce selection bias due to "cohort effects", namely changes in the composition or "quality" of immigrants arriving at different points in time or because of nonrandom return migration or migration to a third country. In addition, the business cycle could affect the results of labor market entrants and bias the estimation (Aslünd

[3] A response rate with respect to the effective sample eligible respondents of 87.4% was obtained. Interviews were conducted face-to-face, and for those informants unable to fill out the questionnaire in Spanish, a telephone line was set up (in Arabic and English).

[4] More detailed information on the design and contents of the ENI can be found at http://www.ine.es/daco/daco42/inmigrantes/inmigra_meto.pdf.

and Rooth, 2007).[5] Considering the period between 1997 and 2007 minimizes these effects. Simón et al. (2011) also stress that during this period immigrant flows into Spain were relatively homogeneous in relation to their regions of origin. Further, the authors point out that the economic growth and strong job creation observed in this period reduce the effects of the economic cycle on immigrants' labor market situations and the importance of return migration relative to economic downturns.

This analysis considers immigrants between 16 and 64 years old at the time of the survey, and older than 16 and less than 57 years at the time of arrival. This selection excludes immigrants who finished their studies in Spain, focuses only on those who emigrated directly from their countries of birth to Spain. This leads to a final sample of 7,377 observations (8,064 observations were dropped) of which 945 individuals never worked in Spain. After excluding those individuals who have never worked, we have a subsample of 6,432 observations. Tables A.1 and A.2 in the Appendix detail the sample selection and provides in-depth definitions of the variables used in this study, respectively.

Table 1 presents summary statistics for the final sample, the subsample and the excluded sample. For the final sample, most immigrants come from Latin America (49%) followed by immigrants from Eastern Europe (25%), are on average 34 years old, and have around four years of residence in Spain. In terms of educational attainment, more than half of immigrants have at least secondary level, while approximately a quarter of the sample reports tertiary education level. In addition, more than 75% declares proficiency in Spanish language, and having legal residence authorization.

In order to capture the strength of the social network, two dummy variables are created. Close ties is a dummy variable equal to one if the individual declares having had at least one relative or friend on arrival to Spain. Weak ties are captured through individual's social participation in organizations. Two dummy variables are created in order to distinguish between individuals participating in organizations devoted exclusively to immigrants (non-mixed organizations) and those not (mixed organizations). More than 80% of the immigrants declare having contacts at arrival while social participation in organizations is, on average, low. Individuals participating

[5] The literature addresses this issue through creating synthetic cohort of immigrants by tracking specific immigrant waves across decennial Censuses or across Current Population Surveys (Borjas, 1994). In the present study, the approach considered is analogous, since the ENI is a single cross-sectional database with a 10-year period of analysis.

in mixed organizations represent 10% of the total sample, while 6% of the individuals are involved in non-mixed organizations.

Columns 2 and 3 of Table 1 present summary statistics for the subsample and excluded observations respectively. The comparison between different samples provides a first insight of the potential sample selection bias that could happen when excluding individuals who have never worked in the Spanish labor market. Main differences are observed in terms of gender composition (79% are women in the excluded sample), age (32 versus 34 years old), region of origin (30% of excluded individuals come from North Africa) and years living in Spain (2 versus 4 years). Also, the proportion of immigrants with proficiency in Spanish language and those with legal residence authorization varies across different samples. In addition, differences between the samples are observed in terms of internal mobility across municipalities (grouped as never moved, moved once, or more than once) and in the declared motives for migration. For instance, family regrouping motives is a dummy variable equal to one if the immigrant declares family reunion as a motive for migration. Labor motive is a dummy variable which refers to individual declaring job searching or looking for a better job.[6] Almost 60% of individuals in the excluded sample declare family regrouping motives for migration, in comparison with less than 30% for the final and sub-samples. Finally, in terms of social network variables, no differences are observed across the different samples.

Table 2 presents the summary statistics for immigrants' who have at least worked once in Spain (80% of the final sample). More than 70% of them obtained their first jobs through social networks while 30% of them got the job through formal channels.[7][8] Throughout this text, 'network jobs' and having obtained the first job

[6] The ENI contains self-reported information on the reason for migration, namely due to the presence of a family member or labor motives. As the question in the ENI allows for multiple responses, regrouping motives considers those immigrants that declare family reunion as a motive for immigration, although they could state another motive for migration. Labor motives is a dummy variable that is equal to one if the immigrant declares job searching or looking for a better job as a motive for migration. Further, migration motives were interacted with the region of origin and gender variables in the first equation and did not change the final estimations obtained.

[7] The mechanisms considered are formal methods and social networks. The translated question of the ENI (2007) reads: *By what means did you obtain your first job?* Respondents can choose many options. If the immigrant only chooses one channel, that is, getting the job *through family, friends, or other contacts*, then we consider that the immigrant obtained the first job through social networks. Otherwise, it is considered as getting the job through formal channels. In this sense, formal sources of information include State and private employment agencies, newspaper advertisements, union hiring halls and school and college placement services.

[8] Following Goel and Lang (2011) two issues need to be noted. First, finding a job through the social network does not necessarily imply the presence of a close tie (relative or friend on arrival). This is

through social networks are used interchangeably, as are 'formal jobs' and having obtained the first job through formal channels. Approximately 31% remain in their first jobs, more than 50% have changed jobs, almost 10% are unemployed, while 7% are inactive. About half of these workers were first employed in non-skilled occupations and a quarter in administrative jobs. The main activities in which immigrants are involved in the first job are household activities, construction, and agriculture. In order to explore if differences in observable characteristics exist between immigrants with some labor experience in Spain, Panel A in Table 3 expose summary statistics for those with and without close (columns 1 and 2); and between those with weak ties (columns 3 and 4). A priori, only slightly differences are observed. Immigrants with close ties are on average more women than men, married, and mainly from Latin America. Conversely, the proportion of immigrants with legal residence authorization is higher for immigrants without close ties than for those with close ties. In terms of education and last occupation in the country of origin, no differences are observed between those with and without close ties. However, the proportion of those with close ties and proficiency in the Spanish language is higher than for those without close ties. This is also observed when analyzing immigrants with and without weak ties. Also, those with weak ties are on average more educated. Finally, regional disparities are observed in terms of gender composition, educational attainment, social network endowment and occupational mobility (Tables A.3 and A.4 in Appendix, respectively). It is worth noting that despite the low participation of immigrants in mixed organizations, the proportion of those from Western Europe is three times that for North Africa. In addition, immigrants from Asia and the rest of the world more than double the sample mean of immigrants involved in non-mixed organizations.

Regarding the occupational mobility of immigrants, it is worth noting that workers from Western Europe experience less downward mobility relative to immigrants from other regions (Table A.3 in Appendix), thus reflecting the limited transferability of human capital between non-Western European countries and the Spanish labor market (Simón et al., 2011).

---

because immigrants may have found their job through a friend made after migrating to Spain, or a relative or friend not living in Spain. Thus, having or not obtained the job through social networks does not imply having or not close ties or vice versa. In addition, in contrast with other studies, we measure network use directly, and therefore, we avoid the need to infer network use from clustering of immigrants.

1.3 Methodology

This section presents the empirical approach and identification strategy. The analysis is conducted in two steps. First, we study to what extent social networks affect the job matching process (Section III.1). Second, we analyze whether wage differences could arise for immigrants who maintain their first job due to the presence of close and weak ties and job access mechanisms (section III.2).

1.3.1 Job match and social networks

We analyze the impact of social networks on the quality of the job matching process for immigrant workers. A "good" match is considered if the skills and qualifications of the worker are those required for the job. Then, if a good match between employers and employees takes place, a longer duration of the worker in the same job is expected. Duration models require information on contract job duration. However, the required information is not provided by the ENI. Therefore, as an alternative to these models and based on the information contained in the ENI, the quality of the matching process is studied through the probability of keeping the first job in Spain in comparison to not keeping it, namely relative to being employed in a different job, being unemployed, or being inactive.[9]

Other studies analyze the quality of the job match through a comparison of the last occupation in the country of origin and the occupation obtained in the country of destination (Mahuteau and Junankar, 2008; Simón et al., 2011); or by studying occupational mobility in the host country, comparing the first and the actual occupation in the host country (Simón et al., 2011). However, this approach excludes from the study those immigrants that (i) do not have previous labor experience in the country of birth and (ii) those immigrants that despite having previous experience in the country of birth and a first job in Spain are actually unemployed or inactive. Therefore, the definition of the job matching process considered in this paper includes those

9 An individual is classified as "keeping the first job" if she declares that the actual job is the first obtained in Spain. Specifically, the ENI (2007) asks for actual labor status in Spain. If the individual declares being employed, then she is asked if this is the first job obtained in Spain. If the answer is "yes", the individual is considered to currently be in the first job. Otherwise, if she answers negatively, then we consider she has had a different job since arrival. Employment stability is observed if the immigrant is employed in the first job obtained in Spain.

immigrants that after having a first job in Spain are now in a different job, unemployed or out of the labor market, thereby reflecting job mismatch.

The hypothesis to test is that the probability of keeping the first job is affected by immigrants' close and weak ties as well as the job search mechanisms used to obtain the first job in Spain. Depending on the relationship (positive or negative) found between social networks and actual labor market status, this would reflect the positive or negative impact of social networks on the job matching process between workers and employers.

To assess this relationship we use the following multinomial logit regression:

$$P(Y = j|X) = \frac{\exp(\beta'_j X)}{\sum_{j=0}^{J} \exp(\beta'_j X)}$$

where $P(Y = j|X)$ is the probability of observing the $j \in \{0, J\}$ outcome of the dependent variable Y conditional on the vector X of independent variables. β_j is the vector of regression coefficients to be estimated by the maximum likelihood method.
In this study, the dependent variable (Y) measures four possible labor market statuses, namely being employed in the first job obtained in Spain, being employed in a different job, being unemployed, or being inactive.[10] The independent variables of interest are the immigrant social networks in the host country and job access mechanisms for the first job.

We consider different measures of the strength of immigrants' networks. Close ties is a dummy variable that refers to whether the immigrant had at least one relative or friend on arrival in Spain. Endogenous network formation and the ensuing problem of reverse causality are important empirical issues that need to be addressed in this analysis. For instance, social networks might be affected by labor market outcomes in that labor market status may influence social interaction and social relationships by creating or limiting interaction opportunities. As Goel and Lang (2011) and Kahanec and Mendola (2008) point out, contacts at arrival are largely exogenous with respect to the individual's subsequent labor market experience. The other two measures used in the literature refer to weak ties: participation in social organizations distinguishing those

[10] Inactive refers to those immigrants actually studying or involved in non-waged household activities, excluding retirees.

devoted exclusively to immigrants and those not, and the proportion of immigrants of the same country of birth living in the same region of the total immigrant population in the region as a proxy of the network size (Munshi, 2003; Kahanec and Mendola, 2008). Because the ENI is only representative at national level, the Municipal Register (*Padrón Municipal de Habitantes*) for 2007 was used to calculate the share of immigrants by country of birth in the different Autonomous Communities of Spain.[11]

Besides the key variables of interest, other control variables include socio-demographic characteristics (age, gender, education, region of origin, region of residence in Spain, proficiency in the Spanish language, legal residence authorization), migration experience (internal migration in Spain), remittance behavior, and first job characteristics in Spain (activity sector and occupation). In addition, variables referring to immigrants' labor market status and last occupation in the country of origin are included. These variables are incorporated in order to control for potential unobserved heterogeneity. Identifying the effect of social networks is difficult because unobserved individual attributes such as being sociable, being ambitious, being responsible, can be correlated with both the probability of having contacts at arrival and their own probability of being at different labor market statuses. In addition, social interactions are more likely to emerge among individuals that share some relevant traits, such as education, occupation or ethnicity. Therefore, the estimated effect could be biased and may not be attributable to a network effect. By controlling for several observable characteristics, we are able to partially remove the potential bias arising from omitted personality traits. A priori, it is not clear the direction of the bias. If omitted personality traits affect both labor market outcomes and social network in the same way, neglecting them leads to an upward bias in the coefficient, and thus an overestimation of the effects of the networks in the multinomial regression. Otherwise, the estimated coefficients will be downward biased. A first insight is provided in Table 3 Panel B, in which we observe that the proportion of workers at different labor market statuses is similar between immigrants with and without close ties, and among those with and without weak ties. In order to disentangle the magnitude and direction of the potential bias, the multinomial regressions are estimated with and without the skills variables such as educational level, proficiency of the Spanish language, and previous labor experience in the host country.

[11] An Autonomous Community is a first-level political and administrative division of Spain (NUTS 2).

Another source of concern could be sample selection as the individuals considered in this analysis are those with some experience in the Spanish labor market. In order to correct for this problem a two-step Heckman procedure adapted to logistic regression is implemented, which consists of a two-step estimator and a maximum likelihood estimator (Durbin and Rivers, 1990). In the first step, the probability of having any experience in the Spanish labor market is estimated. The probability that an individual has worked is modeled as a function of individuals' socio-demographic characteristics, social networks, internal mobility, and motives for migration. From this equation, the Mills ratio is estimated. The second step estimates the probability of those immigrants in the labor market being in one of the four outcomes stated before but including the correction coefficient (obtained through the Mills ratio) as an additional covariate. A key issue in this analysis is that the exclusion restriction should not be directly related with subsequent labor market statuses.

In this study, the exclusion restriction includes two dummy variables which refer to migration motives: family regrouping and labor motives. On the one hand, individuals migrating for family reasons may be less prone to work (as they are expected to engage in non-remunerated household activities). On the other hand, given that they have at least one family member when arriving in the host country, it may be easier for them to access job information. In Section II we observed that individuals with and without labor experience in Spain differs in terms of motives declared for migration. While 70% of the individuals with labor experience declare labor reasons for migrating, 60% of those without labor experience declare family regrouping motives (Table 1). We can expect that migration motives and immigrant' subsequent labor status are related, but only indirectly. A possible channel through which migration motives may affect the quality of the job matching process is through its impact on immigrant legal status, since having or not legal residence authorization determines whether immigrants can freely or not search for a better job. Those who migrated for family reasons may have already a family member with legal residence authorization who could provide information on the legalization process, or facilitate their access to legal status, which in turn affects the subsequent labor market status. Conversely, immigrants declaring labor motives may quickly accept a job, because is the most direct path towards being legalized. Thus, because of their precarious situation, they are more prone to accept any kind of job, even if it does not match with their skills. In addition,

by controlling for a broad set of skill variables, we partially remove the unobserved heterogeneity problem.

Reinforcing the exclusion restriction, Aydemir (2011) shows for the Canadian context that immigrants' labor market outcomes highly depend on their skill levels and on the transferability of those skills rather than on visa categories. For the Spanish case, Rodríguez-Planas and Vegas (2012a) find that Moroccan immigrants who declare regrouping motives are less prone to work than immigrants declaring labor motives. Moreover, the authors find that, once the employment decision is accounted for, no wage differentials arise between immigrants declaring different motives for migrating.[12]

In sum, we can assume that migration motives are not expected to directly affect the quality of job match. In formal terms, a good job match depends on workers' supply-side efforts, the number of workers offering those services in the job market, and the demand for their skills and qualifications. For instance, educational level or prior work experience could affect the job match. For immigrant workers, language proficiency, legal status and years living in the destination country are also important issues.

1.3.2. Social networks and wages

Next, we test whether wage differences could arise between immigrants who maintain their first jobs due to the strength and size of an immigrant network and the job access mechanisms used to find the job. The hypothesis to test is that conditional on keeping the first job in Spain, wage differentials could occur between those who found the job through formal or informal methods and between those with or without close and weak ties. The effect of social networks on wages is still a controversial issue in the empirical literature. While Bentolila et al. (2010) find a wage penalty for those workers who found their jobs through personal contacts, Pellizzari (2010) shows that the use of social networks can lead either to a wage premium or to a wage penalty in different EU countries.

[12] These authors stress the potential endogeneity problem in studies that analyze immigrants' labor market outcomes with different types of visa in countries with a clear immigration policy regime in place, which is very likely to be endogenous to the country's social, economic, and political context, and at the same time affect the settlement process of the different types of immigrants it receives. This issue is not present for the Spanish case, considered as an immigrant-friendly country because of the lax implementation of immigration laws and several generous amnesties granting legal residence to illegal immigrants (p.4).

The study of social networks effects on wages consists of estimating a wage equation of the following type:

$$lnw_i = \left(1, NJ_i, CT_i, NJ_i * CT_i, NS_{ij}, WT_i, X_i\right).\gamma + u_i \qquad (1)$$

where w_i is the hourly wage, network job (NJ_i) is a dummy equal to 1 if individual *i* used personal contacts to find the first job and 0 if used formal channels; while close ties (CT_i) is a dummy equal to 1 if the individual had contacts on arrival and 0 otherwise. An interaction term between NJ_i and CT_i is included in order to capture if wage difference between those who found their job through its networks and those who used formal methods is related to the presence of close ties.[13] The network size (NS_{ij}) is measured by the proportion of immigrants of the country of origin of individual *i* living in region *j* of the total immigrant population residing in region *j*. Weak ties (WT_i) is proxied by a dummy variable equal to 1 if individual *i* participates in social mixed organizations, while X is a set of demographic and socio-economic controls (the same as in previous section except remittance behavior) and γ is a column vector with the parameters of the equation.

Equation (1) is estimated by OLS and QR. QRs, introduced by Koenker and Bassett (1978), estimate the conditional quantile function, namely models in which the quantiles of the conditional distribution of the response variable are defined as functions of observed covariates.[14] QRs are used because OLS implicitly assumes no important differences in terms of the impacts of the exogenous variables along the conditional distribution. Instead, if exogenous variables influence the parameters of the conditional distribution of the dependent variable other than the mean, then the analysis that disregards this possibility will be severely weakened. Unlike OLS, QR models allow for a full characterization of the conditional distribution of the dependent variable, bringing much value added if the relationship between the regressors and independent variables evolves across its conditional distribution. Second, unlike the OLS regression that is sensitive to the presence of outliers and can be inefficient when the dependent variable has a highly non-normal distribution, the QR estimates are more robust. Third, unlike

[13] When interpreting the coefficients on close ties, network job and their interaction, it should be noted that the omitted group is that of immigrants in formal jobs and without close ties.

[14] Similar to the OLS method, the parametric QR can be presented as the solution to a minimization problem. In this case, the asymmetrically weighted value of the residuals is considered to compute the parameters. For more details, refer to Koenker and Bassett (1978) and Koenker and Hallock (2001).

OLS, QR estimators do not require existence of the conditional mean for consistency (Cameron and Trivedi, 2005). This flexibility has thus far been precluded from social networks' effects on wages in empirical studies, which has left unaddressed the possible impact of social networks upon inequality through its within-levels inequality component.

Because the sample is restricted to those immigrants still employed in the first job obtained in Spain, sample selection bias could emerge.[15] The nature of the underlying problem requires sample selection models since the conditional quantile of the observed wages depend on a bias term of an unknown form, a two-stage semiparametric method is used. Specifically, the methodology followed to address this issue is the one proposed by Buchinsky (1998) which is similar to the one proposed for mean regression by Heckman (1979).

This study is conducted separately for women and men in order to account for the different factors that may influence wages by gender.[16] First, we estimate the probability of keeping the first job in Spain (the selection equation). Second, the wage equation regression is estimated. This methodology needs at least one variable which explains the probability of keeping the first job but not directly related with the outcome of interest. As in many other studies, finding suitable instrumental variables is far from straightforward, since almost any regressor that determines the probability of keeping the first job could plausibly affect wages as well. The literature commonly uses as exclusion restriction the number of children at home or the marital status. However, these variables may be correlated with wages.[17] Also, variables on tenant or ownership status are used to account for possible sample selection in the decision of participation (Rodríguez-Planas and Vegas, 2012b). In this study, the exclusion restriction is a dummy variable that indicates whether the immigrant sends remittances to her country of origin or not.[18] This variable reflect immigrant responsibilities in the home country,

[15] The sample is restricted because the ENI (2007) only provides wages for actual employment and does not provide information about the mechanisms through which the worker obtained the job. On the contrary, information on job access mechanisms is only given for the first job in Spain. As the aim of this study relies on both wages and job access mechanisms, the sample is restricted to those who keep the first job obtained in Spain.

[16] As the literature on the participation of women in the labor market points out, women's decisions to participate have important implications on their wages.

[17] There are theoretical arguments that suggest that labor supply, wages and fertility are endogenous. If women with relatively low expected future wages had on average a high fertility, the exclusion restriction would fail.

[18] The translated question of the ENI (2007) reads: *Do you sent money out of Spain?* Respondents can choose yes or no.

such as dependent family members or monetary debts (such as mortgage debts or credit), or investment decisions, which may, in turn influence the individual probability of keeping the first job, change jobs, being unemployed or inactive. [19] Moreover, as we only consider whether the immigrant sends remittances or not instead of considering the amount of money remitted, this variable is expected to be unrelated to current wages, since wages strongly depend on actual labor market conditions in the host country, past labor experience in the country of birth and on the worker's human capital endowments.

The literature on economic integration reinforces the exclusion restriction. This literature relates immigrants' remittance behavior with their economic integration in the host country.[20] Studies that analyze the relationship between labor market status and remittance behavior finds that, on the one hand, employed immigrants are more prone to remit than unemployed or inactive immigrants (Bilgili, 2013; Al-Ali and Koser, 2001; Holst and Schrooten, 2006). On the other hand, Holst and Schrooten (2006) find that income has no effect on the probability to remit while it is only significant for the amount of remittances.

The conventional Heckman correction method is applied to the OLS estimation. However, an analysis of the distribution of the error term in the selection equation is needed for QR because the conventional Heckman correction method assumes a standard normal distribution of the error term in the selection equation. If this assumption is violated, then semi-parametric methods should be applied to estimate the first equation, because this method does not rely on a distributional assumption (Buchinsky, 1998). This model (as the conventional Heckman procedure) highly relies on the assumption that the variables included in the exclusion restriction are not related to the outcome variable in the second equation.

The wage equation with semi-parametric correction for sample selection bias is estimated following Buchinsky (1998) (See the Methodological Appendix for a detailed description of the model).

[19] Since the nature of our sample selection bias is different for the one related to the decision of working or not, these variables may be potentially related with wages in our case, thus violating the exclusion restriction assumption. We instead tried with alternative instruments such as home or land ownership in the country of origin, having or not relatives in the country of birth, proving not to be useful instruments. The estimated coefficients in the first stage were not statistically significant. Nonetheless, because of concerns with endogeneity of our instruments we estimated the wage equation including these variables as controls. When doing so, most of the coefficients of interest remain unaffected.

[20] Economic integration of immigrants is stronger when they have higher participation rates, lower unemployment levels, better jobs and, not directly related to labor market participation, higher income per person at the household level (Bilgili, 2013).

The quantiles of the log wage are given by:

$$Q_\theta(y|x_2) = x_2'\beta_\theta + h_\theta(x_1\gamma_0) \; \forall \; \theta \in (0,1) \qquad (2)$$

The vector x_1 is a set of observable characteristics that may affect the probability that an individual keeps the first job obtained in Spain while x_2 is a subset of x_1, which contains labor market characteristics that could influence on the wage offer. In other words, x_1 must also contain at least one variable that is not included in x_2 (the exclusion restriction). These variable (or variables) should be uncorrelated with the log wage. The term $h_\theta(x_1\gamma_0)$ corrects the selection at the θth quantile. It plays the role that the Mills ratio plays in the usual Heckman (1979) procedure, but it is quantile-specific and more general so not to assume normality.

Buchinsky (1998) suggests a series estimator:

$$\hat{h}_\theta(x_1\gamma_0) = \delta_0(\theta) + \delta_1(\theta)\lambda(x_1\gamma_0) + \delta_2(\theta)\lambda(x_1\gamma_0)^2 + \cdots,$$

where $\lambda(.)$ is the inverse Mills ratio defined as $\lambda = \frac{\phi(.)}{\Phi(.)}$, while $\phi(.)$ and $\Phi(.)$ are the density and the c.d.f. of a standard normal variable, respectively. Thus, first γ_0 needs to be estimated. As wages are only observed when the individual keeps the first job, we only observe whether a dummy indicator *D* equals 1 or 0. This could be written as:

$$D \equiv \Pr(D = 1|x_1) \equiv F(x_1\gamma_0) \qquad (3)$$

where x_1 is a set of individual characteristics and $F(x_1\gamma_0)$ is an unknown function. we estimate $F(x_1\gamma_0)$ using a semiparametric estimator proposed by Klein and Spady (1993). This estimator is asymptotically efficient in the sense that it attains the semiparametric efficiency bound (Melly and Huber, 2008).

The two-step semi-parametric method can be summarized as follows:

1. Estimate the probability of not changing the first job using the semi-parametric index proposed by Klein and Spady (1993).
2. Estimate the parameters in the QR including an approximation of the selective term as stated by Buchinsky (1998).

1.4 Empirical findings

1.4.1 Job match and social networks

Table 4 presents the probability of having some labor experience in Spain (the first step in Heckman's method for binary models). Relative to the key independent variables, close ties increase the probability of having some labor experience. However, network size or social participation in any kind of organization has no impact on the likelihood of having some labor experience. The exclusion restriction has an important effect on the dependent variable and it is estimated with precision. In line with previous studies, immigrants who declare labor motives for migration are more prone to have labor experience in Spain, while those declaring family regrouping motives are less likely to be involved in the labor market.

The other control variables have the expected sign. Being a woman, being married, the number of children in the household[21], are all negatively related to the probability of having some experience in the Spanish labor market. The region of origin has different impacts on the probability of labor experience in Spain. Immigrants from North Africa are less prone to participate while those belonging to an Eastern European or a Latin American country all positively influence the probability of having some labor experience. Variables referring to the social assimilation process, such as years living in Spain, having legal residence authorization, and proficiency in the Spanish language, increase the probability of labor experience. Finally, a positive relation is also observed relative to internal migration in Spain.

The estimated multinomial regression after controlling for sample selection is shown in Table A.5 in the Appendix.[22] Because the coefficients obtained through the multinomial logit model do not measure the effect of the explanatory variables on the outcome probability directly, we focus on the results reported in Table5, which shows the average marginal effects of the independent variables on the probability of each of the four labor statuses from the multinomial logit model.

In light of the hypothesis stated, the results provided below reflect a mismatch in the labor market for immigrants on arrival, showing that upon arrival immigrants prefer to quickly accept a job offered through the social network, even if it is not the most suitable given their education, training, or previous experience. This mismatch is

[21] We distinguish between number of children living in Spain and in the country of origin.
[22] All the results of the multinomial model are interpreted in relation to the omitted labor status: being employed in a different job from the first one obtained in Spain.

observed when analyzing the effects of social networks through different labor market statuses (Table 5). For instance, immigrants with close ties are less likely to keeping the first job while more prone to change jobs (8.9 and 5.2 percentage points, in columns 1 and 2 respectively) reflecting the importance of close ties in terms of job information transmission or financial support when immigrants search for another job. Ottaviano and Peri (2006) argue that job mismatch could happen because jobs found through relatives and friends are often unrelated to the individuals' previous experience or training. This is the case when the social capital accumulated by the network is restricted to a particular segment of the labor market, in which case the new immigrant's job prospects are limited to this segment. Therefore, this great dependence on social networks may also reflect segmentation in the host labor market as well as a lack of access to host labor market institutions.

In addition, a mismatch is more likely to be observed for immigrants who obtained the first job through social networks in comparison to those who used formals channels (6.5 percentage points in column 1). Interesting, those immigrants who have had a first network job are more prone to being unemployed (4.6 percentage points), thus reinforcing the negative effect of informal job access channel on the matching process. Further, for immigrants who took less than a month to find the first job, those who obtained the job through their social networks are more likely to be mismatched in relation to those who used formal search methods. In line with our results, Bentolila et al. (2010) find a mismatch for workers who access their current job through social networks for the US and Europe. According to these authors, workers have a natural talent for a specific occupation, which may not be the one to which their social contacts can provide referrals. In this scenario, workers may have to accept a trade-off; they may find it advantageous to find a job more quickly through their social networks, but they may also work in an occupation that does not maximize their productivity.

The results in Table 5 column (1) show that the probability of keeping the first job decreases for immigrants with weak ties (13.8 percentage points), but it is not statistically significant on the probability of changing jobs. We find that conditional on having obtained the first job through social networks, the probability of keeping the job is independent of the network strength. This effect is measured through the sum of the coefficients on close ties and this variable interacted with network jobs (almost 0). A statistical and significant effect (positive or negative) would imply that immigrants with close ties are better or worse workers, having different proclivities to receive network

and formal offers, than for those without them. The interaction term between close ties and network job measures the causal effect of having close ties on the difference between the expected probabilities of keeping the first job conditional on *choosing* different channels to obtain the first job. Our results show that the probability of keeping a network job is higher for immigrants with close ties than the probability of keeping a formal job for those without close ties (8.7 percentage points).

In addition, the bigger the network size the more likely that the immigrant keeps the first job. No statistical and significant effects of network size, weak ties or the job access mechanisms, on the probability of changing jobs are found (column 2 in Table 5). Loury (2004) points out that differences between industries and employers may also account for ethnic and race variations in contact effects. Ethnics groups have established specific occupational and employment niches that facilitate employment and training of members of their group and that limit access of outsiders. This may explain the positive effects of network size in the probability of keeping the first job. This is also consistent with Veira and Stanek (2011) who find ethnic niches in the Spanish labor market.

Next, we explore the effects of close and weak ties on the probability of being unemployed (column 3 in Table 5). Contacts on arrival and network size do not influence the likelihood of unemployment. However, the probability of unemployment decreases for those immigrants with more years living in Spain participating in social mixed organizations, reflecting a positive effect of individual's social integration. Finally, a positive effect of close ties on the probability of being inactive is observed (3.4 percentage points), while immigrants who got the first job through social networks and with close ties are less prone to being inactive relative to those who got the job through formal channels and without close ties (column 4, Table 5).

While the primary interest of this study is on social networks, a brief look at the results of the control variables is provided. The results reported in Table 5 are consistent with previous findings in the literature. For example, being a woman increases the probability of unemployment or being inactive, while decreases the probability of changing jobs. Immigrants from Western Europe experience better matches in the Spanish labor market than other immigrants groups. Different impacts of regions of destination on labor statuses outcomes are also found, reflecting differences in labor market conditions and opportunities for immigrant workers across Spain. Years living in Spain decreases by almost 5 percentage points the probability of keeping the first employment. Consistent with the idea that legal migrants can search freely in the host

labor market, those with legal residence authorization are more prone to change jobs and less prone to be unemployed.

Statistical and significant effects of human capital endowment on the probability of being in different labor market statuses are found. The probability of changing jobs decreases for immigrants with secondary education, while immigrants with tertiary education are less likely to be unemployed. Proficiency in the Spanish language decreases the probability of unemployment. Immigrants with skilled occupations in the country of origin are more likely to be mismatched upon arrival. Specifically, they are more likely to switch jobs possibly for a better one, more in accordance with their previous labor experience. Being a student before migration also increase the probability of mismatch upon arrival.

Overall, our results support previous studies that stress the difficulties in transferring immigrants' previous labor experience and credentials. Once established in the host country, immigrants search for a new job more in accordance with their levels of education, previous experience, and training. Also in line with our results, Simón et al. (2011) and Veira and Stanek (2009) find a U-shaped pattern in terms of occupational mobility for immigrants in Spain, characterized by occupational downgrading on arrival and a gradual improvement as the duration of residence in the host country increases.

First job characteristics also influences on the job matching probability. Immigrants employed in qualified occupations, such as managers or skilled workers are more likely to experience a good match, as well as those employed in any sector in comparison to those employed in agriculture.

Finally, immigrants sending remittances to their country of birth are less prone to keep the first job while are more likely to changing jobs. It is well addressed in the literature that immigrants' remittances are very important to financially support stayers in the country of origin, namely own children, parents or other family members. Considering these motives, immigrants probably put more effort in searching for better jobs, more stable or with better labor conditions.

In order to be more confident in the presented results, some robustness checks are made. First, a separate analysis is conducted for women and men. The magnitudes of the coefficients of the key independent variables varies across gender, however the relationship between social networks and job matches exposed above remain (Table 6). In addition, in order to get some insight of the sign and magnitude of the potential bias due to unobservable characteristics, we estimate the marginal effects excluding

measured skill variables, such as educational level, proficiency in the Spanish language, labor status before migration and last occupation in the country of birth. As can be seen in Table 7, the magnitude and sign of the key independent variables are similar to the ones provided in Table 5. The estimated coefficients without controlling for these variables would be downward biased for close and ties and informal search mechanism, while upward biased for network size and weak ties coefficients. Finally, in order to remove any concern with endogeneity of the variables included in the exclusion restriction, we re-estimate the multinomial logit model and the average marginal effects including the motives for migrating as controls. The estimated coefficients of our key variables do not change significantly (Table 8).

While it remains possible that there is an important measure of skill that is correlated with the immigrant' social networks, the fact that excluding these extensive set of variables does not alter the results in an important way, gives a reasonable level of confidence in the presented results.

1.4.2 Social networks and wages

This subsection aims to study the impact of social networks on wages for male and female immigrants keeping the first job.[23] For this purpose, a two-step procedure is used in order to control for the possible selection bias arising from selecting workers that keep their first jobs in Spain (see Section III.2). This analysis is conducted separately for women and men.

In the first step, the probability of keeping the first job, the selection equation, is estimated using a standard probit model. After probit regression, the hypothesis of the normality of the residuals is rejected for women but not for men (Table 9). Hence, for women, the selection equation is estimated using the semiparametric estimator proposed by Klein and Spady (1993). For men, a standard probit model is estimated. The exclusion restriction is a dummy variable equal to one if the immigrant sends remittances to her country of birth and zero otherwise (see Section III.2).

The factors influencing the probability of keeping the first job was largely analyzed in the previous section. Therefore, for the sake of brevity, we do not describe the results obtained for the selection equation, which are reported in Tables A.6 and A.7

[23] Recall that in this case, the sample is restricted to those immigrants who keep the first job, since the ENI provides information on the job access mechanisms only for the first job, while wage information is provided for current job.

in the Appendix, for women and men respectively.[24] However, it is worth noting that the exclusion restriction, the dummy variable that indicates if the individual sends remittances, is statically significant and negatively related to the probability of keeping the first job for both women and men.[25]

The results presented in Table 10 reflect different impacts of social networks on wages by gender and across the observed wage distribution. For instance, the job access mechanism influences wages. Both women and men who have obtained the job through social networks present a wage penalty in comparison to those who used formal channels. This penalty is present across the distribution, observed for the 25th and 50th percentile for women and among different percentiles for men. However, some important dissimilarities between female and male appears, for instance in the magnitude and significance of the coefficients. Having obtained the job through social networks has a lower negative impact for women. At the 25th percentile the gap is 3.7% and statistically significant at 10%, and this pattern is observed until the 50th percentile where the coefficient is 0.034. For men, the wage gap is around 11.3% at the 25th percentile, 10.8% at the 50th percentile, and 11.7% at the 75th percentile, and statistically significant at 1% in each percentile. These estimates evidence how the penalty for being employed in a network job has also a gender dimension that favors female.

These results are in line with Bentolila et al. (2010) who find a wage penalty across workers who obtained the job through informal channels. According to Pellizzari (2010), the positive or negative effects of social networks on wages could be related to employer characteristics, which in turn determine the context in which job search methods operate. It could be the case that for some employers, desired applicant characteristics may be easily discernible through formal channels rather than relying in recommendations from trusted sources. Pellizzari (2010) finds substantial variations in the effects of social networks on earnings. This author states that wage penalties are likely to happen in industries where firms invest substantially in formal recruitment activities. Firms are more likely to undertake such investments for high productivity

[24] Strictly speaking, the estimated coefficients of the semi-parametric model are not comparable with the ones obtained in the previous section through the multinomial logit regression. This is so because the coefficients estimated in the semi-parametric model only indicate the sign of the effect, but not the elasticity, which could be obtained through the estimation of average marginal effects.

[25] An important difference between women and men, is that the probability of keeping the first job for women decreases with the number of children in the country of birth and in Spain, while for men this is not statistically significant. So, for men we only consider the total number of children.

jobs where the cost of turnover is substantial. When large investments are made, workers found through formal channels average higher productivity than those found through other means. An alternative explanation is that referred workers are segregated into low wage types of jobs with respect to no referred workers. Then, workers who access job through social networks earn less than those who used formal mechanisms.

Looking at the strength of the social network, we observe that close ties only affect men' but not women' wages. For women, the estimated coefficients of close ties and the interaction term between close ties and network job is not statistically significant, meaning that regardless of the channel of access to employment, the presence of close ties does not have statistically significant effect on wages. Conversely, conditional on having found a job through formal channels, a glass ceiling effect of close ties on men' wages is observed. This effect refers to a wider wage gap at the top of the distribution, suggesting that those men who obtained the job through formal channels with close ties in the high-income jobs earn less than workers without close ties. In other words, hourly wages decreases with close ties throughout the conditional wage distribution For instance, the return to having close ties decreases from 8.4% to 11.5% between the 25th and 75th percentile. This could be interpreted as a negative ability returns relationship as evidence that having close ties and ability are related, which if true suggest that less able individuals benefit less from the presence of close ties. However, because individuals' abilities are unobserved for the researcher, it is difficult to isolate the effect that drives the heterogeneous pattern of returns to personal contacts across the wage distribution.

When interacting close ties and network job variables, the coefficient shows that the returns to the channels of search differ for men with and without close ties. The positive and statistically significant coefficient observed at different quintiles of the distribution shows that immigrants who got the job through social networks and with close ties are higher than for those who obtained the job through formal channels and do not have close ties. In other words, a network premium (understood as the difference of wages between network jobs and formal jobs) is observed across the wage distribution. Moreover, this wage premium increases for higher percentiles reflecting a sticky floor effect. This effect is observed when the gap widens at the lower percentiles of the wage distribution.

Next, the role of weak ties on wage distribution is analyzed. The estimated coefficients show great differences across genders. For women no statistically

significant effects are found. Conversely, a wage penalty is observed in the 25th and 50th percentiles of the distribution for men. For the highest percentile, the estimated coefficient is still negative but no statistically significant. However, this penalty is reversed as the length of time living in Spain increases possibly reflecting the positive effects of the social integration process in the host country (Table A.9).

The network size penalizes both women and men wages. This effect is observed for the median of the distribution for both genders, and in the 75th percentile of the distribution only for men. This is consistent with Calvó-Armengol and Jackson (2007) who state that in the short run, network size has a negative impact on labor market outcomes due to competition for job information within the network, which negatively affect immigrants' wages. Other explanations points out the strong presence of immigrants from the same country of origin may indicate the presence of immigrant enclaves and, therefore, segmentation in some occupations in the labor market, which results in wage penalties (Chiswick and Miller, 2005). This possible explanation is the counterpart of the results exposed above, that social integration (as opposite to enclaves' formation) in the host country positively affects wages.

In the case of the estimates of the control variables, the results reported in Tables A.8 and A.9 in the Appendix are in the direction one would expect. Covariates referring to socio-demographic characteristics, such as marital status and number of children living in the immigrant' country of birth has different impacts on wages across gender. While being married penalizes women wages (statistically significant in the 50th percentile of the distribution), a wage premium is observed for men across different percentiles of the distribution. In addition, the marital status and the number of children loses significance for higher percentiles of women' wage distribution.

The region of origin also impact on wages. Immigrants from Western Europe present a wage premium in comparison to other immigrant groups. This wage premium is observed in the 50th and 75th quantiles of the distribution for women, and across the whole distribution for men. In addition, wage differentials are observed within the Spanish territory. This could be reflecting regional disparities in terms of productive structures and labor markets dynamics in Spain.

In line with the literature, immigrants with legal residence authorization present a wage premium present across the distribution. Differences in power negotiation between immigrants with and without legal residence might explain this result. Human capital endowments positively affect wages. Immigrants with tertiary education present

a wage premium at different quintiles (statistically significant for women at the 50th percentile of the distribution, and for different percentiles for men). Men with proficiency in Spanish language earn more than men without it. Conversely, proficiency in Spanish language does not affect women wages.

Variables referring to actual occupation and last occupation in the country of origin are also relevant on affecting wages. Men in skilled occupations present a wage premium across the different percentiles. For women in professional and managerial activities, positive and statistically significant returns on wages are observed for the 50th and 75th percentile of the distribution. Similar effects of last occupation in the country of origin are observed for both genders. These result are not surprising, it is expected that more qualified occupations pays better, and premiums those workers that have the human capital endowments and previous experience required for the job. However, for less skilled jobs, other factors such as the region of origin, the legal status, or the years living in the host country seems to be important individual attributes and more relevant than those referring to human capital endowments or previous experience in the country of origin. It could be also the case that for employers these socio-demographic factors are relevant for screening workers.

Finally, the actual sector of activity has different returns on wages and across genders. For women, the only sector that is significant is the household activity sector in comparison to agriculture. In this case, a wage penalty is observed for the 50th percentile of the distribution. For men employed in construction, returns are higher than wages in the agriculture sector, and this is observed across the wage distribution. This is consistent with the construction boom that took place in this period in Spain, and the consequent high labor demand of this sector. In addition, men working in industry or in firm services present wage premiums across the wage distribution. The other activities namely, trade, education and health services, and transportation, present a wage premium in the 50th and 75th percentiles. The only sector that presents a wage penalty is the household activity.

1.5 Conclusion

This paper investigates the extent to which social networks influence immigrants' labor market outcomes in Spain. Using micro-data from the ENI, we identify the effect of social networks by examining the effect of close and weak ties, network size and job access mechanisms on immigrants' labor market outcomes. The empirical strategy is conducted in two steps. First, we study the impact of social networks on the probability of being in different labor market statuses. Second, for those immigrants who keep the first job, we study whether wage differentials could arise due to the presence of social networks. Because sample selection could arise in this study, the analyses are conducted in a two-step procedure similar to the one proposed by Heckman. In addition, a broad set of control variables are included in order to control for potential unobserved heterogeneity.

The findings reported in this paper indicate that a mismatch takes place in the labor market for immigrants on arrival. Immigrants tend to quickly accept a job offered through the social network, even if it is not the most suitable job given their levels of education, training, and previous experience. Once established in the host country, immigrants search for another job possibly more in accordance with their human capital endowment. Second, different effects of social networks on wages by gender and across the wage distribution are observed for immigrants who keep the first job. Workers who obtained the job through social networks present a wage penalty in comparison to those who used formal channels. This is observed for the 25th and 50th percentile for women and among different percentiles for men. In addition, the strength of the network only penalizes men' wages but do not influence women's wage. As the length of time living in Spain increases, men' participating in social mixed organizations present wage premium in comparison to those not participating. The network size also penalizes both women's and men's wages. Conditional on having obtained the first job through social networks, men with close ties present wage premium in comparison to those who got the job though formal channels and without close ties. This effect is not statistically significant for women.

To sum up, two main factors influence immigrants' labor market outcomes. First, their great reliance on personal contacts as a job access mechanism is reflected in a mismatch in the labor market and in wage penalties across the distribution for both women and men. The positive effect of network size on job match and its negative impact on wages may be reflecting the presence of segmentation in some occupations in

the labor market. Second, human capital endowment are partially transferred to the host country, negatively affecting the matching process upon arrival.

In light of these results, some considerations are made. First, it is important to stress that policies whose objectives are to accelerate the assimilation process or improve the labor market outcomes of immigrants not only have to focus on the individual (such as improving human capital endowments), but might also influence individuals' social backgrounds and the social networks within which an immigrant is embedded. If this strong dependence on social networks persists over time, the integration process of immigrants in Spain may be compromised. Second, the adaptation process of immigrants to labor institutions and transferability of previous experience and education should be addressed.

Acknowledgements

We appreciate the comments and suggestions made by the participants of the 2013 EEA-ESEM meeting, session "Social Netwroks II", held at the University of Gothenburg; participants of the "Annual Meeting on Equaliity and Poverty: Implications and Methods", at Universitat Autònoma de Barcelona, Spain, December, 2012; and the participants of the "Doctoral Day XTREPP" workshop, at Universidad de Barcelona, November 2012. We are especially grateful to Cristina López Mayan for her insightful comments and to Javier Vázquez Grenno for carefully reading this essay and for his helpful comments and suggestions.

References

Aguilera, M. (2003) "The Impact of the Worker: How Social Capital and Human Capital Influence the Job Tenure of Formerly Undocumented Mexican Immigrants", *Sociological Inquiry*, 73(1): 52-84.

Aguilera, M. and Massey, D. (2003) "Social capital and the Wages of Mexican Migrants: New Hypothesis and Tests", *Social Forces*, 82(2): 671-701.

Al-Ali, N., Black, R.; and Koser, K. (2001) "Refugees and transnationalism: The experience of Bosnians and Eritreans in Europe". *Journal of Ethnic and Migration Studies,* 27 (4), 615-634.

Alcobendas, M., and Rodríguez-Planas, N. (2009) "Occupational Assimilation After a Recent Immigration Boom", IZA DP No. 4394.

Amuedo- Dorantes, C., and de la Rica, S. (2007) "Labor Market Assimilation in Spain", *British Journal of Industrial Relations* 45(2): 257-285.

Aslund, O., and Rooth, D-O. (2007) "Do when and where matter? Initial Labor Market Conditions and Immigrants Earnings", *The Economic Journal* 117(March): 422-448.

Aydemir, A. (2011) "Immigrant Selection and Short-Term Labor Market Outcomes by Visa Category", *Journal of Population Economics*, 24: 451-475.

Bentolila, S., Michelacci, C., and Suarez, J. (2010) "Social Contacts and Occupational Choice", *Economica*, 77: 20-45.

Bertoli, S., Fernández- Huertas, J.; and Ortega, F. (2010) "Immigration Policies and the Ecuatorian Exodus". IZA DP No. 4737.

Bertrand, M.; Luttmer, E.; and Mullainathan (2000) "Network Effects and Welfare Cultures", *Quarterly Journal of Economics*, 115(3): 1019-1055.

Bilgili, Ö. (2013). "The links between economic integration and remittances behaviour of migrants in the Netherlands", UNU-MERIT WP 037.

Borjas, G. (1985) "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants", *Journal of Labor Economics*, 3(4): 463-489.

Borjas, G. (1994) "The Economics of Immigration", *Journal of Economic Literature*, 32(4): 1667-1717.

Borjas, G. (1995) "Ethnicity, neighborhoods, and human capital externalities", *American Economic Review*, 85(3): 365-390.

Borjas, G. (2000) "The Economic Progress of Immigrants", (in) *"Issues in the Economics of Immigration",* National Bureau of Economic Research, Inc.

Buchinsky, M. (1998) "The dynamics of changes in the female wage distribution in the USA: a quantile regression approach", *Journal of Applied Econometrics*, 13: 1-30.

Buchinsky, M. (2001) "Quantile regression with sample selection: Estimating women's return to education in the U.S.", *Empirical Economics*, 26: 87-113.

Calvó-Armengol, A. (2004) "Job Contact Networks", *Journal of Economic Theory*, 115: 191-206.

Calvó-Armengol, A., and Jackson. M. (2004) "The effects of social networks on employment and inequality", *American Economic Review* 94(3): 426-454.

Calvó-Armengol, A., and Jackson. M. (2005) "Job matching and word-of-mouth communication", *Journal of Urban Economics* 57: 500-522.

Calvó-Armengol, A., Patacchini, E.; and Zenou, Y. (2009) "Peer effects and social networks in education", *Review of Economic Studies* 76: 1239-1267.

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

Cappellari, L., and Tatsimaros, K. (2010) "Friends' networks and job finding rates". CESifo WP, No. 3243.

Carrasco, R., Jimeno, J.F.; and Ortega, C. (2008) "The effect of immigration on the labor market performance of native-born workers: some evidence for Spain", *Journal of Population Economics*, 21: 627-648.

Chiswick, B.R., and Miller, P.W. (2005) "Do Enclaves Matter in Immigrant Adjustment?", *City and Community*, 4: 5-35.

De Luca, G. (2008) "SNP and SML estimation of univariate and bivariate bynary-choice models", *The Stata Journal* 8(2): 190-220.

Dubin, J., and Rivers, D. (1990) "Selection Bias in Linear Regression, Logit and Probit Models", *Sociological Methods and Research*, 18(2 & 3): 360-390.

Dustmann, C., Glitz, A., and Schonberg, U. (2010) "Referral based Job Search Networks", unpublished paper, Department of Economics, University College London.

Edin, P., Fredriksson, P.; and Aslund, Ö. (2003) "Ethnic Enclaves and the Economic Success of Immigrants. Evidence from a Natural Experiment", *Quarterly Journal of Economics*, 118(1): 329-357.

Eichhorst, W., Escudero, V., Marx, P.; and Tobin, S. (2010) "The impact of the Crisis on Employment and the Role of the Labour Market Institutions", IZA DP No. 5320.

Elliot, J. (1999) "Social Isolation and Labor Market Isolation: Network and Neighborhood Effects on Less Educated Urban Workers", *Sociological Quarterly*, 40: 199-216.

Espinosa, K., and Massey, D. (1999) "Undocumented Migration and the Quantity and Quality of Social Capital", (in) *"Migration and Transnational Social Spaces. Research in Ethnic Relations",* Pries, L. (ed.) Hants, Ashgayr Publishing.

Fernández-Huertas, J. (2008) "Wealth Constraints, Skill Prices or Networks: What Determines Emigrant Selection?", UFAE and IAE WP 741.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).

Fernandez Kelly, P. (1995) "Social and Cultural Capital in the Urban Ghetto: Implications for the Economic Sociology of Immigration", (in) *"Essays on Networks, Ethnicity amd Entrepreunership"*, Portes, A. (ed.), New York, Russel Sage Foundation.

Goel, D., and Lang, K. (2011) "Social Ties and the Job Search of Recent Immigrants".
http://www.econ.upf.edu/docs/seminars/lang.pdf

Goel, D., and Lang, K. (2012). "Social Ties and the Job Search of Recent Immigrants". http://people.bu.edu/lang/network.pdf

Granovetter, M. (1973) "The Strength of Weak Ties", *American Journal of Sociology*, 78(6): 1360-1380.

Granovetter, M. (1974; 1995) *"Getting a Job: A Study of Contacts and Careers"*, first edition, Harvard University Press, second edition, The University of Chicago Press, Chicago, Illinois.

Hanson, G. (2006) "Illegal migration from Mexico to the United States", *Journal of Economic Literature*, 44(4): 869-924.

Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1): 153-161.

Holst, E., and Schrooten, M. (2006). "Migration and Money: What determines Remittances? Evidence from Germany", DIW Berlin, DP 566.

Ioannides, Y., and Loury, L. (2004) "Job Information Networks, Neighborhood Effects and Inequality", *Journal of Economic Literature*, 42(4): 1056-1093.

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

Jasso, G., and Rosenzweig, M. (1995) "Do Immigrants Screened for Skills Do Better than Family Reunification Immigrants?", *International Migration Review*, 29: 85-111.

Kahanec, M., and Mendola, M. (2008) "Social Determinants of Labor Market Status of Ethnic Minorities in Britain", Centro Studi Luca d'Agliano, No. 253.

Jackson, M. (2008) *"Social and Economic Networks"*, Princeton University Press.

Klein, R., and Spady, R. (1993) "An efficient semiparametric estimator of the binary response model", *Econometrica*, 61(2): 387-421.

Koenker, R., and Basset, G. (1978) "Regression Quantiles", *Econometrica*, 46: 33-50.

Loury, L. (2006) "Some Contacts Are More Equal Than Others: Earnings and Job Information Networks", *Journal of Labor Economics*, 24(2): 299-318.

Llul, J. (2008) "The impacts of immigration on productivity", CEMFI WP 0802.

Mahuteau, S., and Junankar, P.N. (2008) "Do Migrants get Good Jobs in Australia? The Role of Ethnic Networks in Job Search", *The Economic Record*, 84(Special Issue): S115-S130.

Manski, C. (1995) "*Identification Problems in the Social Sciences"*, Harvard University Press.

Manski, C. (2003) "*Partial Identification of Probability Distributions"*, Springer-Verlag.

Montgomery, J. (1991) "Social Networks and Labor-Market Outcomes: Toward and Economic Analysis*", American Economic Review*, 81(5): 1408-1418.

Montgomery, J. (1992) "Job Search and Network Composition: Implications of the Strength-of-Weak-Ties Hypothesis", *American Sociology Review*, 57(5): 586-596.

Munshi, K. (2003) "Networks in the modern Economy: Mexican migrants in the US labor market", *Quarterly Journal of Economics*, 549-599.

Nee, V., and Sanders, J. (2001) "Understanding the diversity of immigrant incorporation: a forms-of-capital model", *Ethnic and Racial Studies*, 24(3): 386-411.

OECD (2010) "International Migration Outlook"

Ottaviano, G., and Peri, G. (2006) "The Economic Value of Cultural Diversity: Evidence from U.S. Cities", *Journal of Economic Geography*, 6: 9-44.

Patacchini, E., and Zenou, Y. (2008) "Ethnic networks and employment outcomes". IZA DP No. 331.

Pellizzari, M. (2010) "Do Friends and Relatives Really Help in Getting a Good Job?", *Industrial & Labor Relations Review*, 65(3), article 7.

Rees, A. (1966) "Information Networks in Labor Markets", *American Economic Review*, 56(1-2): 559-566.

Reher, D. (2008): Informe Encuesta Nacional de Inmigrantes (ENI-07), INE, DT 2-08.

Rodríguez-Planas, N., and Vegas, R. (2012a) "Moroccans' Assimilation in Spain: Family-Based versus Labor-Based Migration", IZA DP No. 6368.

Rodríguez-Planas, N., and Vegas, R. (2012b) "Moroccans', Ecuadorians' and Romanians' Assimilation in Spain" IZA DP No. 6542.

Simón, H., Ramos, R.; and Sanromá, E. (2011) "Occupational Mobility of Immigrants in a Low Skilled Economy: The Spanish Case", IZA DP No. 5581.

Smith, S. (2000) "Mobilizing social resources: Race, ethnic, and gender differences in social capital and persisting wage inequalities", *Sociological Quarterly*, 41(4): 509-537.

Stark, O., and Wang, Y. (2002) "Migration Dynamics", *Economic Letters*, 76(2): 159-164.

Topa, G. (2001) "Social Interactions, Local Spillovers and Unemployment", *Review of Economic Studies*, 68: 261-295.

Veira, A., and Stanek, M. (2009) "Occupational transitions and social mobility at migration to Spain", Grupo de Estudios "Población y Sociedad" (Universidad Complutense de Madrid), DT No. 4 (III).

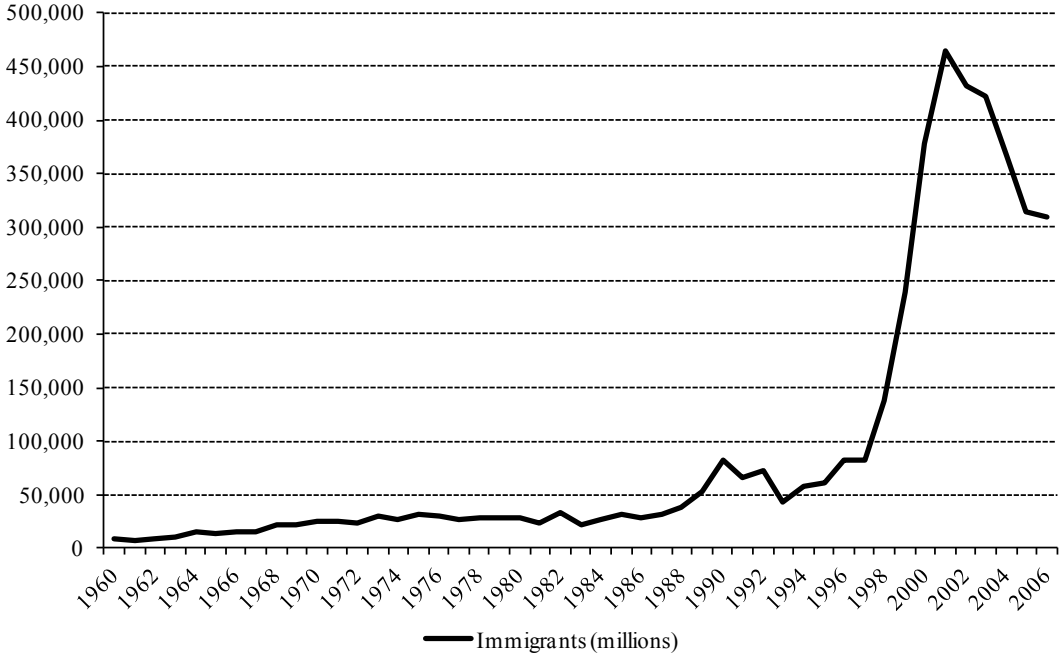
Veira, A., Stanek, M.; and Cachón, L. (2011) "Los determinantes de la concentración étnica en el mercado laboral español", *Revista Internacional de Sociología*, 69(M1): 219-242.

Wahba, J., and Zenou, Y. (2005) "Density, social networks and job search methods: theory and application to Egypt", *Journal of Development Economics* 78: 443-473.

Zenou, Y. (2009) *"Urban Labor Economics"*. New York: Cambridge University Press.

TABLES AND FIGURES

Graph 1. Evolution of the new immigrants in Spain by year of arrival



Source: ENI (2007)

Table 1 Descriptive statistics for socio-demographic variables

Variables	Sample (1)	Subsample (2)	Excluded (3)
Female	0.57	0.53	0.79
Age (years)	34	34	32
Years since arrival	4.11	4.31	2.69
Married	0.52	0.50	0.64
Number of children	1.27	1.25	1.39
Residence authorization	0.75	0.76	0.67
Education			
Primary level	0.19	0.18	0.22
Secondary level	0.55	0.57	0.44
Tertiary level	0.26	0.24	0.35
Speaks spanish	0.76	0.80	0.47
Region of origin			
Western Europe	0.08	0.07	0.14
Eastern Europe	0.25	0.26	0.16
Latin America	0.49	0.52	0.31
North Africa	0.13	0.10	0.30
Asia	0.02	0.02	0.03
Rest of the world	0.03	0.03	0.05
Migration between municipalities. Frecuency (%)			
1. Never moved	0.29	0.24	0.63
2. Moved once	0.35	0.37	0.22
3. More than one	0.36	0.39	0.15
Motives for migration [1]			
Labor motives	0.64	0.69	0.27
Family regrouping	0.27	0.22	0.59
Social networks			
Contacts at arrival (Close ties)	0.83	0.83	0.86
Social participation (exclusive for immigran	0.06	0.06	0.06
Social participation (mixed organization)	0.10	0.10	0.09
Remmitances	0.56	0.61	0.21
Observations	7,377	6,432	945

1. More than one motive could be chosen. The options given in the ENI (2007) are: being unemployed, search for a better job, jubilation, better quality of life, family regrouping, politic motives, religious motives, others. Labor motives include being unemployed or search for a better job.

Table 2 Descritpive Statistics. Labor outcome in Spain

Variable	Freq.
Labour experience in Spain	87.19
Obs.	7,377
Dependent variables	
Maintain first job	29.71
Actual job different first job	53.73
Unemployed	9.87
Inactive[1]	6.70
First job characteristics (dummy variables)	
Job access mechanisms	
Social Networks	0.70
Formal methods	0.29
Occupation	
Manager	0.01
Professional	0.06
Paraprofessional[2]	0.27
Skilled workers[3]	0.18
Unskilled workers	0.48
Sector of activity	
Agriculture	0.16
Industry	0.08
Construction	0.15
Trade	0.07
Hotel sector	0.15
Transportation	0.03
Business services	0.06
Education- Health	0.06
Household activities	0.25
Public administration	0.00
Time before finding the first job (dummy variables)	
Jobs proposal before migration	0.16
Less than one month	0.40
Between 1 and 3 months	0.19
Between 4 and 12 months	0.17
More than one year	0.04
Not known	0.03
Last occupation in the country of birth (dummy variables)	
Manager	0.04
Professional	0.17
Paraprofessional[2]	0.27
Skilled workers[3]	0.24
Unskilled workers	0.12
Never worked at origin	0.15
Obs.	6,432

1. Inactive excludes those immigrants thar are retired.
2. Includes administrative workers, comercial salers, personal service workers.
3. Includes qualified workers employed in industrial or agricultural activities.

Table 3 Observable differences across immigrants' network strength

Variables	With CT (1)	No CT (2)	With WT (3)	No WT (4)
Panel A				
Socio-demographic characteristics				
Female	0.55	0.44	0.51	0.54
Age (years)	34	35	35	34
Years since arrival	5.1	5.7	5.5	5.1
Married	0.50	0.46	0.50	0.50
Number of children	1.25	1.29	1.19	1.26
Residence authorization	0.76	0.81	0.78	0.76
Education				
Primary level	0.19	0.17	0.11	0.19
Secondary level	0.58	0.55	0.56	0.57
Tertiary level	0.24	0.27	0.32	0.23
Speaks spanish	0.82	0.73	0.86	0.79
Region of origin				
Western Europe	0.06	0.11	0.13	0.07
Eastern Europe	0.26	0.27	0.18	0.27
Latin America	0.54	0.39	0.57	0.51
North Africa	0.09	0.13	0.06	0.10
Asia	0.02	0.03	0.02	0.02
Rest of the world	0.02	0.07	0.04	0.03
Last occupation in the country of birth (dummy variables)				
Manager	0.04	0.05	0.25	0.16
Professional	0.17	0.17	0.26	0.27
Paraprofessional[2]	0.27	0.27	0.20	0.25
Skilled workers[3]	0.24	0.25	0.11	0.09
Unskilled workers	0.12	0.14	0.09	0.13
Never worked at origin	0.15	0.13	0.13	0.15
Panel B				
Labor market status				
Keep job	0.30	0.29	0.27	0.30
Change job	0.53	0.55	0.59	0.53
Unemployed	0.10	0.10	0.08	0.10
Inactive	0.07	0.05	0.07	0.07
Observations	5344	1088	656	5776

Table 4 Probability of labor experience in Spain. Logit regression

Variable	Coefficient	SE
Key independent variables		
Close ties	0.679***	(0.173)
Social participation. Non mixed organizations	0.076	(0.246)
Social participations. Mixed organizations	-0.041	(0.208)
Migrant proportion	0.563	(0.704)
Motives for migration (reference: other motives)		
Labor	1.260***	(0.144)
Family regrouping	-0.683***	(0.136)
Other controls		
Female	-0.915***	(0.138)
Age	0.215***	(0.045)
Age^2	-0.003***	(0.001)
Married	-0.472***	(0.134)
No. children Spain	-0.262***	(0.059)
No. children origin	0.100	(0.086)
Years since arrival (years)	0.387***	(0.036)
Residence authorization	0.901***	(0.143)
Educational attainment (reference: Primary level or less)		
Secondary level	0.128	(0.170)
Terciary level	-0.251	(0.182)
Speaks spanish	0.933***	(0.157)
Region of origin (reference: Western Europe)		
Eastern Europe	1.407***	(0.247)
Latin America	0.776***	(0.214)
North Africa	-0.269	(0.241)
Asia	0.053	(0.415)
Rest of the world	-0.221	(0.371)
Region of destination (reference: Madrid)		
Andalucía	-0.389	(0.283)
Aragon	-0.778***	(0.278)
Asturias	-0.409	(0.417)
Balears	-0.623**	(0.259)
Canarias	-0.715**	(0.287)
Cantabria	-0.802***	(0.301)
Castilla Leon	-0.772**	(0.312)
Castilla la Mancha	-0.919***	(0.286)
Catalonia	-0.799***	(0.205)
Valencian Community	-0.678***	(0.232)
Extremadura	-1.029***	(0.310)
Galicia	-1.178***	(0.348)
Murcia	-0.247	(0.239)
Navarra	-0.518**	(0.253)
Basque Country	-0.749**	(0.352)
La Rioja	-0.442	(0.304)
Internal mobility (reference: never moved)		
Moved once	0.768***	(0.153)
More than once	1.050***	(0.167)
Activity before migration		
Unemployed at origin	0.251	(0.205)
Student at origin	0.151	(0.226)
Last occupation in the origin country (reference: unskilled worker)		
Manager	-0.405	(0.363)
Professional	-0.501*	(0.267)
Paraprofessional	-0.511**	(0.252)
Skilled workers	-0.400	(0.276)
Never worked	-1.272***	(0.253)
Constant	-4.090***	(0.841)
Observations		7,377
Pseudo R2		0.404

Standard error s in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 5 Marginal effects

	Keep job (1)	Different job (2)	Unemployed (3)	Inactive (4)
Independent interest variables				
Close ties (CT)	-0.089***	0.052*	0.003	0.034**
Network job (NJ)	-0.065*	-0.001	0.046**	0.019
CT*NJ	0.086**	-0.033	-0.013	-0.040**
Network size (NS)	0.218**	-0.127	-0.039	-0.052
Weak ties (WT)	-0.138***	0.075	0.053	0.009
WT*years	0.024**	-0.008	-0.015*	-0.001
Time before finding the first job (less one month)	0.119***	-0.053	-0.032	-0.035*
Time before finding the first job (less one month)*NJ	-0.094**	0.050	0.013	0.031
Other independent variables				
Female	-0.008	-0.124***	0.036***	0.096***
Age	0.004	0.000	-0.002	-0.003
Age^2	-0.000	-0.000	0.000	0.000
Married	0.002	-0.005	-0.004	0.007
No.of children origin	-0.013	0.011	0.011**	-0.009*
No. of children Spain	-0.005	-0.007	0.002	0.009***
Residence authorization	-0.008	0.052***	-0.037***	-0.007
Years since arrival (years)	-0.048***	0.040***	0.003	0.005***
Educational attainment (Reference: primary level or less)				
Secondary level	-0.032	0.042*	-0.009	-0.001
Terciary level	0.009	0.029	-0.034**	-0.003
Spanish language	-0.006	0.026	-0.021	0.001
Region of origin (Reference: Western Europe)				
Eastern Europe	-0.086***	0.068*	0.025	-0.007
Latin America	-0.082***	0.084**	0.002	-0.005
North Africa	-0.070*	-0.008	0.068***	0.009
Asia	0.098	0.080	-0.183***	0.005
Rest of the world	-0.045	-0.021	0.065**	0.000
Region of destination (Reference: Madrid)				
Andalucía	0.058**	-0.062*	0.007	-0.003
Aragon	0.010	-0.029	0.003	0.016
Asturias	0.094**	-0.117***	0.031	-0.008
Balears	0.044	-0.093***	0.036*	0.013
Canarias	0.084**	-0.068*	-0.014	-0.001
Cantabria	-0.037	-0.007	0.043	0.001
Castilla Leon	0.050	-0.065*	0.003	0.011
Castilla la Mancha	0.062*	-0.095***	0.035*	-0.002
Catalonia	0.060**	-0.040	-0.002	-0.019
Valencian Community	0.023	-0.040	0.027	-0.010
Extremadura	0.055	-0.104*	0.039	0.010
Galicia	0.075*	-0.145***	0.039	0.032**
Murcia	0.011	-0.057*	0.032*	0.013
Navarra	0.011	-0.008	0.018	-0.021
Basque Country	0.007	-0.051	0.043*	0.002
La Rioja	0.034	-0.014	-0.021	0.001

Table 5 Marginal effects (cont.)

	Keep job (1)	Different job (2)	Unemployed (3)	Inactive (4)
Mobility (Reference: never moved)				
1. Moved once	-0.125***	0.129***	-0.002	-0.001
2. More than one	-0.230***	0.206***	0.017	0.008
First occupation (Reference: unskilled occupation)				
Manager	0.283***	-0.271***	-0.019	0.007
Professional	0.010	-0.039	0.003	0.026
Paraprofessional	-0.048**	0.044*	-0.001	0.005
Skilled workers	0.080***	-0.059**	-0.025	0.004
Sector of activity (Reference: Agriculture)				
Industry	0.168***	-0.173***	-0.003	0.007
Construction	0.158***	-0.183***	0.017	0.009
Trade	0.198***	-0.214***	0.020	-0.004
Hotel sector	0.135***	-0.189***	0.030	0.024*
Transportation	0.165***	-0.138***	-0.021	-0.006
Firm services	0.243***	-0.224***	-0.003	-0.016
Education- Health	0.291***	-0.225***	-0.063**	-0.004
Household activities	0.183***	-0.124***	-0.041**	-0.018
Public administration	0.331***	-0.490***	0.226***	-0.067
Mill's ratio	-0.009	-0.018	0.011	0.016
Activity before migration				
Unemployed at origin	-0.019	-0.031	0.059***	-0.009
Student at origin	-0.089***	0.034	0.039**	0.017
Last occupation in the origin country (reference: unskilled worker)				
Manager	-0.128***	0.100**	0.027	0.001
Professional	-0.051*	0.063**	-0.005	-0.007
Paraprofessional	-0.048*	0.059**	-0.003	-0.008
Skilled workers	-0.030	0.061**	-0.023	-0.009
Never worked at origin	0.086***	-0.073**	0.001	-0.014
Remittances	-0.027*	0.058***	-0.006	-0.025***
Observations				6,432

* p<0.1, ** p<0.05, *** p<0.01

Table 6 Robustness checks Marginal effects by gender

	Women				Men			
	Keep job	Different job	Unemployed	Inactive	Keep job	Different job	Unemployed	Inactive
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Close ties (CT)	-0.110**	0.026*	0.021	0.062**	-0.098**	0.093**	-0.003	0.008
Network job (NJ)	-0.068*	-0.014	0.032	0.051	-0.057*	0.008	0.056**	-0.006
CT*NJ	0.067*	0.008	0.006	-0.081**	0.102*	-0.065	-0.031	-0.006
Network size (NS)	0.122**	-0.069	0.047	-0.101	0.283**	-0.164	-0.114	-0.005
Weak ties (WT)	-0.138*	0.112	0.072	-0.046	-0.112	0.059	0.034	0.018*
WT*years	0.017	-0.009	-0.016	0.007	0.024	-0.009	-0.015	-0.001
Time before finding the first job (less one month)	0.013	0.065	-0.041	-0.037	0.253***	-0.019	0.013	-0.246***
Time before finding the first job (less one month)*NJ	0.004	-0.044	0.019	0.021	-0.205***	-0.016	-0.030	0.250***
Observations	3429	3429	3429	3429	3003	3003	3003	3003

*Other controls used are the same as in Table 5.

Table 7 Robustness checks Marginal effects No skills variables

	All Controls				No controls at all			
	Keep job	Different job	Unemployed	Inactive	Keep job	Different job	Unemployed	Inactive
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Independent interest variables								
Close ties (CT)	-0.089***	0.052*	0.003	0.034**	-0.085***	0.052*	-0.005	0.038**
Network job (NJ)	-0.065*	-0.001	0.046**	0.019	-0.058	-0.003	0.041*	0.020
CT*NJ	0.086**	-0.033	-0.013	-0.040**	0.082**	-0.036	-0.005	-0.041**
Network size (NS)	0.218**	-0.127	-0.039	-0.052	0.241***	-0.146	-0.045	-0.049
Weak ties (WT)	-0.138***	0.075	0.053	0.009	-0.141***	0.072	0.059*	0.010
WT*years	0.024**	-0.008	-0.015*	-0.001	0.024**	-0.007	-0.016**	-0.001
Time before finding the first job (less one month)	0.119***	-0.053	-0.032	-0.035*	0.121***	-0.051	-0.034	-0.036*
Time before finding the first job (less one month)*NJ	-0.094**	0.050	0.013	0.031	-0.099**	0.054	0.012	0.033
Observations	6432				6432			

* p<0.1, ** p<0.05, *** p<0.01

**Other controls used are the same as in Table 5, but excluding skill variables (educational level, proficiency in the Spanish language, labor status before migration, and last occupation in the country of birth).

Table 8 Robustness checks Marginal effects (including motives for migrating)

	Keep job (1)	Different job (2)	Unemployed (3)	Inactive (4)
Independent interest variables				
Close ties (CT)	-0.091***	0.057*	0.000	0.034**
Network job (NJ)	-0.063*	-0.002	0.046**	0.020
CT*NJ	0.084**	-0.030	-0.014	-0.040**
Network size (NS)	0.229**	-0.143	-0.042	-0.044
Weak ties (WT)	-0.153***	0.073	0.071*	0.009
WT*years	0.023**	-0.007	-0.016**	-0.001
Time before finding the first job (less one month)	0.116***	-0.051	-0.032	-0.034*
Time before finding the first job (less one month)*NJ	-0.094**	0.047	0.013	0.034
Observations	6432	6432	6432	6432

* p<0.1, ** p<0.05, *** p<0.01

**Other controls used are the same as in Table 5 and adding motives for migration.

Table 9 Test for normality of the residuals

Lagrange Multiplier Test for Normality after Probit	
Women	*Men*
Chi2(2) = 5.1442	Chi2(2) = 2.7341
Prob > chi2 = 0.0764	Prob > chi2 = 0.2549
Ho: Normality	
Ha: No Normality	

Table 10 Wage regression. Women versus men

Dependent variable: ln(wages per hour)

	QR 25	**QR50**	**QR75**	**OLS**
Women				
Network job (NJ)	-0.037*	-0.034*	-0.048	-0.055*
Close ties (CT)	0.015	-0.002	-0.031	-0.012
Network size (NS)	-0.147	-0.194**	0.094	-0.087
Weak ties (WT)	0.074	-0.032	0.024	0.064
Observations	912	912	912	912
Men				
Network job (NJ)	-0.113***	-0.108***	-0.117***	-0.235***
Close ties (CT)	-0.084***	-0.097***	-0.115***	-0.210***
CT*NJ	0.034*	0.078***	0.087***	0.195***
Network size (NS)	-0.009	-0.195***	-0.105**	0.126
Weak ties (WT)	-0.080***	-0.080***	-0.031	-0.096
Observations	862	862	862	862

* p<0.1, ** p<0.05, *** p<0.01

**Tables A.8 and A.9 in the Appendix present the wage regressions for full specifications for women and men respectively.

APPENDIX

Table A.1 Database elaboration

	Dropped observations	Total
Total sample		**15,441**
Missing Age/ not recorded	41	
Missing Year of arrival/ not recorded	212	
Missing Years of residence/ not recorded	1	
Subtotal		15,187
Subsample - Data restricted to:		
Year of arrival>1996	5,226	
Age between 16 and 65 years	242	
Age at arrival (between 14 and 56 years)	411	
Not finish studies in Spain	595	
Missings	241	
Country before migrarion: country of birth	1,095	
Inactives (retired) / Missings	241	
Subtotal		**7,377**
Without labour experience in Spain	945	
Final Subsample		**6,432**

Table A.2 Definition of independent variables

Female	1 if respondent is a woman; 0 otherwise
Man	1 if respondent is a man; 0 otherwise
Age	Age in years
Age^2	Age square
Years since arrival	Years
Married	1 if the respondent is married; 0 otherwise
Number of children	Number of daughters and sons
Residence authorization	1 if the respondent declares having any of the following documents: Permanent residency authorisation; temporary residency authorisation, EU residence permit (except in the case of Romanian and Bulgarian workers who, despite being EU citizens could not become legally contracted workers in Spain temporarily at the time of the survey); refugee status or assylum application. This cathegory also includes immigrants whose nationailty is Spanish, from other EU member state (excluding Bulgaria and Romania) or from non-EU members of thr Free Trade Association (i.e., Lichtenstein, Iceland, Switzerland and Norway); 0 otherwise
Education level attained (dummies variables)	
Primary level	1 if the respondent has primary level attained or less; 0 otherwise
Secondary level	1 if the respondent has secondary level complete or incomeplete; 0 otherwise
Tertiary level	1 if the respondent has tertiary level complete or incomeplete; 0 otherwise
Language	
Speaks spanish	1 if respondent declares having spanish as her mother tongue or, if she states can speak Spanish 'well' or 'very well'; 0 otherwise
Region of origin	
Western Europe	1 if country of birth is in Western Europe; 0 otherwise
Eastern Europe	1 if country of birth is in Eastern Europe; 0 otherwise
Latin America	1 if country of birth is in Latin America; 0 otherwise
North Africa	1 if country of birth is in North Africa; 0 otherwise
Asia	1 if country of birth is in Asia; 0 otherwise
Rest of the world	1 if country of birth is in Oceania, rest of Africa, ; 0 otherwise

Table A.2 (Cont.)

Migration between municipalities. Frecuency (%)	
1. Never moved	1 if respondent declares have lived in the same municipality since arrival; 0 otherwise
2. Moved once	1 if respondent declares have lived in two different municipalities; 0 otherwise
3. More than one	1 if respondent declares have lived in more than two different municipalities; 0 otherwise
Motives for migration	
Labor	1 if respondent declares moved because being unemployed in the country of origin or declares looking for a better job; 0 otherwise
Family regrouping	1 if respondent declares family regrouping; 0 otherwise
Social networks	
Contacts at arrival (Close ties)	1 if respondent has contacts at arrival; 0 otherwise
Social participation in organizations exclusive for immigrants	1 if respondent participates in: immigrant assistance organizations specifically to foreigners, associations and sports clubs specifically targeting foreigners, educational and cultural groups specifically targeting foreigners, religious organizations and groups specifically targeting foreigners, other groups specifically targeting foreigners; 0 otherwise
Social participation in mixed organizations	1 if respondent participates in: NGO´s Political organizations, unions, or neighborhood activities, Religious groups, Sport clubs, educational and cultural groups, Other social groups; 0 otherwise
Migrant proportion	Proportion of immigrants of the same country of birht living in the same Autonomous Community on the total immigrant population in the Autonomous Community (%)
Network job	1 if respondent has found the job through family and friends; 0 otherwise
Formal job	1 if respondent has found the job through State and private employment agencies, newspapers´ advertisements, union hiring halls as well as school and college placement services; 0 otherwise

Table A.2 (Cont.)

Sector of activity	
Agriculture	1 if respondent' first job is in: Agriculture, Hunting, and Forestry Fishing, Minning; 0 otherwise
Industry	1 if respondent' first job is in: Manufacture industries, Production and distribution of electricity, gas and water; 0 otherwise
Construction	1 if respondent' first job is in Construction; 0 otherwise
Trade	1 if respondent' first job is in: Trade, repair of motor vehicles and motorcycles and personal articles and electronic products for household; 0 otherwise
Hotel sector	1 if respondent' first job is in: Hotel sector; 0 otherwise
Transportation	1 if respondent' first job is in: Transport, storage and communications; 0 otherwise
Firm services	1 if respondent' first job is in: Financial intermediation Real estate, renting and business services; 0 otherwise
Education- Health	1 if respondent' first job is in: Education, Health and veterinary activities, social service, Other social and community services, personal services; 0 otherwise
Household activities	1 if respondent' first job is in: Household activities; 0 otherwise
Public administration	1 if respondent' first job is in: Public administration, defense and compulsory social security; 0 otherwise

Table A.2 (Cont.)

Occupation	
Manager	1 if respondent declares: Management of companies and public administrations; 0 otherwise
Professional	1 if respondent declares: Technical and scientific professionals and intellectuals, Technicians and associate professionals; 0 otherwise
Paraprofessional	1 if respondent declares: Administrative workers, Workers in catering services, personal services, protection services, and comercial salers; 0 otherwise
Skilled workers	1 if respondent declares: Qualified workers in fishing and agriculture activities. Craftsmen and skilled manufacturing, construction, and mining, except plant and machinery operators. 0 otherwise
Unskilled workers	1 if respondent declares: Unskilled occupation; 0 otherwise
Time before finding the first job	Dummy variable equal to 1 if respondent declares spending less than a month before finding the first job; 0 otherwise.
Remmitances	Dummy variable equal to 1 if respondent declares sending remmitances to the country of brith; 0 otherwise.

Notes: 1. Weak ties refer to immigrants participating in mixed organizations.
2. Migrant proportion is the network size.

Table A.3 Descriptive statistics. Socio-demographic characteristics by region of origin

	Western Europe	Latin America	Eastern Europe	North Africa	Asia	Rest of the world	Total
Variables							
Female	0.47	0.59	0.57	0.28	0.35	0.27	0.53
Age	36	34	33	33	33	33	34
Year of arrival	2002	2002	2002	2001	2001	2001	2002
Years since arrival	4	4	4	5	5	5	4
Married	0.37	0.47	0.54	0.59	0.59	0.56	0.50
Number of children	0.89	1.49	1.02	0.98	1.09	1.15	1.25
No. children origin	0.56	0.43	0.76	0.36	0.61	1.21	0.41
No. children Spain	1.25	1.19	1.28	1.98	1.32	1.03	0.86
Residence authorization	1.00	0.74	0.70	0.86	0.90	0.85	0.76
Educational level attained (dummies variables)							
Primary level	0.14	0.19	0.13	0.28	0.31	0.32	0.18
Secondary level	0.50	0.58	0.67	0.39	0.39	0.39	0.57
Tertiary level	0.36	0.23	0.20	0.33	0.31	0.29	0.24
Speaks spanish	0.64	0.98	0.64	0.55	0.39	0.50	0.80
Migration between municipalities. Frecuency (%)							
1. Never moved	40.69	21.12	23.57	27.23	32.12	25.41	24.12
2. Moved once	29.65	39.99	37.59	30.99	28.47	34.05	37.31
3. More than one	29.65	38.88	29.65	41.78	39.42	40.54	38.56
Motives for migration [1]							
Labor motives	0.13	0.51	0.68	0.28	0.22	0.50	0.64
Family regrouping	0.40	0.31	0.27	0.39	0.48	0.30	0.27
Social networks							
Contacts at arrival (Close ties)	0.75	0.87	0.82	0.78	0.76	0.58	0.83
Social participation (exclusive for immigrants)	0.04	0.06	0.05	0.07	0.14	0.17	0.06
Social participation (mixed organization)	0.18	0.11	0.07	0.06	0.10	0.14	0.10
Frecuency (region of birth) Subsample (%)	7.18	51.82	26.06	9.93	2.13	2.88	100.00
Observations	3,644	6,059	2,386	2,018	437	643	6432

Table A.4 Occupational mobility between actual occupation and last occupation in the country of origin

Region	occupation in the	Actual occupation in Spain					
		Manager	Professional	raprofessio	alified work	skilled work	Total
Total sample	Manager	**11.2**	9.1	27.5	16.7	35.5	100
	Professional	1.7	**20.0**	30.5	11.2	36.7	100
	Paraprofessic	0.9	3.5	**38.7**	10.0	46.9	100
	Qualified wor	0.1	1.2	11.4	**39.5**	47.7	100
	Unskilled wo	0.1	1.0	16.8	13.0	**69.1**	100
	Total	1.3	5.9	25.3	19.7	47.8	100
Western Europe	Manager	**59.5**	10.8	18.9	5.4	5.4	100
	Professional	6.9	**60.3**	25.2	3.1	4.6	100
	Paraprofessic	7.0	17.4	**55.7**	4.4	15.7	100
	Qualified wor	1.3	6.7	17.3	**62.7**	12.0	100
	Unskilled wo	0.0	13.3	23.3	16.7	**46.7**	100
	Total	10.3	28.9	32.0	16.2	12.6	100
Latin America	Manager	**3.4**	9.0	32.8	18.1	36.7	100
	Professional	1.0	**14.8**	36.3	12.0	35.8	100
	Paraprofessic	0.4	2.4	**43.5**	9.3	44.4	100
	Qualified wor	0.0	1.4	12.9	**41.3**	44.4	100
	Unskilled wo	0.3	0.3	23.6	12.5	**63.4**	100
	Total	0.6	5.0	31.5	18.2	44.7	100
Eastern Europe	Manager	**8.1**	8.1	16.2	13.5	54.1	100
	Professional	0.5	**8.4**	20.3	10.9	59.9	100
	Paraprofessic	0.8	1.1	**24.6**	12.2	61.4	100
	Qualified wor	0.2	0.4	9.4	**39.0**	51.2	100
	Unskilled wo	0.0	0.6	9.5	11.8	**78.1**	100
	Total	0.6	2.0	15.4	23.3	58.7	100
North Africa	Manager	**0.0**	5.9	23.5	23.5	47.1	100
	Professional	0.0	**10.8**	24.3	24.3	40.5	100
	Paraprofessic	0.0	5.2	**18.2**	15.6	61.0	100
	Qualified wor	0.0	0.6	7.8	**28.5**	63.1	100
	Unskilled wo	0.0	0.8	5.7	14.5	**79.0**	100
	Total	0.0	2.5	11.1	21.7	64.8	100
Asia	Manager	**0.0**	0.0	0.0	0.0	0.0	100
	Professional	6.3	**18.8**	31.3	12.5	31.3	100
	Paraprofessic	0.0	3.3	**63.3**	10.0	23.3	100
	Qualified wor	0.0	4.0	36.0	**16.0**	44.0	100
	Unskilled wo	0.0	5.9	11.8	5.9	**76.5**	100
	Total	1.1	6.8	39.8	11.4	40.9	100
Rest of the world	Manager	0.0	12.5	12.5	37.5	37.5	100
	Professional	0.0	50.0	0.0	16.7	33.3	100
	Paraprofessic	0.0	7.7	23.1	12.8	56.4	100
	Qualified wor	0.0	0.0	7.1	40.5	52.4	100
	Unskilled wo	0.0	0.0	12.8	18.0	69.2	100
	Total	0.0	8.9	12.3	24.0	54.8	100

Table A.5 Multinomial regression (base outcome: employed in a different job)

	Keep job	Unemployed	Inactive
Ommited: Employed in a different job			
Key independent variables			
Close ties (CT)	-0.478***	-0.081	0.591*
Network job (NJ)	-0.258	0.517*	0.402
CT*NJ	0.422*	-0.082	-0.751*
Migrant proportion	1.165**	-0.152	-0.847
Weak ties (WT)	-0.731**	0.437	0.002
WT*years	0.114*	-0.149	0.011
Time before finding the first job (less one month)	0.608***	-0.247	-0.548
Time before finding the first job (less one month)*NJ	-0.498**	0.036	0.488
Remmitances	-0.216**	-0.220*	-0.538***
Control variables			
Female	0.238**	0.713***	2.236***
Age	0.017	-0.019	-0.068
Age^2	0.000	0.000	0.001
Married	0.006	-0.019	0.089
No. Children in Spain	-0.006	0.044	0.203***
No. Children in origin	-0.076	0.091	-0.204*
Years since arrival (years)	-0.291***	-0.059*	-0.009
Residence authorization	-0.154	-0.525***	-0.284*
Educational attained (Reference: primary level or less)			
Secondary level	-0.227*	-0.194	-0.131
Terciary level	-0.032	-0.447**	-0.142
Spanish language	-0.086	-0.291*	-0.052
Region of origin (Reference: Western Europe)			
Eastern Europe	-0.516**	0.133	-0.363
Latin America	-0.531***	-0.162	-0.323
North Africa	-0.270	0.780**	0.168
Asia	0.199	-2.228***	-0.114
Rest of the world	-0.121	0.761*	0.120
Region of destination (Reference: Madrid)			
Andalucía	0.366**	0.225	0.076
Aragon	0.097	0.097	0.382
Asturias	0.648***	0.607	0.144
Balears	0.386**	0.610**	0.473*
Canarias	0.486**	-0.005	0.115
Cantabria	-0.133	0.478	-0.005
Castilla Leon	0.343	0.194	0.393
Castilla la Mancha	0.460**	0.612**	0.157
Catalonia	0.330**	0.065	-0.309
Valencian Community	0.183	0.393*	-0.107
Extremadura	0.447	0.677*	0.406
Galicia	0.627**	0.769**	0.971***
Murcia	0.175	0.489**	0.435
Navarra	0.058	0.220	-0.420
Basque Country	0.137	0.593**	0.130
La Rioja	0.165	-0.196	0.035

Table A.5 Multinomial regression (cont.)

	Keep job	Unemployed	Inactive
Mobility (Reference: never moved)			
1. Moved once	-0.804***	-0.312*	-0.336
2. More than one	-1.400***	-0.273	-0.317
First occupation (Reference: unskilled occupation)			
Manager	1.742***	0.402	0.736
Professional	0.124	0.138	0.581
Paraprofessional	-0.295**	-0.108	-0.018
Skiled workers	0.456***	-0.143	0.200
Activity sector (Reference: Agriculture)			
Industry	1.069***	0.361	0.562
Construction	1.049***	0.607**	0.603
Trade	1.284***	0.712**	0.406
Hotel sector	0.975***	0.772***	0.972***
Transportation	0.974***	0.074	0.203
Firm services	1.492***	0.476	0.200
Education- Health	1.686***	-0.182	0.453
Household activities	1.027***	-0.175	-0.057
Public administration	2.447***	3.629***	-0.332
Mill's ratio	0.018	0.201	0.377
Activity before migration			
Unemployed at origin	-0.001	0.716***	-0.085
Student at origin	-0.429**	0.348	0.311
Last occupation in the origin country (reference: unskilled worker)			
Manager	-0.742***	0.069	-0.219
Professional	-0.348**	-0.192	-0.296
Paraprofessional	-0.329**	-0.170	-0.308
Skilled workers	-0.256*	-0.392*	-0.315
Never worked	0.511***	0.179	-0.141
Remittances	-0.246**	-0.208*	-0.653***
Constant	0.758	-1.071	-1.687
Observations			6432
Pseudo R2			0.159

* p<0.1, ** p<0.05, *** p<0.01

Table A.6 Probability of keeping the first job. Semiparametric model. Women

	Coef	SE
Key independent variables		
Network job (NJ)	-0.575***	(0.155)
Close ties (CT)	-0.201*	(0.134)
CT*NJ	0.184	(0.146)
Migrant proportion	4.372***	(0.662)
Weak ties (WT)	-0.011	(0.119)
Other controls		
Age	0.185***	(0.039)
Age^2	-0.001***	(0.000)
Married	-0.110	(0.085)
No. Children origin	-0.145**	(0.062)
No. Children Spain	-0.166***	(0.052)
Residence authorization	-0.374***	(0.082)
Years since arrival (years)	-1.678***	(0.254)
Educational level attained (Reference: primary level or less)		
Secondary level	0.775***	(0.154)
Terciary level	0.720***	(0.150)
Spanish language		
Region of origin (Reference: Western Europe)		
Eastern Europe	-0.524***	(0.119)
Latin America	-0.708***	(0.115)
North Africa	-2.16***	(0.157)
Asia	4.123***	(0.226)
Rest of the world	5.443***	(0.244)
Region of destination (reference: Madrid)		
Andalucía	0.154	(0.114)
Aragon	0.188	(0.137)
Asturias	0.305*	(0.158)
Balears	0.088	(0.112)
Canarias	0.016	(0.159)
Cantabria	-0.062	(0.159)
Castilla Leon	0.055	(0.139)
Castilla la Mancha	0.044	(0.133)
Catalonia	0.175*	(0.100)
Valencian Community	-0.007	(0.112)
Extremadura	0.023	(0.179)
Galicia	0.190	(0.158)
Murcia	-0.151	(0.119)
Navarra	-0.016	(0.115)
Basque Country	-0.015	(0.141)
La Rioja	0.064	(0.134)

Table A.6 Probability of keeping the first job. Semiparametric model. Women (cont.)

	Coef	SE
First occupation (Reference: unskilled occupation)		
Manager	13.682***	(2.040)
Professional	0.014	(0.136)
Paraprofessional	-1.230***	(0.226)
Skilled workers	-0.945***	(0.261)
Activity sector (Reference: Agriculture)		
Industry	-1.211***	(0.235)
Construction	-1.689***	(0.321)
Trade	2.119***	(0.414)
Hotel sector	0.651***	(0.133)
Transportation	1.719***	(0.443)
Firm services	2.124***	(0.350)
Education- Health	2.089***	(0.320)
Household activities	1.539***	(0.280)
Public administration	0.950**	(0.420)
Activity before migration		
Unemployed at origin	0.426***	(0.100)
Student at origin	-0.284***	(0.106)
Last occupation in the origin country (reference: unskilled worker)		
Manager	-4.502***	(0.499)
Professional	-0.990***	(0.146)
Paraprofessional	-1.436***	(0.179)
Skilled workers	-1.524***	(0.201)
Never worked at origin	0.706***	(0.165)
Remittances	-1.268***	(0.195)
Observations	3429	

* p<0.1, ** p<0.05, *** p<0.01

Table A.7 Probability of keeping the first job. Probit model. Men

	Coef	SE
Key independent variables		
Network job (NJ)	-0.240*	(0.125)
Close ties (CT)	-0.209*	(0.111)
CT*NJ	0.201	(0.143)
Migrant proportion	0.618*	(0.323)
Weak ties (WT)	-0.448**	(0.194)
WT*years	0.086**	(0.039)
Other controls		
Age	-0.013	(0.023)
Age^2	0.000	(0.000)
Married	0.012	(0.063)
Number of children	-0.015	(0.028)
Residence authorization	0.147**	(0.071)
Years since arrival (years)	-0.178***	(0.016)
Maximum educational level attained (Reference: primary level or less)		
Secondary level	-0.159**	(0.073)
Terciary level	-0.086	(0.088)
Spanish language	-0.072	(0.076)
Region of origin (Reference: Western Europe)		
Eastern Europe	-0.234*	(0.128)
Latin America	-0.204*	(0.119)
North Africa	-0.041	(0.137)
Asia	0.254	(0.193)
Rest of the world	-0.161	(0.174)
Region of destination (reference: Madrid)		
Andalucía	0.263*	(0.135)
Aragon	-0.167	(0.157)
Asturias	0.279	(0.214)
Balears	0.108	(0.133)
Canarias	0.445***	(0.154)
Cantabria	-0.172	(0.199)
Castilla Leon	0.162	(0.158)
Castilla la Mancha	0.306**	(0.138)
Catalonia	0.202*	(0.110)
Valencian Community	0.128	(0.122)
Extremadura	0.069	(0.224)
Galicia	0.055	(0.209)
Murcia	0.060	(0.125)
Navarra	0.060	(0.129)
Basque Country	0.043	(0.172)
La Rioja	0.071	(0.160)

Table A.7 Probability of keeping the first job. Probit model. Men

	Coef	SE
First occupation (Reference: unskilled occupation)		
Manager	1.574***	(0.251)
Professional	0.663***	(0.145)
Paraprofessional	0.457***	(0.137)
Skilled workers	0.447***	(0.071)
Activity sector (Reference: Agriculture)		
Industry	0.523***	(0.109)
Construction	0.474***	(0.084)
Trade	0.427***	(0.124)
Hotel sector	-0.213	(0.165)
Transportation	0.164	(0.162)
Firm services	0.339**	(0.156)
Education- Health	0.600***	(0.174)
Household activities	-0.569*	(0.306)
Public administration	1.313***	(0.420)
Mobility (Reference: never moved)		
1. Moved once	-0.510***	(0.070)
2. More than one	-0.880***	(0.074)
Time before finding the first job (less one month)	0.258***	(0.071)
Activity before migration		
Unemployed at origin	-0.090	(0.081)
Student at origin	-0.317**	(0.125)
Last occupation in the origin country (reference: unskilled worker)		
Manager	-0.341**	(0.155)
Professional	-0.083	(0.117)
Paraprofessional	-0.132	(0.103)
Skilled workers	-0.004	(0.087)
Never worked at origin	0.273**	(0.129)
Remittances	-0.163***	(0.061)
Constant	0.653	(0.457)
Observations		3003
Pseudo R2		0.224

* p<0.1, ** p<0.05, *** p<0.01

Table A.8 Wage regression. Women

	QR 25	QR50	QR75	OLS
Independent interest variables				
Network job (NJ)	-0.037*	-0.034*	-0.048	-0.055*
Close ties (CT)	0.015	-0.002	-0.031	-0.012
Network size (NS)	-0.147	-0.194**	0.094	-0.087
Weak ties (WT)	0.074	-0.032	0.024	0.064
Time before finding the first job	0.090	0.063***	0.024	0.038
Other independent variables				
Age	-0.008	0.001	-0.002	-0.002
Age^2	0.000	-0.000	-0.000	0.000
Married	-0.025	-0.031*	-0.015	-0.017
No. Children in Spain	0.005	0.006	-0.001	0.011
No. Children in origin	-0.009**	-0.020**	-0.032	-0.029*
Residence authorization	0.102	0.061***	0.085	0.071**
Years since arrival (years)	0.007	0.005	-0.005	0.000
Educational level attained (Reference: primary level or less)				
Secondary level	-0.075	-0.077***	-0.084	-0.069
Terciary level	0.080	0.060**	0.093	-0.066
Spanish language	-0.065	-0.006	0.024	0.014
Region of origin (Reference: Western Europe)				
Eastern Europe	-0.021	-0.079**	-0.101	-0.087
Latin America	-0.099	-0.198***	-0.173*	-0.166***
North Africa	-0.176	-0.223***	-0.240*	-0.257***
Asia	-0.179	-0.359***	-0.258	-0.247*
Rest of the world	-0.369	-0.461***	-0.263	-0.366**
Region of destination (Reference: Madrid)				
Andalucía	-0.039	-0.010	0.091	-0.029
Aragon	-0.098	0.020	0.170	0.006
Asturias	0.028	-0.012	0.167	-0.005
Balears	0.176	0.142***	0.169*	0.124*
Canarias	0.105	0.106**	0.152	0.100
Cantabria	-0.078	-0.133**	0.010	-0.136
Castilla Leon	0.008	-0.083*	0.040	-0.053
Castilla la Mancha	-0.074	-0.042	-0.003	-0.150**
Catalonia	0.199	0.166***	0.194**	0.148***
Valencian Community	-0.004	0.015	0.031	-0.071
Extremadura	0.042	-0.069	-0.071	-0.085
Galicia	-0.135	-0.079	-0.010	-0.109
Murcia	0.026	0.005	0.026	-0.016
Navarra	0.154	0.107***	0.153	0.090
Basque Country	0.031	0.116***	0.167	0.060
La Rioja	0.154	0.076*	0.070	0.049

Table A.8 Wage regression. Women (cont.)

	QR 25	QR50	QR75	OLS
First occupation (Reference: unskilled occupation)				
Manager	0.176	0.141**	0.615***	0.311**
Professional	0.198	0.264***	0.406***	0.248***
Paraprofessional	-0.026	-0.020	-0.070	-0.056
Skilled workers	-0.166	-0.177	-0.118	-0.142
Sector of activity (Reference: Agriculture)				
Industry	-0.054	-0.083	-0.027	-0.081
Construction	0.047	-0.037	-0.104	0.002
Trade	0.124	0.007	0.025	0.070
Hotel sector	0.040	-0.031	0.022	0.015
Transportation	-0.027	-0.110	0.280	0.102
Firm services	0.013	-0.039	0.209	0.087
Education- Health	-0.005	-0.021	0.071	0.012
Household activities	-0.173*	-0.166***	-0.028	-0.139*
Public administration	-0.040	-0.123	-0.178	-0.072
Activity before migration				
Unemployed at origin	-0.065	-0.086***	-0.066	-0.044
Student at origin	-0.126	-0.004	0.008	-0.063
Last occupation in the origin country (reference: unskilled worker)				
Manager	-0.041	-0.049	-0.019	-0.026
Professional	0.153	0.154***	0.157*	0.175***
Paraprofessional	0.031	0.075***	0.045	0.077
Skilled workers	-0.016	-0.008	-0.023	0.013
Never worked at origin	0.137	0.119***	0.045	0.092
Mill's ratio	0.020**	0.019**	0.026*	0.015**
Constant	2.961	3.126***	3.265***	3.147***
Observations	912	912	912	912

* p<0.1, ** p<0.05, *** p<0.01

Table A.9 Wage regression. Men

	QR 25	QR50	QR75	OLS
Network job (NJ)	-0.113***	-0.108***	-0.117***	-0.235***
Close ties (CT)	-0.084***	-0.097***	-0.115***	-0.210***
CT*NJ	0.034*	0.078***	0.087***	0.195***
Migrant proportion	-0.009	-0.195***	-0.105**	0.126
Weak ties (WT)	-0.080***	-0.080***	-0.031	-0.096
WT*years	0.009**	0.020***	0.015**	0.038*
Other independent variables				
Age	0.013***	0.019***	0.006*	0.005
Age^2	-0.000***	-0.000***	-0.000	-0.000
Married	0.049***	0.028***	0.034***	0.053*
No. of children Spain	-0.008*	-0.014***	-0.024***	-0.009
No. Children origin	0.010**	0.003***	-0.016***	0.004
Residence authorization	0.055***	0.050***	0.052***	0.092***
Years since arrival (years)	-0.013***	0.013***	-0.002	0.000
Educational level attained (Reference: primary level or less)				
Secondary level	0.008	-0.016***	-0.015	-0.014
Terciary level	0.080***	0.070***	0.043***	0.089**
Spanish language	0.047***	0.065***	0.038***	0.033
Region of origin (Reference: Western Europe)				
Eastern Europe	-0.129***	-0.110***	-0.157***	-0.166***
Latin America	-0.203***	-0.172***	-0.193***	-0.207***
North Africa	-0.239***	-0.255***	-0.297***	-0.260***
Asia	-0.162***	-0.218***	-0.259***	-0.219***
Rest of the world	-0.294***	-0.251***	-0.312***	-0.344***
Region of destination (Reference: Madrid)				
Andalucía	0.083***	0.038***	0.089***	0.092
Aragon	0.014	-0.056***	-0.075***	-0.068
Asturias	0.031	-0.054***	0.021	-0.005
Balears	0.082***	-0.026***	0.068***	0.100
Canarias	0.092***	-0.000***	0.170***	0.164**
Cantabria	-0.204***	-0.187***	-0.137***	-0.260***
Castilla Leon	-0.026	-0.162***	-0.102***	-0.043
Castilla la Mancha	0.038*	0.025***	0.053**	0.035
Catalonia	0.156***	0.094***	0.122***	0.124**
Valencian Community	0.014	0.068***	0.026	0.043
Extremadura	-0.219***	-0.105***	-0.152***	-0.034
Galicia	-0.423***	-0.338***	-0.263***	-0.231**
Murcia	0.103***	0.009***	-0.020	0.016
Navarra	0.132***	0.089***	0.091***	0.089
Basque Country	-0.015	-0.113***	-0.114***	-0.085
La Rioja	0.135***	0.074***	0.096***	0.086

Table A.9 Wage regression. Men (cont.)

	QR 25	QR50	QR75	OLS
First occupation (Reference: unskilled occupation)				
Manager	0.596***	0.398***	0.651***	0.751***
Professional	0.439***	0.313***	0.552***	0.521***
Paraprofessional	0.087***	0.088***	0.110***	0.159**
Skilled workers	0.156***	0.059***	0.127***	0.197***
Sector of activity (Reference: Agriculture)				
Industry	0.082***	0.069***	0.110***	0.177**
Construction	0.181***	0.211***	0.226***	0.238***
Trade	0.029	0.052***	0.147***	0.081
Hotel sector	-0.050**	0.137***	-0.061**	-0.129
Transportation	0.005	0.145***	0.328***	0.201**
Firm services	0.109***	0.123***	0.122***	0.154**
Education- Health	0.030	0.130***	0.216***	0.232**
Household activities	-0.195***	-0.157***	-0.372***	-0.277
Public administration	0.053	-0.031***	0.101	0.012
Mobility (Reference: never moved)				
1. Moved once	-0.040***	-0.024***	-0.094***	-0.142***
2. More than one	-0.077***	-0.030***	-0.144***	-0.225**
Time before finding the first job (less one month)	0.110***	0.070***	0.195***	0.200***
Activity before migration				
Unemployed at origin	-0.144***	-0.150***	-0.135***	-0.116
Student at origin	-0.173***	-0.094***	-0.124***	-0.130
Last occupation in the origin country (reference: unskilled worker)				
Manager	-0.118***	-0.046***	-0.012	-0.112
Professional	0.004	0.097***	0.075***	0.061
Paraprofessional	-0.035	-0.001	-0.003	-0.041
Skilled workers	-0.071***	0.001***	0.011	-0.044
Never worked at origin	0.036*	0.052***	0.149***	0.118*
Mill's ratio	0.201***	0.025***	0.216***	0.402***
Constant	2.705***	2.879***	3.050***	2.859***
Observations	862	862	862	862

* p<0.1, ** p<0.05, *** p<0.01

METHODOLOGICAL APPENDIX

Buchinsky (1998)

Buchinsky (1998) was the first to consider the difficult problem of estimating quantile regression in the presence of sample selection. We summarize this methodology as if follows:

First, the reservation wage equation is considered as follows:

$$y_i^R = x_i'\alpha_0 + v_i \qquad \text{(A.1)}$$

The reservation wage of each individual is a function of her characteristics (x_{1i}) in addition to an idiosyncratic term.

The wage offer (y_i^*) is assumed to be linearly dependent on a set of labor market characteristics (x_{2i}; a subset of x_{1i}) equation (2) in Buchinsky (1998) is:

$$y_i^* = c + x_{2i}'\beta_0 + \varepsilon_i \qquad \text{(A.2)}$$

where β is the vector of slope coefficients and ε_i is the error term.

In what follows, in order to simplify the notation, we omit the *i* subscript.

Equation (A.2) can be rewritten in the QR form considered by Koenker and Bassett (1978) as:

$$y^* = c + x_2'\beta_\theta + u_\theta \;\; 0 \le \theta \le 1 \qquad \text{(A.3)}$$

where $u_\theta \equiv x_2'(\beta_0 - \beta_\theta) + u$. It is assumed that the conditional quantile of y^*, conditional on x_2, satisfies $Q_\theta(y|x_2) = x_2'\beta_\theta$ so that $Q_\theta(u_\theta|x_2) = 0$

Since wage offer is observed only if it exceeds the reservation wage, we have $y = d \cdot y^* = d(x_2'\beta_\theta + u_\theta)$, where $d \equiv I(y^* \ge y^R)$ and *I(.)* is the usual indicator function.

In the presence of this selection mechanism the conditional quantile of the observed wage is given by

$$Q_\theta(y|x_2) = Q_\theta(y^*|x_2, d = 1) = x'\beta_\theta + Q_\theta(u_\theta|x_2, d = 1)$$

and in general $Q_\theta(u_\theta|x_2, d = 1) \neq 0$. Nevertheless, if $Q_\theta(u_\theta|x_2, d = 1)$ is only a function of a known index *f*, then the observed wage equation can be written as

$$y = x_2'\beta_\theta + h_\theta(f) + \varepsilon_\theta \qquad \text{(A.4)}$$

where $h_\theta(f) \equiv Q_\theta(u_\theta|x_1, y^* \geq y^R|x_1)$ and, by construction, $Q(\varepsilon_{\theta i}|x_1, d = 1) = 0$

The probability of working can be written as

$$PW \equiv \Pr(y^* \geq y^R|x_1) = \Pr(\varepsilon \leq -x_i'\alpha_0 + x_2'\beta_\theta)|x_1) \qquad \text{(A.5)}$$

In order to ensure that PW is only a function of f and the representation of the equation (A.4) holds, two additional assumptions are made by Buchinsky (1998). First, assumes that $w \equiv (v, u)'$ has a continuous density; and second dependence of w and x_1: $g_w(.|x_1) = g_w(.|f(x_1; \gamma_0)$

These assumptions on the joint distribution of these unobservables, both unconditionally and conditional on x_1, that justifies the single-index representation.[27] These assumptions, while sufficient for the single-index representation, does not reveal the functional form of *h(.)*. Buchinsky (1998) suggests using the following series estimator

$$\hat{h}_\theta(x_1\gamma_0) = \delta_0(\theta) + \delta_1(\theta)\lambda(x_1\gamma_0) + \delta_2(\theta)\lambda(x_1\gamma_0)^2 + \cdots,$$

where $\lambda(.)$ is the inverse Mills ratio defined as $\lambda = \frac{\phi(.)}{\Phi(.)}$, while $\phi(.)$ and $\Phi(.)$ are the density and the c.d.f. of a standard normal variable, respectively. Thus, for appropriate values of the δ's $\hat{h}_\theta(x_1\gamma_0) \to h_\theta(x_1\gamma_0)$ as the number of terms goes to infinity.

Finally, in order to estimate γ, we use the semi-parametric estimator suggested by Klein and Spady (1993).

Klein and Spady (1993)

First, the dichotomous realization of the participation equation is specified:

$$D_i = \begin{cases} 1 & if\ v_i \geq x_{1i}'\gamma_0 \\ 0 & otherwise \end{cases}$$

Taking the conditional expectation of *D* conditional on x_1, we obtain

$$E(D_i|x_{1i}) = 1 \times \Pr(v_i \geq -x_{1i}'\gamma_0|x_{1i}) = \Pr(v_i < x_{1i}'\gamma_0|x_{1i}) = F_{v|x}(x_1'\gamma_0)$$

So, it is obtained:

[27] Assumptions C and E in Buchinsky (1998) pp.4.

$$E(D_i|x_{1i}) = F_{v|x}(x_1'\gamma_0) \qquad (1)$$

Klein and Spady (1993) proposes a semi-parametric estimation γ_0 in which assume that the model satisfies the index restriction

$E(D_i|x_{1i}) = E(x_1'\gamma_0)$ (2)

Equations (1) and (2) implies that:

$$E(D_i|x_{1i}) = F_{v|x}(x_1'\gamma_0) = G(x_1'\gamma_0)$$

Where G is an unknown function whose range in contained in [0,1].

γ_0 is computed by maximizing the equation (2) replacing the true but unknown distribution $F_{v|x}(.)$ by $G_n(.)$ that is a non parametric estimated of the function $G(.)$ which is a kernel estimate giveb by:

$$G_n(\vartheta_i) = \frac{\sum_{j=1} Z_j K[(\vartheta_i - \vartheta_j)/h_n]}{\sum_{j=1} K[(\vartheta_i - \vartheta_j)/h_n]}$$

The semi-parametric estimator of γ_0, γ_0^{SP}, is obtained by maximizing the quasi likelihood function given by

$$LnL = \sum_{i=1}^{N} Z_i G_n\,(x_1'\gamma_0^{SP}) + (1 - Z_i)(1 - G_n(x_1'\gamma_0^{SP}))$$

γ_0^{SP} is consistent, asymptotically normally distributed and achieves the semiparametric efficiency bound. In addition, in Monte Carlo experiments γ_0^{SP} performed well relative to probit, and can, in models sufficiently perturbed from the usual probit specification dominate the probit estimator (Klein and Spady, 1993).

Essay 2
The Long-Term Effect of Inequality on Entrepreneurship and Job Creation

The Long-Term Effect of Inequality on Entrepreneurship and Job Creation[*]

Abstract

We assess the extent to which historical levels of inequality affect the probability of businesses being created, surviving and of these creating jobs overtime. For this end, we build a pseudo-panel of entrepreneurs across 48 countries using the Global Entrepreneurship Monitor Survey over 2001-2009. We complement this pseudo-panel with historical data of income distribution and current indicators of business regulation. We find that countries with higher levels of inequality in the 1700s and 1800s, their businesses today are more likely to die young and create fewer jobs. Our evidence support theories that argue initial wealth distribution influences development path, thereby with important policy implications for wealth distribution.

[*] This essay has been co-written with Roxana Gutiérrez Romero (Departament d'Economia Aplicada – Universitat Autònima de Barcelona).

2.1 Introduction

To foster development it is crucial to understand the reasons why entrepreneurship struggles or flourishes. Whilst the literature has developed complex theoretical models on what might drive entrepreneurship over time, these theories have not been empirically tested (Naudé, 2010). Instead, the empirical literature has focused on analyzing separately the individual, economic or institutional factors that might affect entrepreneurship.

We contribute to the literature by testing empirically one of the main mechanisms highlighted in the theoretical literature that suggest affect entrepreneurship over time. The theoretical occupational choice model proposed by Banerjee and Newman (1993) guides our work. This model suggests that initial conditions, understood as the historical distribution of wealth, can be detrimental for economic development if credit constraints are such that they prevent poor individuals from investing in profitable entrepreneurial activities. The model shows that a country can converge to a different family of equilibriums, depending on the initial wealth distribution. Countries that start with a high proportion of non-credit constrained people will grow over time aided by a high share of people being able to start-up business, of these surviving over time and with an active labor market paying high salaries. A contrasting equilibrium could be reached if a country starts with a high proportion of credit constrained people. In this case, only a small share of the population will be able to start-up new businesses, whilst the rest will remain as workers, earning low wages over time, in which there is (almost) only self-employment at small scale.

Based on this model, the main goal of this paper is to test whether initial conditions, proxied by the income distribution prevailing in the 1700s and 1800s, and taking into account the current business environment, have a detrimental effect on today's chances of businesses being created, surviving, and creating jobs over time.

Since our interest is to look at the effect of initial conditions on the dynamics of entrepreneurship, ideally we would want to follow firms over time. Unfortunately, empirically it is difficult to follow the same firms over time, especially if firms die in large numbers creating substantial attrition bias and if surveys are being censored by not representing newly created firms. We overcome these limitations by constructing a pseudo-panel of entrepreneurs using the Global Entrepreneurship Monitor (GEM)

survey, the largest comparable dataset covering 70 countries over 2001-2009.[1] The GEM datasets are drawn from a new sample in each country every year. However, the surveys include nationally representative information on how many people claimed to be entrepreneurs, whether they are involved in nascent, young, established firms, or have shut down businesses over the last year; as well as information on firm's size at each of these different stages of entrepreneurship.[2] Thus, using this information we build a pseudo-panel of cohorts of people based on their age and gender for each country following the methodology proposed by Deaton (1985). In doing so, we are able to track generations of people over time and assess whether initial conditions and current business environment affect the creation, survival of firms, as well as job creation.

We complement the GEM survey with historical data of income distribution from the 1700s and 1800s as estimated by Morrisson and Murtin (2011) and Bourguignon and Morrisson (2002) respectively. We also use historical indicators of GDP per capita prevailing in the 1800s, obtained from the historical databases estimated by Maddison. In addition, we use the index of credit protection provided by the World Bank, which measures the degree to which laws protect the right of borrowers and lenders, thus proxing the extent to which laws are designed to expand access to credit.

We combine the pseudo-panel methodology with instrumental variables given that the index of law protection of borrowers and lenders we use could be endogenously determined by the proportion of people involved in entrepreneurial activities, who for instance may lobby having better laws. As instrumental variables we use the legal code of origin and the colonial origin, both variables frequently used in the literature when dealing with the endogeneity of business regulation (La Porta, 1998; 1999). In addition, we use the average blood pressure and cholesterol, instruments that have been found in the literature to be correlated with the physiology responses to economic stress, such as credit constraints (Ezzati *et al*., 2005; O'Neil *et al*., 2005).

We find that initial conditions have a detrimental effect on development, even when taking into account current regulation in the credit market. Countries that started

[1] Although the survey covers 70 countries we include in our analysis only 48 as are the ones we could obtain data on historical income distribution.
[2] Nascent firms are those recently created that have not payed wages for more than three months; young firms have been running for up to 3.5 years and established firms have been running for more than 3.5 years.

with a high ratio of rich to poor people during the 1700s or 1800s currently are less likely to open new firms, and of these to survive, and create more jobs over time.

Although several articles have tested whether inequality has a detrimental effect on growth, our central contribution to the literature relies on testing an overlooked mechanism as why this might be the case (Banerjee and Duflo, 2000; Benabou, 1996). Specifically, our results suggest that high levels of inequality prevent people from taking up business thereby affecting job creation and development in the long-run.

Our findings also suggest that improvements in the regulation of current credit market promote the creation of both businesses and jobs. This effect however is of lower magnitude in Africa than in other regions, perhaps because some African households lack property rights of their land, thus prevented from providing a collateral and accessing credit.

The article proceeds as follows. Section 2 discusses the literature on entrepreneurship, including the model by Banerjee and Newman. Section 3 describes the dataset and the construction of the pseudo-panel. Section 4 presents the econometric results. Section 5 presents robustness tests. Section 6 concludes.

2.2 Institutions and Initial Conditions and Entrepreneurship

This paper is related to the large literature analyzing the factors that foster or constraint the success of entrepreneurs. According to the interdisciplinary surveys on entrepreneurship by Naudé (2008, 2010) the literature has experienced three important developments over the last decades, which we describe below.

First, while the traditional research in entrepreneurship has focused on empirically assessing the entrepreneur themselves, there has been a shift from analyzing their personality traits and individual characteristics towards their behavior and cognitive issues that enable them to recognize and exploit opportunities (Blanchflower *et al*., 2001; Caliendo and Kritikos, 2011; Shane and Venkartarman, 2000). The same has happened in economics where there has been a shift towards developing theoretical models of occupational choice (Evans and Jovanovic, 1989; Lucas, 1978).

The second development in the literature has been to examine how business environment influences the creation of firms and its relationship with long-run development (Throton, 1999). Within this literature one can distinguish two veins. The

first one analyzes the extent to which historical institutions affect current ones which in turn influence today's entrepreneurial sector and growth. These studies, for instance, examine the development path of former colonies.[3] The second vein studies the impact of current business regulation (such as investor protection and regulation of entry) on entrepreneurship (Djankov *et al*., 2002; Glaeser *et al*., 2004; La Porta *et al*., 1998). Within this vein, there is no consensus on whether business regulation always favors entrepreneurship. For instance, business regulation could impose a burden on firms if the regulation is aimed at extracting rents for the benefit of bureaucrats or certain industries. However, the public interest theory of regulation argues entrepreneurship can be fostered if regulation reduces market failures, by for instance allowing lenders to seize the collateral in case borrowers default (Ardagna and Lusardi, 2008).

The third development in the literature has been the theoretical analysis on the relationship between initial conditions, specifically wealth distribution, and development on the long-run. This literature, within the neoclassical viewpoint, analyzes whether initial conditions, such as country's past inequality, can affect entrepreneurship and economic growth in the long-run (Galor, 2011; Murphy *et al*., 1989). [4] There is no consensus to the extent initial conditions can affect development. On the one hand, the supporters of the "big push" hypothesis, argue that if there is the possibility of coordination of investment across various sectors in the economy, which can be promoted with public policy, countries can get out of no-industrialization/development traps (Murphy *et al*., 1984; Rosenstain-Rodan, 1943). On the other hand, other articles argue that initial conditions can determine development path. For instance, inequality, it is argued, can have a long-term detrimental effect on growth if the wealthier individuals lobby against changes in policies or institutions that could distribute wealth and foster a more inclusive growth.[5] Inequality can also have a detrimental effect on entrepreneurship if a large proportion of individuals are prevented from taking up profitable investments, thus perpetuating inequality and low levels of economic growth in the long-run. This negative effect of inequality on long-run development could be enhanced whenever accompanied by credit market imperfections

[3] For instance, Acemoglu *et al*. (2001) show that settler colonies perform better than former extractive colonies because they inherited institutions that better protect private property rights.
[4] See Benabou (1996) and Galor (2011) for a complete literature review on the effect of inequality on development.
[5] For an extensive overview of the dynamic interaction between political institutions and the development process see Acemoglu *et al*. (2005).

(Aghion and Bolton, 1997; Banerjee and Newman, 1993; Galor and Zeira, 1993; Ghatak and Jiang, 2002).

Within the third development in the literature, there are few empirical papers testing the effect of wealth distribution on entrepreneurship, and among the existing ones usually done in a static way and for a single country. Nonetheless, supportive evidence has been found in the USA that wealthier individuals are more likely to become entrepreneurs (Hurst and Lusardi, 2004). There is however, mixed evidence on whether inequality affects entrepreneurship, or the other way around. For instance, Mesnard and Ravallion (2001) show for the case of Tunisia the number of business start-ups is an increasing function of aggregate wealth and that the greater the initial inequality of wealth, the lower the overall rate of new business start-ups.[6] In contrast, Yanya (2012) concludes that firm establishment causes poverty and income inequality, but not the other way around using a panel data of the 76 provinces in Thailand over 1997-2008.[7]

2.2.1 Banerjee and Newman's Occupational Choice Model

In this paper we follow the theoretical model by Banerjee and Newman (1993) as it encloses the three main developments described above in the literature: analyzing why people choose to become entrepreneurs, and how business institutions along with initial wealth distribution can affect entrepreneurship and development in the long-run.

Specifically, Banerjee and Newman's model assume that because of imperfections in the credit market, people can borrow only limited amounts. These imperfections are derived from the possibility that borrowers may renege their debt. To prevent this, lenders will limit borrowing and will require a collateral, such that:

$$L < w + (\pi F / \bar{r}) \qquad (1)$$

where L is the amount borrowed, w is the borrower's wealth, π is the probability of the borrowers being caught if renege their debt, F is the nonmonetary punishment of being caught, and $\bar{r}$ represents the return from a divisible safe asset which the model assumes

[6] Initial wealth is captured by the amount of wealth accumulated by returned migrants from past savings while abroad.

[7] Income inequality is measured through the Gini index and poverty with the lowest income quintile at the province level. The causal relationship is assessed using the granger causality test

requires no labor. The model assumes that anyone that invests only in this safe asset is said to be idle or subsisting.

To become an entrepreneur people need to make an up-front investment. Thus, entrepreneurship is only available to those individuals that are wealthy enough to make this investment or provide the required collateral to accessing credit. Those poorer individuals that do not have enough wealth to provide collateral have two occupation choices: they can become employees, and for those individuals with individuals with wealth between w^* and w^{**} they can also become self-employed. Self-employment is assumed that requires some up-front investment but of lower level than the required to become entrepreneur. As entrepreneurship requires an up-front investment is available only to wealthy people or those who can provide the required collateral, whereas poorer individuals credit constrained their choices are limited to becoming employees and if have wealth between w^* and w^{**} will be able to become self-employed if they chose to.

The expected return to self-employment and subsistence are given exogenously by the model's parameters. Wage *v,* is determined endogenously in the model such that it clears the labor market, and in turns determines the returns of entrepreneurs and workers.

The equilibrium wage can take a low value $\underline{v}$ if $G_t(w^*) > \mu[1 - G_t(w^{**})]$, a high value $\bar{v}$ if $G_t(w^*) < \mu[1 - G_t(w^{**})]$ and a value within the range $[\underline{v}, \bar{v}]$ if $G_t(w^*) = \mu[1 - G_t(w^{**})]$.

where $G_t(w^*)$ is the proportion of the population that has no other choice but to become a worker, as does not have enough wealth to provide a collateral to become entrepreneurs. $\mu[1 - G_t(w^*)]$ is the proportion of the population that can become entrepreneurs. Then, the pattern of occupational choice that is generated in equilibrium is summarized as:

1) individuals with initial wealth less than w^* will be a worker unless wages are exactly the minimum wage $\underline{v}$,

2) individuals with initial wealth between w^* and w^{**} can become self-employed.

3) individuals with $w \geq w^{**}$ will be an entrepreneur if $v < \bar{v}$. In the case $v = \bar{v}$, then $1 - G_t(w^*)/\mu - G_t(w^{**})$ of them will opt becoming self-employed for the labor market to clear.

Then the pattern of occupational choice is determined by the initial distribution of wealth, and the structure of occupational choice determines in turn, how much people can save and leave a bequest. These factors, in turn give rise to a new distribution of wealth affecting long-run development.

The model predicts that the fate of the economy depends on the initial wealth distribution. Countries with an initially high proportion of non-credit constrain people will grow over time aided by a high share of people being able to start-up business, of these surviving over time and with an active labor market paying high salaries. A contrasting equilibrium could be reached if a country starts with a high proportion of credit constrained people. In this case, the process of development ends up in a situation of low wages, in which there is (almost) only self-employment at small scale.

Based on Banerjee and Newman model, we will test the following two hypotheses:

Hypothesis 1: Countries that have a historical high ratio of wealthy to poor people, a proxy for being non-credit to credit-constrained, have a lower probability of firms being created, surviving and of these creating jobs over time.

Hypothesis 2: Countries that currently have more efficient credit markets have a higher probability of people being involved in entrepreneurship and higher job creation.

2.2.2 Endogeneity between Credit Regulation and Entrepreneurship

When testing our second hypothesis we are likely to encounter an endogenity problem. We would expect that more efficient credit markets will benefit entrepreneurs. However, it is also possible that the degree of imperfections in the credit market change as response to the number of entrepreneurs in the economy, for instance if by lobbying for better regulation (Besley and Gathak, 2010; Manski, 2000). We address this potential endogeneity by using instrumental variables.

We use four instrumental variables across all the regressions presented in section 4. Two of these variables (origin of country's legal code and colonial origin), are drawn from the institutional literature that has used these instruments to deal with the endogeneity of current business environment. The other two instrumental variables used (average blood pressure and cholesterol level) are drawn from the recent literature on physiological responses to economic stress that can prove to be constrained from credit. We explain below the literature supporting the use of these instruments.

Based on the pioneer work of La Porta *et al*. (1998, 1999) several authors have addressed the likely endogeneity of current business environment using as instrumental variables the country's historical legal origin (Ardagna and Lusardi, 2008; Djankov *et al*., 2003; Gleasser *et al*., 2004; Levine *et al*., 2000). La Porta *et al*. show that the legal rules protecting investors are greatly dependent on the legal traditions or origins. For instance, they find that countries under the English common law are more protective of investor rights and contractual enforcements than the laws originated in the French civil code. Thus, countries with "better" legal origins are more likely to develop institutions in which property rights are protected and less distortionary policies are implemented, which in turn favor investment and economic growth.[8] Other studies have also found that, the colonial origin of the country is a strong predictor of current's institutions (Acemoglu, *et al*., 2001). These authors stress that different types of colonization policies created different sets of institutions which persisted over time. In one extreme, whenever colonizers aimed at exclusively draining resources from the colony developed "extractive" institutions with poor emphasis on protecting private investment.[9] In contrast, whenever colonizers intended to settle in these colonies in the long-run, they tried to replicate European institutions, protecting property rights.[10]

Recent literature has found that people who find hard to gain access to credit can experience physiological responses to stress. For instance, people experiencing financial distress are less likely to follow recommended health maintenance practices such as eating a healthy diet, thus elevating risk of cardiovascular diseases, elevated blood pressure, and cholesterol (O'Neill *et al*., 2005). Also, cardiovascular diseases and their nutritional risk factors such as overweight and obesity, elevated blood pressure, and cholesterol, have been predicted to rise with economic development and hence to vary across regions, an important aspect since the credit market regulation we analyze vary sharply across countries (Ezzati *et al*., 2005).

[8] La Porta *et al*. (1998) stress that countries under the English common law have the best investor right protection and contractual enforcements, followed by those under German or Scandinavian civil law, and of these followed by countries with French civil law.

[9] Belgian colonization in the Congo is an example of extractive institutions, whilst the Great Britain colonization of Australia, New Zealand, United States and Canada are examples of pro-European institutions (Acemoglu *et al*., 2001).

[10] Acemoglu *et al*. (2001) argues that former British colonies prospered relative to former French, Spanish, and Portuguese colonies because of the good economic and political institutions and culture they inherited from Britain.

2.3 Data and Methodology

2.3.1 Historical Income Distribution and Current Credit Regulation

In our regression models presented in Section 4 we control for countries' initial wealth per capita. For such purpose, we use the GDP per capita prevailing in the 1800 estimated by Angus Maddison's historic income database.[11]

We also use the historical data on income distribution prevailing in the 1700s and 1800s as estimated by Morrisson and Murtin[12] (2011) and Bourguignon and Morrisson (2002) respectively. These estimates provide the income share for each decile, which we use to build different indicators to proxy the historical ratio of people that were credit and non-credit constrained, such as the Gini coefficient and different ratios of income shares across different deciles. We do so as Banerjee and Newman stress that income inequality is the main factor preventing poor people to invest in entrepreneurial activities. Moreover, previous research has shown that people in the lower deciles are less likely to have access to credit, as they might not to have enough wealth to provide a collateral or are living away from a banking institution, thus affecting their chances of obtaining credit (Baliamoune-Lutz *et al*., 2011; Berg, 2013).

Since we are interested in assessing the impact that credit market imperfection have on the creation of firms and jobs over time, we use indicators on credit protection from the Doing Business database gathered by the World Bank from 2004 to 2009.[13] Specifically, we use the strength of legal right index which "measures the degree of which collateral and bankruptcy laws protect the right of borrowers and lenders and thus facilitate lending." The index ranks from 0 to 10, where higher scores indicating that collateral and bankruptcy laws are better designed to expand access to credit.[14] This variable is particularly suitable for our analysis as it is a proxy of the extent to which better credit rules can enhance investment incentives by improving

[11] Online data available at: Maddison Project website http://www.ggdc.net/maddison/maddison-project/home.htm

[12] We thank Fabrice Murtin for having provided us these datasets.

[13] Since the Doing Business dataset covers the year 2004 until 2009, we imputed the values for the years 2001 and 2002 taking the information for the year 2004 or for the closest year we had information on. We did so to retain as much information as possible for earlier years, and given the little change in business environment observed for the years we have.

[14] Data on the legal rights of borrowers and lenders are gathered through a questionnaire administered to financial lawyers and verified through analysis of laws and regulations as well as public sources of information on collateral and bankruptcy laws. A detailed description of the elaboration of this index can be found in: http://www.doingbusiness.org/methodology/getting-credit

collateralizability of assets and limiting its seizing. All those aspects improve property rights thereby reducing imperfections in the market (Besley and Gathak 2010).

2.3.2 GEM Survey

We use the Global Entrepreneurship Monitor (GEM) survey, the largest study on entrepreneurial activity over 2001-2009.[15] The surveys are representative of the adult population and are collected annually. In developing countries the survey is conducted in face-to-face interviews, and otherwise through random telephone interviews.

We use the definition of "entrepreneurs" proposed by the GEM network: "*adults in the process of setting up a business they will (partly) own and or currently owning and managing and operating young businesses*" (Reynolds *et al*., 2005 p. 209). To study the dynamics of entrepreneurship we consider the four stages in the life-cycle of businesses, as defined by the GEM network. In the first stage, *nascent entrepreneurs*, are those actively involved in setting up a business they will own or co-own, but who have not paid salaries, wages or any other payments to the owners for more than three months. In the second stage are the owners of *young firms*, defined as those who have paid salaries for more than three months and up to 3.5 years. In the third stage, *established firms*, are those who have paid salaries or wages for more than 3.5 years. In the fourth and last stage are firms that in the past 12 months, have been sold, shut down or discontinued.

For our analysis, we focus on 48 countries surveyed in GEM for which we also have gathered historic information on income distribution and GDP per capita. The countries analysed are listed in Table A.1 in the Appendix. In total, we have 1,001,458 individuals interviewed over 2001-2009. From these, 37,136 were in nascent entrepreneurship, 32,359 in young firms, 62,514 in established firms and 25,183 had recently shot down their firm.

In the Appendix we report the basic descriptive statistics of the pseudo-panel and other aggregate indicators used for the whole sample, and differentiating between groups of countries –low-middle and high income countries (Tables A.3 and A.4). A characterization of firms at different stages of entrepreneurship, such as size and sector, and grouped by region are presented in Table A.6. In Table A.7, we report the

[15] The chosen period of analysis refers to that for which the GEM datasets are publicly available.

correlation matrix among all the dependent and explanatory variables used, which show that we have no problems of multi-colinearity.

Figure 1 shows the percentage of the population engaged in the various states of entrepreneurship analysed over 2001-2009. The onset of the economic crises reduced the percentage of the population involved in entrepreneurial activities across all stages (nascent, young and established firms) particularly in 2009.

2.3.3 Pseudo-Panel

Since GEM draws new samples each year, the surveys remain representative of the population engaged (or that were engaged) in entrepreneurial activities over time, avoiding an attrition bias. Since a new sample is drawn each year, we cannot study the decision of the *same* individuals to become or remain in entrepreneurial activities over time. To overcome this limitation, we construct a pseudo-panel using the GEM surveys and the methodology proposed by Deaton (1985). We describe next the construction of the pseudo-panel.

GEM consist of a set of T independent cross-sections of i individuals that belong to a new and most likely different set of I individuals in each period. Equation (2) denotes the factors that affect whether a person is an entrepreneur, if we were to stack together all the cross-section observations, typically known in the literature as pooled-cross section.

$$y_{it} = \beta x_{it} + \delta_i + \varepsilon_{it} \qquad (2)$$

where y_{it} denotes whether the individual is engaged in an entrepreneurial stage, x_{it} denotes a vector of explanatory variables, δ_i and ε_{it} are the individual-specific time-constant unobserved heterogeneity; and the unobserved idiosyncratic error that varies over individuals and time.

OLS estimates using this pooled-cross section data will be biased and inconsistent if the individual unobserved characteristics (such as personal traits, risk aversion or cognitive abilities) were correlated with some or all of the explanatory variables. To solve this potential endogeneity problem, Deaton (1985) proposed building a pseudo-panel, which yields consistent β estimators, even when the individual unobservables characteristics are correlated with explanatory variables. Pseudo-panels

have the additional advantage of avoiding attrition problem that plagues genuine panels since data is collected from random samples drawn from cross sections.[16]

To build the pseudo-panel Deaton (1985) proposes to average observations with similar characteristics that are stable over time (such as gender, year of birth) in a sequence of repeated cross-sectional datasets. These synthetic observations can be therefore thought as cohorts of generations being "followed" over time, just as if pure panel surveys were available.

Following Gutiérrez-Romero (2012) who built a pseudo-panel using the GEM survey for the case of Spain, we build the pseudo-panel by defining the cohorts within countries in terms of gender and year of birth, as these are observable and do not change over time.[17] In total, we have nine time periods (2001-2009) and 10 cohorts in each. Five of these cohorts are for males, and five for females. Within each gender we further defined five cohorts of age: those who in 2001 were 28 years old or less, 29-38, 39-48, 49-58 and 58 or over.[18] The average sample size for each cohort is shown in Table A.5.

We produce the pseudo-panel by averaging observations over individuals in each of the cohorts *C* described above and *T* periods, as shown in equation (3).

$$\bar{y}_{ct} = \beta \bar{x}_{ct} + \bar{\delta}_{ct} + \varepsilon_{ct} \qquad (3)$$

where the bars denote the average value of all individuals in cohort *c* at time *t.* The average of the fixed effects of those members belonging to cohort *c* in the sample $\bar{\delta}_{ct}$ varies over time. Since $\bar{\delta}_{ct}$ is unobserved it might be correlated with $\bar{x}_{ct}$ therefore leading to inconsistent estimates.[19] In addition, treating $\bar{\delta}_{ct}$ as a fixed effect can lead to

[16] The pseudo-panel approach is especially useful for life-cycle models, and has been recently taken in empirical studies for which panel data is not available, largely used in social mobility analysis (Antman and Mckenzie, 2005) and previously used for studying entrepreneurial success of the Spanish case in Gutiérrez-Romero (2012).

[17] We also define cohorts following age and gender as the literature has found evidence of the probability of being engaged in entrepreneurial activities differs considerably with regard to these two variables and allows to explicitly recognizing the life-cycle stage a firm is in (Bergmann and Sternberg 2007).

[18] For instance, individuals are considered to belong to the first cohort of age if they were aged 30 in year 2001, 31 in 2002, 32 in 2003 and so on.

[19] This is likely in our case because we consider a number of explanatory variables that might be correlated with the error term, such as individuals' personality traits like risk aversion and cognitive abilities. Since these characteristics are unobservable and might be correlated with our outcome of interest, the estimated effect could be biased.

an identification problem, unless it is assumed that the individual error is time invariant, that is $\bar{\delta}_{ct} = \bar{\delta}_{c}$.

Baltagi (2005) argues that pseudo-panels estimations could be biased if cohorts do not have enough observations to eliminate a potential unobserved heterogenity bias. Verbeek and Nijman (cited by Gutiérrez-Romero, 2012) show that if each cohort has greater than 100-200 observations, as it is our case, then the cohorts will be large enough to eliminate the unobserved heterogeneity bias if assumed the individual error is time invariant. In that case, equation (3) can be estimated using cohort dummy variables yielding unbiased estimators.

To ensure that the estimators are also efficient, we control for the likely problem of heteroskedasticity, which could occur if the number of observations per cohort varies substantially. To correct for this we use weighted least squares (WLS) by weighting by the square root of the number of observations in each cohort, as it is recommended in the literature (Dargay, 2007).

2.4. Econometric Results

2.4.1. Firm's Life Cycle: Birth, Maturity and Death

To test our two hypotheses and to deal with the potential endogeneity of the degree of credit protection we extend equation (3) and estimate instead equation (4). We correct for this endogeneity in a two-stage process, as it is standard in the literature, so equation (4) represents the IV-second-stage least squares estimation.

$$E[\bar{y}_{ct}|Z] = \alpha + \beta_1 INEQ_{1820} + \beta_2 Lindex + \beta_3 \bar{x}_{ct} + \beta_4 X + \bar{\delta}_{ct} + \varepsilon_{ct} \qquad (4)$$

where $\bar{y}_{ct}$ measures the dependent variable in the second-stage least square, as the proportion of individuals involved in a specific stage of entrepreneurship, namely nascent, young, established or recently closed firm. $INEQ_{1820}$ represents the historical ratio of wealthy people (income share of top 9th decile) to poor people (bottom 1st decile) prevailing in 1820. We use this indicator as a proxy of the ratio of non-credit to credit constrained people. *Lindex* represents the strength of legal right index[20], X is a set

[20] Note that the legal right index ranks from 0 to 10, however this index is not equal to 0 for none of the countries over the period time considered in the analysis, then being possible to make this log transformation.

of characteristics, which includes GDP per capita in 1800, regional and year dummy variables to control for unobserved regional and time effects. At cohort level, in $\bar{x}_{ct}$ we include the proportion of people in cohort *c* at time *t* with secondary education or more, and control for cohort fixed effects $\bar{\delta}_{ct}$. Z is the instrument used in the first-stage least squares, which is a dummy variable for whether the country's legal origin's code is English or not. All variables are measured in logarithms except the generation cohort, the instrumental variable *Z*, regional and time dummy variables.

Table A.8 (in Appendix) shows the results of the first-stage regressions. This table includes the coefficients associated with our instrument, whether the origin of the legal code is English, and our endogenous variable, the legal right index. We find that the instrument is positive and statistically significant across all models presented. We also include the summary statistics for the first stage regressions, in which the F-statistics test of the excluded instrument, is greater than 10 and statistically significant across all models ran, which suggest our instrument is not weak

Table 1 presents the results of the IV-second-stage least squares. There we also include the endogeneity test which confirm that the legal right index is endogenous with the our dependent variable $\bar{y}_{ct}$, the proportion of people involved in different entrepreneurial stages. The Kleibergen-Paap Wald F statistic test confirms the instrument is correlated with the endogenous variable, the legal right index.[21]

Our results confirm the first hypothesis. The higher the ratio of wealthy to poor people in the 1820 the lower the probability that people were engaged in entrepreneurial activities across all stages, nascent, young and established firms, during the period 2001-2009 (Table 1, columns 1-4). The lower the income share of the poor relative to the wealthy, the less the share of people involved in firms of any type. For instance, a 1% increase in the historic ratio of wealthy to poor reduces the proportion of people involved in nascent firms by 0.2%, the proportion of people involved in young firms by 0.17% and the proportion of people involved in established firms by 0.08%.

We also find evidence to support our second hypothesis. The higher the index of legal rights, a proxy we use for efficiency in the credit market, the higher the proportion of people involved in entrepreneurial activities. Specifically a 1% increase in the legal right index, increases the proportion of people involved in nascent firms by almost 1%, the proportion of people involved in young firms by 0.8%, and the proportion of people

[21] We do not present the exogeneity test, which test the null hypothesis that the instruments are jointly exogenous, since this test can only be conducted with more than one (Baum, 2006).

involved in established firms by 0.22%. These results suggest that the strength of the legal right index is more important in the early stages of entrepreneurship than those already established. There are potential reasons for this. For instance, already established firms might have had time to generate their own financial resources (from previous profits) and had enough time to develop networks, other than with financial markets, that could enable them to stay afloat in case of requiring prompt credit. This argument is in line with previous research that shows small and medium firms are more likely to be more credit constrained than larger firms (Claessens *et al*., 2007). For instance, Kuntchev *et al*. (2013) show that the firms' perception of being credit is negatively correlated with firm's size and age: smaller and younger firms tend to find access to credit to be more of stringent constraint to carry out their operations than larger and older firms.

We also find the higher the historical GDP per capita, the less people would be involved in different stages of entrepreneurial activity over time. It is unclear why this might be the case. One potential reason, and in line with the predictions of Banerjee and Newman model, is that countries that started with higher historical GDP per capita over time developed a more active labor market, paying higher wages. As wages rise, more people would prefer becoming workers, instead of entrepreneurs.

The cohort effects on entrepreneurial activity show that in general, older individuals are more prone to be involved in established business, while younger people are engaged in young firms. This result is consistent with previous studies that show that because knowledge, capital accumulation, and experience increase with age, over time individuals are more likely to have an established firm (Bergmann and Sternberg, 2007).

In addition, we find evidence that the higher the proportion of people with high school or more, the less likely would be engaged in entrepreneurial activities, for all nascent, young and established firms during the period 2001-2009. A number of studies have found a positive correlation between education and degree of entrepreneurship, suggesting that education helps people identify opportunities in the market place and provide them with needed managerial abilities (Simón-Moya et al. 2014). Our findings instead, support the other vein in the literature that has found education to be negatively related to the probability of being self-employed (Blanchflower, 2004; Reynolds et al., 2003). These studies argue that education is not necessarily correlated with being an entrepreneur as specific entrepreneurial knowledge is what matters more, managerial

abilities and knowledge in accounting and finance (Man *et al*., 2002). Other empirical studies have found that employees in Spain and Portugal value more having higher level of educations, whilst self-employed people have lower levels of education (Garcia-Mainar and Montuenga-Gomez 2005).

To conclude this sub-section, we focus on the regional differences on firm's life cycle. We find that Africa was less likely to create firms and of these to survive over time than firms located in the rest of the world over 2001-2009. These results might reflect the structural and institutional differences supporting entrepreneurship in Africa and the rest of the world.

2.4.2 Job Creation: Firms' Size

We next move on to analyze the extent to which initial conditions and credit market affect the number of employees hired by firms, as shown in equation (5)

$$E[\bar{s}_{ct}|Z] = \alpha + \beta_1 INEQ_{1820} + \beta_2 Lindex + \beta_3 \bar{x}_{ct} + \beta_4 X + \beta_5 Lindex * region + \bar{\delta}_{ct} + \varepsilon_{ct} \quad (5)$$

where $\bar{s}_{ct}$ represents the average number of employees hired by firms in each stage of entrepreneurship in the cohort *c* at time *t*. In addition, we interact the legal right index with a regional variable ($Lindex * region$) to take into account for regional differences in the credit regulation. We also add in $\bar{x}_{ct}$ a categorical variable denoting the sector of the firm and a dummy variable denoting whether the firm has a medium/high level of technological intensity, both variables provided in the GEM surveys.

We chose these explanatory variables following the literature on the determinants of firms' size. Our key explanatory variable affecting firm size over time is the historical ratio of wealthy to poor. We include this variable based on the theoretical model of Banerjee and Newman, expecting that the higher the historical income inequality the smaller the firms will be. We also include in our regressions the legal right index, as the literature predicts that countries with better institutions and more access to credit to be more likely to develop larger firms (Beck *et al*., 2003; 2005; Kumer *et al.*, 2001). In addition, we control for sector fixed effects and technology intensity as the literature has found these variables play a crucial rule on firm's size (Aghion *et al.* 2007; .Kumar *et al*., 2001). Finally, we take account of market size, as the literature predicts that firms will expand in size depending on the expected profits of the market (Lucas. 1978). Since we are interested in studying the impact of initial

conditions, and to avoid a potential endogenity issue with current market size, we use as a proxy of market size the GDP per capita prevailing in 1800 and not actual GDP per capita. We measure in logarithm our dependent variables, the ratio of wealthy to poor, GDP per capita in 1800 and the legal right index.[22]

The legal right index is likely to be endogenous with the size of the firms, as well as the interaction of this legal right index with the seven regional dummies used.[23] Thus, we require at least eight instruments, one for our proxy to access to credit, and seven for this variable interacted with the dummy regional variables. The instruments Z we use are: the country's origin of legal code (one dummy for each legal code: English common law, French commercial code, Socialist/Communist law, German commercial code and Scandinavian commercial code); the colonial origin of the country (a dummy variable equal to one if the country's colonial origin is Spanish, and zero otherwise) and two variables that measures blood pressure and cholesterol at country level.[24]

In Tables A.9.1 to A.9.3 (in the Appendix) we provide the first stage regressions. This table includes the coefficients associated with our instruments and our endogenous variables, the legal right index and its interaction with the regional variables. We find that the instruments are statistically significant across all models presented. The F-statistics test of the excluded instruments are greater than 10 and statistically significant across all models ran, which suggest our instruments are not weak.

In Table 2 we present the IV second-stage least squares. There we also include the endogeneity test which shows our dependent variables are endogenous. As before, we include the Kleibergen-Paap rank Wald F statistic test which confirms that our instruments are not weak. All models are just identified.

We find mixed evidence to support our first hypothesis. On the one hand, the higher the historical ratio of wealthy to poor, the bigger the nascent firms were over 2001-2009 (Table 2, column 1). On the other hand, and in line with our first hypothesis, the higher the historical ratio, the smaller the young and established firms are over time

[22] Given that firms could have no workers hired, if taking the logarithm of our dependent variable would lose several obseravtions. To prevent this, we trasnsform our dependent variable by adding one to the number of hired workers. Then we take the logarithm of that number, and that is the variable we use as dependent variable.

[23] The regions considered in the analysis are: Africa, Asia, Western Europe, Latin America, North America, Oceania and Eastern Europe.

[24] Table A.2 in the Appendix shows in further detail the instrumental variables definitions and data sources.

(Table 2, columns 2-3). This evidence suggests that as the income share of the poor shrinks (the higher the historical ratio of wealthy to poor) the bigger the nascent firms aided perhaps by low salaries. But, once firms get older they shrink in size. This apparent mixed evidence is however consistent with the predictions of Banerjee and Newman (1993). Their model predicts that countries with high ratio of rich to poor people will fail in the long-run building a real demand for the local market production, thus affecting the size of firms as they mature. In similar line, Murphy, Shleifer and Vishny (1989b) show that countries with high income inequality will have a low demand for labor as they do not have a critical mass in their markets to justify firms of bigger size.

We find evidence to support our second hypothesis. The higher the legal right index the bigger the firm's size, across at stages of entrepreneurship.[25] The effect of this index is greater for the nascent firms, and decreases for young and established firms. Which confirms, as earlier shown, once firms are already established they might be less dependent of external credit resources than those firms that just started.

We also find that improving credit protection increases the firm size to lesser extent in Africa than compared to other regions. Which suggests that even if regulation is improved in Africa, its effect will be of lesser impact than in other regions, perhaps because fewer people in Africa will be take advantage of the improved institution if do not have the required collateral. Thus, policy interventions aiming to reduce barriers to access to credit should take into account the specificities of the different regions. In some regions, the problem could be the lack of resources or competition of the banking system, the lack of protection to lenders; while in others could be the excess of collateral requirements. For instance, Baliamoune-Lutz *et al*. (2011) point out that a major issue for African countries is the collateral needed to secure bank loans. Some households in these countries do not have formal titles of their lands, and the constraint is particularly severe for women-headed households.

[25] We obtain the total effect of this legal index by adding up the coefficients of the legal right index and the interactions between this variable and the regional dummies, which turned statistically significant across all specifications in Table 2.

2.5. Robustness Checks

We conducted three main robustness checks to assess the validity and consistency of the results so far presented.

First, we re-run our IV-pseudo-panel regressions but excluding from the analysis self-employed people, in other words, people who claimed were not hiring workers. We do so as the model by Banerjee and Newman (1993) distinguishes between self-employed and entrepreneurship. Table A.10 tests our two hypotheses on the probability of people being engaged in entrepreneurial activities, and Table A.11 on the size of the firms. Both tables confirm our previous results inequality is detrimental for the creation of business, of these surviving and creating jobs overtime, whilst better legal right index if beneficial.

Second, we test alternative inequality measures, four different ratios of wealthy to poor and other indicators such as the Gini index, finding no differences with the results so far presented.[26] Tables A.12 and A.13 in the Appendix show the overall, the detrimental effect of income inequality remained on firm's life cycle and job creation across the alternative indices used. For instance, when using the Gini index, we observe that the higher this index, the less people involved in nascent and established firms. However, we find a positive relationship between the Gini index and the proportion of people involved in young firms, which is opposite to what is observed in Table 1.

Across all regressions presented in Table 1 and Table 2 we also tested the ratio wealthy to poor but for 1700. This ratio yields practically identical results to once present using the ratio 1820s, hence we omitted them.

Third, we consider different instrumental variables in our estimations, such as, religion and language fractionalization (Alesina *et al*., 2002), instruments commonly used in the literature. However, all of these variables proved to be weaker instruments than the ones presented in our estimations. Tables A.14 to A.17 presents the estimated coefficients of the key independent variables and a summary of first stage regressions, weak identification test and endogeneity test. Overall, due to the weakness of the

[26] These four ratios are defined as: The income share of the 1st decile to the average income (bottom 10); income share of the 9th decile to average income (top 90); income share of the median to the average income (middle50); the income share of the 8th decile to the income share of the bottom 2nd decile (top20/bottom20). We also use the sum of the income shares of the 2nd, 3rd and 4th quintiles (middle).

instruments we obtain inconsistent estimations in comparison to the ones obtained with strong instruments.

2.6. Conclusion

The aim of this article was to test the influence of historical income inequality along with the current business environment on the probability of creating new businesses and of these surviving over time and creating jobs at different stages of the firm's life cycle. For this purpose, we built a pseudo-panel of cohort of people across 48 countries over 2001-2008, using the Global Entrepreneurship Monitor Survey and the pseudo-panel methodology proposed by Deaton (1985).

We draw two main conclusions from our results. First, initial inequality, understood as the inequality prevailing in the 1700s or 1800s, has a persistent and detrimental effect on the creation, survival of firms as well as job creation over time. Second, countries with worse credit markets, proxied in our analysis by an index that measures the easiness is to lend in the market, the less likely that firms will be created, survive and create jobs over time.

Our findings are consistent with the prediction of the model by Banerjee and Newman (1993). This model suggests that if the initial wealth distribution, is such that a large percentage of the population are credit constrained, then fewer firms will be created and survive over time especially, under the presence of credit market imperfections.

Despite the extensive research on the relationship between inequality and economic growth, there still remains considerable disagreement about the sign of this relationship in the literature. Banerjee and Duflo (2000) argue previous studies are far from being conclusive of this relationship because of identification problems and data limitations in cross-country studies. Moreover, most empirical papers have assessed the impact of inequality by using not so distant indicators of inequality, instead of historical ones, limiting our understanding of the extent that early inequality conditions, affect economic development over time.

To the best of our knowledge, this is the first empirical paper that tests the predictions of Banerjee and Newman model and other similar theoretical models that suggest initial conditions, understood as the wealth distribution prevailing in the distant past, can affect entrepreneurship and development in the long-run. Our results, have

important policy implications. Although we did not specifically test for convergence, our findings suggest that since some countries are predisposed by their initial conditions to be trapped into a firms-die-young equilibrium whilst others are in a different type of equilibrium with businesses thriving over time, thus, economic convergence across countries is unlikely to occur. Our findings, in line with the theoretical literature, suggest that to foster the creation of jobs and businesses, policies should focus on addressing long-standing differences in wealth within countries as well as reducing credit constraints. Incidentally, these policies could foster convergence across countries as well, an issue that deserves further research.

Acknowledgements

We thank Fabrice Murtin for having shared his estimations of income distribution prevailing in the 1700s and 1800s. We are grateful to Professors Maitreesh Ghatak and Elias Papaionnou for informal discussions on early stages of work. We thank Cristina López-Mayan, Adam Pepelasis, and the participants of the EDIE workshop, the GEM-Barcelona conference, UAB PhD seminar, Universidad Tecnológica Metropolitana de Mérida, the LACEA/IADB/WB/UNDP Research Network of Inequality and Poverty for their comments and suggestions on earlier stages of this paper. Finally, we are grateful to Isabel Busom for her comments on an earlier version of this paper.

References

Acemoglu, D., Johnson, S.; and Robinson, J. (2001) "The colonial origins of comparative development: An empirical investigation", *American Economic Review*, 91(5): 1369-1401.

Acemoglu, D., Johnson, S.; and Robinson, J. A. (2005) "Institutions as the fundamental cause of long-run growth", (in) P. Aghion and S. N. Durlauf (eds), *Handbook of Economic Growth,* Vol. IA, Elsevier North-Holland, Amsterdam, The Netherlands.

Aghion, P., and Bolton, P. (1997) "A theory of trickle-down growth and development", *Review of Economic Studies* 64(2): 151-172.

Aghion, P.; Fally, T., and Scarpetta, S. (2007) "Credit constraints as a barrier to the entry and post-entry growth of firms", *Economic Policy*, 22(52): 731–779.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S.; and Wacziarg, R. (2002) "Fractionalization", *Journal of Economic Growth,* 8(2): 155-194.

Antman, F., and McKenzie, D. (2005) "Earnings mobility and measurement error: A pseudo-panel approach", Stanford, United States: Stanford University WP.

Ardagna, S., and Lusardi, A. (2008) "Explaining international differences in entrepreneurship: The role of individual characteristics and regulatory constraints", NBER WP 14012.

Baliamoune-Lutz, M., Brixiová, Z.; and Ndikumana, L. (2011) "Credit constraints and productive entrepreneurship in Africa", Political Economy Research Institute WP No. 276.

Baltagi, B. H. (2005) "*Econometric analysis of panel data"* (3ed.). Chichester, Hoboken, N.J: John Wiley & Sons.

Banerjee, A., and Duflo, E. (2000) "Inequality and growth: What can the data say?", NBER WP No.7793.

Banerjee, A., and Newman, A.F. (1993) "Occupational choice and the process of development", *Journal of Political Economy*, 101(2): 363-394.

Baum, C. (2006) "*An introduction to modern econometrics using stata",* StataCorp LP.

Beck, T., Demirguc-Kunt, A.; and Levine, R. (2005) "SMEs, Growth and Poverty: Cross-Country Evidence", *Journal of Economic Growth*, 10: 199-229.

Beck, T., A. Demirguc¸-Kunt, and Maksimovic, V. (2003) "Financial and Legal Institutions and Firm Size", World Bank mimeo.

Benabou, R. (1996) "Equity and efficiency in human capital investment: the local connection", *Review of Economic Studies* 63(2): 237-264.

Berg, E. (2013) "Are poor people credit-constrained or mypic? Evidence from a South African panel", *Journal of Development Economics,* 101(3):195-205.

Bergmann, H., and Sternberg, R. (2007) "The Changing face of entrepreneurship in Germany", *Small Business Economics*, 28(2/3): 205–221.

Besley, T., and Ghatak, M. (2010) "Property rights and economic development" (in) D. Rodrik and M. Rosenzweig (Eds.), *Handbook of development economics* (vol. V, Chap. 68, pp. 4525–4595). Amsterdam: North-Holland.

Blanchflower D., Oswald, A.; and Stutzer, A. (2001) "Latent entrepreneurship across nations", *European Economic Review*, 45(4-6): 680-691.

Blanchflower, D. (2004) "Self-employment: More may not be better", NBER WP No. 10286.

Bourguignon, F., and Morrisson, C. (2002) "Inequality among world citizens: 1820–1992", *American Economic Review*, 92(4): 727–744.

Caliendo, M., and Kritikos, A. (2011) "Searching for the entrepreneurial personality: New evidence and avenues for further research", IZA DP No. 5790.

Claessens, S., and Perotti, E. (2007) "Finance and inequality: Channels and evidence", *Journal of Comparative Economics* 35: 748-773.

Dargay, J. (2007) "The effect of prices and income on car travel in the UK", *Transportation Research Part A*, 41(10): 949-960.

Deaton, A. (1985) "Panel data from time series of cross-sections", *Journal of Econometrics*, 30(1-2): 109-26.

Djankov, S., La Porta, R., López-de-Silanes, F.; and Shleifer, A. (2002) "The regulation of entry", *Quarterly Journal of Economics*, 117(1):1-37.

Djankov S., R. La Porta, Lopez-De-Silanes F.; and Shleifer A. (2003) "Courts", *Quarterly Journal of Economics*, 118(2): 453-517.

Evans D., and Jovanovic, B. (1989) "An estimated model of entrepreneurial choice under liquidity constraints", *Journal of Political Economy*, 97(4): 808-827.

Ezzati M, Vander Hoorn S, Lawes C., Leach R; and James W. (2005) "Rethinking the 'Diseases of affluence', Paradigm: Global Patterns of Nutritional Risks in Relation to Economic Development", *PLoS Med* 2(5).

Galor, O. (2011) "Inequality, human capital formation and the process of development", prepared for the *Handbook of the Economics of Education*, North-Holland.

Galor, O., and Zeira, J (1993) "Income distribution and macroeconomics", *Review of Economic Studies* 60(1): 35-52.

Garcia-Mainar, I, and Montuenga-Gomez, V. (2005) "Education returns of wage earners and self-employed workers: Portugal vs. Spain", *Economics of Education Review*, 24: 161-170.

Ghatak, M., and Jiang. N.H (2002) "A simple model of inequality, occupational choice and development", *Journal of Development Economics*, 69(1): 205-226.

Glaeser, E.; La Porta, R; Lopez-de-Silanes, F.; and Shleifer, A. (2004) "Do institutions cause growth?", *Journal of Economic Growth*, 9(3): 271-303.

Gutiérrez-Romero, R. (2012) "Determinants of Spanish firms' life cycle and job creation: A pseudo-panel approach", Universidad Autónoma de Barcelona WP 12.09.

Hurst E., and Lusardi, A. (2004) "Liquidity constraints, household wealth, and entrepreneurship", *Journal of Political Economy*, 112(2): 319-47.

Kuntchev, V., Ramalho, R., Rodríguez-Mesa, J.; and Yang, J. (2013) "What Have We Learned from the Enterprise Surveys Regarding Access to Credit by SMEs?", World Bank WP No.6670.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny. (1998) "Law and finance", *Journal of Political Economy* 106(6): 1113-1155.

La Porta R., F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny. (1999) "The quality of government", *Journal of Law, Economics and Organization*, 15(1): 222-279.

Levine, R.; Loayza, N., and Beck, T. (2000) "Financial intermediation and growth: Causality and causes", *Journal of Monetary Economics*, 46(1):31-77.

Lucas, R.E. (1978) "On the size distribution of business firms", *Bell Journal of Economics*, 9(2): 508-523.

Man, T.; Lau, T., and Chan, K. (2002) "The competitiveness of small and medium enterprises A conceptualization with focus on entrepreneurial competencies", *Journal of Business Venturing* 17: 123–142.

Manski, C. (2000) "Economic analysis of social interactions", *Journal of Economic Perspectives*, 14(3): 115-136.

Mesnard, A., and Ravallion, M (2001) "Is inequality bad for business?", Policy Research WP 2527, World Bank.

Morrisson, C., and Murtin, F. (2011) "Internal income inequality and global inequality", Foundation pour les etudes et recherches sur le développpement international, WP No. 26.

Murphy, K., Shleifer, A.; and Vishny, R. (1989a) "Industrialization and the big push", *Journal of Political Economy*, 97(5): 1003-1026.

Murphy, K., Shleifer, A.; and Vishny, R. (1989b) "Income Distribution, Market Size, and Industrialization", *Quarterly Journal of Economics*, 104(3): 537-564.

Naudé, W. (2010) "Entrepreneurship, developing countries, and development economics: new approaches and insights", *Small Business Economics*, 34(1): 1-12.

Naudé, W. (2008) "Entrepreneurship in economic development", UNU-Wider Research Paper No. 2008/20

O'Neill, B., Sorhaindo, B., Xiao, J. J.; and Garman, E. T. (2005) "Health, financial well-being, and financial practices of financially distressed consumers", *Consumer Interests Annual*, 51.

Rajan, R.; Zingales, L.; and Kumar, K. (2001) "What Determines Firm Size?" CRSP WP No. 496.

Reynolds, P., Bosma, N., Autio, E., Hunt, S., De Bono, N., Servais, I.; and Lopez-Garcia, P. (2005) "Global entrepreneurship monitor: data collection design and implementation 1998-2003*'*, *Small Business Economics*, 24(3): 205–31.

Reynolds, P., Autio, E.; and Hay, M. (2003) "*Global Entrepreneurship Monitor Report*", Kansas City, MO, US: E.M. Kauffmann Foundation

Rosenstein-Rodan, P. N. (1943) "Problems of industrialization of Eastern and South-Eastern Europe", *The Economic Journal,* 53(210/211):202-211.

Shane, S., and Venkataraman, S. (2000) "The promise of entrepreneurship as a field of research", *Academy of Management Review*; 25(1): 217-226.

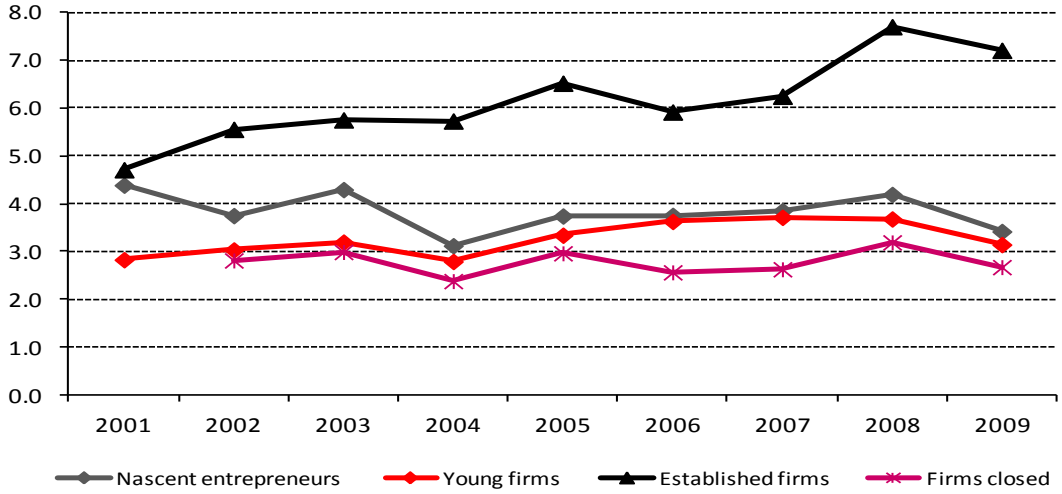
Simón-Moya, V., Revuelto-Taboada, L.; and Fernández-Guerrero, R. (2014) "Institutional and economic drivers of entrepreneurhsip: An international perspective", *Journal of Business Research*, 67:715-721.

Thornton, P. (1999) "The sociology of entrepreneurship", *Annual Review of Sociology.* 25(25):19-46

Yanya, M. (2012) "Causal relationship between entrepreneurship poverty and income inequality in Thailand", *International Journal of Trade, Economics and Finance*, 3(6): 436-440.

TABLES AND FIGURES

Figure 1 Proportion of people engaged in entreprenerual stages



Source: GEM 2001-2009

Table 1 IV Second Stage Pseudo-Panel Regression: Impact of inequality on firm's life cycle

	(1) Nascent IV		(2) Young IV		(3) Established IV		(4) Closed IV	
Initial conditions								
Log (Ratio 90/10)	-0.197***	(0.005)	-0.175***	(0.005)	-0.087***	(0.004)	-0.177***	(0.011)
Log (GDPpc1800)	-0.749***	(0.006)	-0.698***	(0.006)	-0.500***	(0.006)	-0.683***	(0.009)
Institutional environment								
Log (IndexCreditProtection)	0.997***	(0.011)	0.799***	(0.011)	0.222***	(0.010)	0.707***	(0.011)
Region (reference group: Africa)								
Asia	0.206***	(0.010)	1.073***	(0.011)	1.625***	(0.010)	0.727***	(0.019)
Western Europe	0.209***	(0.010)	0.664***	(0.011)	1.325***	(0.010)	-0.004	(0.011)
Latin America	1.445***	(0.011)	1.541***	(0.012)	1.476***	(0.011)	1.310***	(0.012)
North America	0.892***	(0.010)	1.029***	(0.011)	1.443***	(0.010)	0.475***	(0.015)
Oceania	0.122***	(0.009)	0.570***	(0.010)	1.384***	(0.009)	-0.203***	(0.010)
Eastern Europe	0.297***	(0.010)	0.433***	(0.011)	0.880***	(0.010)	-0.018	(0.012)
Individual characteristics								
% of individuals with high school or more (at cohort level)	-0.142***	(0.006)	-0.345***	(0.006)	-0.317***	(0.005)	-0.121***	(0.026)
Cohort (Male aged 16-28 reference group)								
Male 29-38	0.138***	(0.004)	0.182***	(0.004)	1.143***	(0.004)	0.308***	(0.004)
Male 39-48	-0.105***	(0.004)	-0.136***	(0.004)	1.355***	(0.004)	0.273***	(0.004)
Male 49-58	-0.570***	(0.005)	-0.515***	(0.005)	1.237***	(0.004)	0.247***	(0.005)
Male 59-64	-1.456***	(0.006)	-1.420***	(0.006)	0.453***	(0.005)	0.074**	(0.037)
Female 16-28	-0.609***	(0.004)	-0.593***	(0.004)	-0.637***	(0.004)	-0.384***	(0.004)
Female 29-38	-0.464***	(0.004)	-0.347***	(0.004)	0.369***	(0.004)	-0.117***	(0.004)
Female 39-48	-0.686***	(0.004)	-0.687***	(0.004)	0.594***	(0.004)	-0.186***	(0.004)
Female 49-58	-1.172***	(0.005)	-1.172***	(0.004)	0.367***	(0.004)	-0.375***	(0.006)
Female 59-64	-2.200***	(0.009)	-2.136***	(0.009)	-0.333***	(0.005)	-0.553***	(0.013)
Year (reference: 2001)								
2002	-0.178***	(0.005)	-0.054***	(0.005)	0.105***	(0.005)	-0.346***	(0.015)
2003	0.023***	(0.005)	0.171***	(0.006)	0.257***	(0.005)	-0.106***	(0.020)
2004	-0.333***	(0.006)	-0.072***	(0.005)	0.248***	(0.005)	-0.223***	(0.018)
2005	-0.134***	(0.005)	0.004	(0.005)	0.349***	(0.005)	-0.308***	(0.019)
2006	-0.064***	(0.005)	0.147***	(0.005)	0.417***	(0.005)	-0.217***	(0.035)
2007	-0.114***	(0.005)	0.091***	(0.005)	0.435***	(0.005)	-0.238***	(0.023)
2008	-0.116***	(0.005)	0.166***	(0.005)	0.687***	(0.005)	-0.103***	(0.021)
2009	1.142***	(0.007)	1.418***	(0.007)	1.102***	(0.005)		
Constant	0.772***	(0.027)	0.085***	(0.028)	-1.636***	(0.026)	0.259***	(0.066)
No. Observations	959,199		942,535		973,873		914,094	
R-squared	0.509		0.506		0.603		0.469	
F test	31198.78***		31095.09***		30728.22***		27843.20***	
K-P Wald rk F statistic (weak identification test)	150,000***		130,000***		140,000***		150,000***	
Endogeneity test	5520***		3866.9***		150.286***		2591.045***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 2 IV Second Stage Pseudo-Panel Regression: Impact of inequality on job creation

	Nascent IV		Young IV		Established IV	
Initial conditions						
Log (Ratio 90/10)	0.605***	(0.126)	-0.304***	(0.066)	-0.165***	(0.024)
Log (GDPpc1800)	-0.792***	(0.127)	0.093	(0.060)	0.087***	(0.022)
Institutional environment						
Log(IndexCreditProtection) Total effect [1]	7.023***	(0.703)	1.996***	(0.283)	2.224***	(0.218)
*Ommited: Log(IndexCreditProtection)*Africa*						
Log(IndexCreditProtection)	0.529***	(0.162)	0.508***	(0.079)	0.332***	(0.051)
Log(IndexCreditProtection)*Asia	2.028***	(0.216)	0.514***	(0.090)	0.344***	(0.049)
Log(IndexCreditProtection)*Western Europe	1.354***	(0.147)	0.113*	(0.058)	0.352***	(0.045)
Log(IndexCreditProtection)*Latin America	1.941***	(0.196)	0.505***	(0.066)	0.436***	(0.044)
Log(IndexCreditProtection)*North America	-0.390	(0.284)	0.060	(0.085)	0.181***	(0.055)
Log(IndexCreditProtection)*Oceania	0.280**	(0.136)	-0.313***	(0.078)	0.118**	(0.047)
Log(IndexCreditProtection)*Eastern Europe	1.282***	(0.148)	0.609***	(0.058)	0.461***	(0.044)
Individual characteristics						
% of individuals with high school or more (at cohort level)	-0.669***	(0.117)	-0.114**	(0.051)	0.066***	(0.023)
Cohort (Male aged 16-28 reference group)						
Male 29-38	-0.119*	(0.063)	-0.073**	(0.030)	0.024	(0.020)
Male 39-48	0.118	(0.079)	0.013	(0.037)	0.066***	(0.019)
Male 49-58	-0.584***	(0.122)	-0.204***	(0.047)	-0.043*	(0.023)
Male 59-64	0.018	(0.164)	-0.171*	(0.091)	-0.270***	(0.030)
Female 16-28	-0.548***	(0.099)	-0.393***	(0.034)	-0.339***	(0.025)
Female 29-38	-0.802***	(0.079)	-0.308***	(0.029)	-0.369***	(0.021)
Female 39-48	-0.676***	(0.072)	-0.601***	(0.039)	-0.337***	(0.022)
Female 49-58	0.328	(0.212)	-0.359***	(0.103)	-0.472***	(0.025)
Female 59-64	-0.510*	(0.310)	-0.620***	(0.093)	-0.577***	(0.031)
Technology sector (reference: No/ Low technology sector)						
Medium or high	-0.003	(0.068)	0.069*	(0.036)	0.025	(0.020)
Sector (reference: Extractive sector)						
Transforming sector	0.095	(0.088)	0.057	(0.039)	0.064***	(0.015)
Business services	0.146	(0.092)	0.024	(0.041)	0.099***	(0.016)
Consumer oriented	0.041	(0.088)	0.030	(0.039)	0.014	(0.014)
Year (reference: 2001)						
2002	-0.378	(0.245)	-0.258***	(0.098)	0.146***	(0.025)
2003	0.831***	(0.270)	0.038	(0.094)	0.314***	(0.028)
2004	0.479**	(0.243)	-0.363***	(0.093)	0.084***	(0.023)
2005	0.465*	(0.255)	-0.147	(0.094)	0.188***	(0.024)
2006	0.063	(0.273)	-0.316***	(0.093)	0.170***	(0.024)
2007	0.009	(0.220)	-0.157*	(0.093)	0.141***	(0.025)
2008	-0.615**	(0.241)	-0.277***	(0.095)	0.132***	(0.025)
2009					0.436***	(0.029)
Constant	2.758***	(1.046)	0.660	(0.514)	0.168	(0.169)
No. Observations	6,952		22,119		53,067	
F test	933.11***		1833.33***		3332.82***	
K-P Wald rk F statistic (weak identification test)	27.24***		106.994***		317.925***	
Endogeneity test	28.58***		62.53***		489.05***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

APPENDIX

Table A.1 Independent variables

Country	GDPpc 1800	Ratio90/10 1800	Gini 2000	Ratio 90/10 Circa2000	GDPpc 2000	Freq.	Percent	Region
Argentina	871.63	6.00	47.90	10.73	5472.71	18,070	1.80	Latin America
Australia	671.48	7.75	44.60	.	30869.66	12,646	1.26	Oceania
Austria	1434.51	5.42	29.20	4.46	35027.30	4,199	0.42	Western Europe
Bosnia and Herzegovina	490.82	12.00	26.00	.	2241.70	4,016	0.40	Eastern Europe
Brazil	509.20	9.38	58.61	17.84	4406.71	20,000	2.00	Latin America
Canada	1159.50	7.75	32.40	5.59	33373.13	7,008	0.70	North America
Chile	702.10	6.00	59.50	16.03	6551.57	19,005	1.90	Latin America
China	985.89	5.22	40.30	6.79	1122.26	14,443	1.44	Asia
Colombia	522.98	9.38	57.22	20.99	3074.32	8,157	0.81	Latin America
Croatia	1227.06	12.00	30.00	4.26	8141.48	16,013	1.60	Eastern Europe
Czech Republic	1622.74	5.42	23.82	3.06	10378.64	2,001	0.20	Eastern Europe
Denmark	1342.84	6.19	35.00	8.47	45339.69	26,083	2.60	Western Europe
Egypt	748.79	5.50	54.20	12.02	1140.12	2,603	0.26	Africa
Finland	1037.69	6.19	27.00	3.59	33217.07	18,044	1.80	Western Europe
France	1388.32	6.11	28.20	3.75	32392.16	17,947	1.79	Western Europe
Germany	1695.68	5.42	29.20	4.50	32662.13	58,535	5.84	Western Europe
Greece	934.41	12.00	32.30	5.33	18040.57	13,970	1.39	Western Europe
Hungary	1390.67	5.42	24.96	3.44	8810.08	17,726	1.77	Eastern Europe
Iceland	801.90	6.19			46985.73	16,025	1.60	Western Europe
India	562.88	4.80	36.00	4.76	578.22	10,751	1.07	Asia
Indonesia	514.12	4.30	36.50	4.45	1086.05	2,000	0.20	Asia
Ireland	1213.17	10.00	34.10	5.06	41953.95	15,888	1.59	Western Europe
Italy	1339.84	6.11	35.80	7.08	29872.29	20,744	2.07	Western Europe
Japan	1055.06	6.75	31.88	.	33956.81	17,270	1.72	Asia
Korea	596.01	5.22	36.90	11.06	15162.35	8,023	0.80	Asia
Macedonia	800.76	12.00	35.14	9.12	2698.55	1,971	0.20	Eastern Europe
Mexico	1053.60	9.38	53.50	15.22	7689.10	9,448	0.94	Latin America
Netherlands	2412.43	5.92	25.50	3.50	37546.78	29,625	2.96	Western Europe
New Zealand	541.62	7.75	40.20	9.63	24260.06	8,868	0.89	Oceania
Norway	950.00	6.19	27.40	3.55	60726.25	19,921	1.99	Western Europe
Peru	697.30	9.38	49.62	14.41	2309.83	9,985	1.00	Latin America
Philippines	626.97	6.47	42.20	6.47	1060.55	2,000	0.20	Asia
Poland	1198.39	12.00	34.18	7.04	6824.47	6,001	0.60	Eastern Europe
Portugal	1284.41	6.11	34.70	5.00	17891.38	5,023	0.50	Western Europe
Romania		12.00	30.27	4.76	3326.62	6,218	0.62	Eastern Europe
Russia	823.99	12.00	48.40	9.02	3870.32	9,378	0.94	Eastern Europe
Serbia	1308.87	12.00	.	4.25	2588.74	6,776	0.68	Eastern Europe
Slovenia	1357.95	12.00	24.84	3.46	15033.47	21,138	2.11	Eastern Europe
South Africa	759.05	10.83	60.10	8.93	4652.34	24,865	2.48	Africa
Spain	1443.02	6.11	32.60	4.57	23920.93	158,307	15.81	Western Europe
Sweden	1100.00	6.19	27.20	3.91	36576.19	38,786	3.87	Western Europe
Switzerland	1612.48	5.92	31.80	.	50188.18	13,632	1.36	Western Europe
Taiwan	871.27	5.22	31.90	4.51	.	2,236	0.22	Asia
Thailand	496.98	6.47		33.33	2205.78	7,043	0.70	Asia
Turkey	869.92	9.58	39.80	.	6119.23	7,217	0.72	Asia
UK	2716.87	10.00	34.60	5.17	34058.66	197,518	19.72	Western Europe
United States	1912.62	6.84	40.10	8.93	40965.03	36,848	3.68	North America
Venezuela	442.02	9.38	45.80	11.61	5255.69	7,487	0.75	Latin America
Total						**1,001,458**	**100**	

Sources : Historical data on GDPpc Madisson's database, GDP per capita 2000 (constant 2005 US$) World Bank Indicators.

Table A.2 Variable definitions and sources

Variable notation	Definition	Source
Depedent variables		
Entrepreneurial stages:		
Nascent firms	% proportion of individuals involved in setting up a business they will own or co-own, but has not paid any payments for more than 3 months (in natural logarithms).	GEM
Young firms	% proportion of individuals that owners-manages firms, defined as having paid salaries for more than 3 months and less than 3.5 years (in natural logarithms).	GEM
Established firms	% proportion of individuals that owners-manages firms, defined as having paid salaries for more than3.5 year (in natural logarithms)s.	GEM
Closed firms	% proportion of individuals that owned-managed firms that in the past 12 months have been sold, shut down, discontinue or quit business (in natural logarithms).	GEM
Firm size at different stages:		
Nascent firms	Number of employees of nascent firms (Log transformation: 1+ number of jobs)	GEM
Young firms	Number of employees of young firms (Log transformation: 1+ number of jobs)	GEM
Established firms	Number of employees of established firms (Log transformation: 1+ number of jobs)	GEM
Independent variables		
Historical data		
Log (Ratio 90/10)	The 90/10 ratio measures the income of those individuals at the 90th and those at the 10th percentiles.Higher values of the ratio measures greatest income inequality.	Bourguignon and Morrison (2002)
Log (GDPpc1800)	Gross Domestic Product per capita in 1820	Angus Maddison's historic income database
Business environment		
Log(IndexCreditProtection)	Measures the degree of which collateral and bankruptcy laws protect the right of borrowers and lenders and thus facilitate lending. The index ranks from 0 to 10; higher scores indicating that collateral and bankruptcy laws are better designed to expand access to credit.	World Bank
Low_medium	Dummy variable: 1 if the country is classified as low or medium income country; 0 otherwise	Low-medium countries are those which mean gdp per capita for the considered period are below 13,000 USD dollars. Classification according to the World Bank.

Table A.2 Variable definitions and sources (cont.)

Variable notation	Definition	Source
Regional dummies		
Africa	Dummy variable: 1 Africa; 0 otherwise	Own classification
Asia	Dummy variable: 1 Asia; 0 otherwise	
Western Europe	Dummy variable: 1 Western Europe; 0 otherwise	
Latin America	Dummy variable: 1 Latin America; 0 otherwise	
North America	Dummy variable: 1 North America; 0 otherwise	
Oceania	Dummy variable: 1 Oceania; 0 otherwise	
Eastern Europe	Dummy variable: 1 Eastern Europe; 0 otherwise	
Individual variables at cohort levels		
% of individuals with high school or more (at cohort level)	Proportion of individuals in the cohort c with post-secondary level or more living in country *i in year* j	GEM
% of individual that provided credit to network (at cohort levels)	Proportion of individuals in cohort *c* that provided credit to others (excluding family members) living in country *i in year* j	GEM
Male aged 16-28	Proportion of males aged 16-38 years living in country *i in year j*	
Male 29-38	Proportion of males aged 29-38 years living in country *i in year j*	GEM
Male 39-48	Proportion of males aged 39-48 years living in country *i in year j*	GEM
Male 49-58	Proportion of males aged 49-58 years living in country *i in year j*	GEM
Male 59-64	Proportion of males aged 59-64 years living in country *i in year j*	GEM
Female 16-28	Proportion of females aged 16-28 years living in country *i in year j*	GEM
Female 29-38	Proportion of females aged 29-38 years living in country *i in year j*	GEM
Female 39-48	Proportion of females aged 39-48 years living in country *i in year j*	GEM
Female 49-58	Proportion of females aged 49-58 years living in country *i in year j*	GEM
Female 59-64	Proportion of females aged 59-64 years living in country *i in year j*	GEM
Sector		
Extractive sector	Dummy variable: 1 if the firm is involved in extractive activities; 0 otherwise	GEM
Transforming sector	Dummy variable: 1 if the firm is involved in transforming activities; 0 otherwise	GEM
Business services	Dummy variable: 1 if the firm is involved in business services; 0 otherwise	GEM
Consumer oriented	Dummy variable: 1 if the firm is involved in consumer oriented activities; 0 otherwise	GEM
Medium or high	Dummy variable: 1 if the firm is intensive in techonology sector (medium orhigh); 0 otherwise	GEM

Table A.2 Variable definitions and sources (cont.)

Variable notation	Definition	Source
Instrumental variables		
English Common Law	Dummy variable: 1 if the country has english legal origin; 0 otherwise	QOG The Quality of Government Institute (Original source: La Porta, López-de- Silanes, Shleifer & Vishny). http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
French Commercial Code	Dummy variable: 1 if the country hasfrench legal origin; 0 otherwise	QOG The Quality of Government Institute (Original source: La Porta, López-de- Silanes, Shleifer & Vishny). http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
Socialist/Communist Laws	Dummy variable: 1 if the country has socialist/communist legal origin; 0 otherwise	QOG The Quality of Government Institute (Original source: La Porta, López-de- Silanes, Shleifer & Vishny). http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
German Commercial Code	Dummy variable: 1 if the country has german legal origin; 0 otherwise	QOG The Quality of Government Institute (Original source: La Porta, López-de- Silanes, Shleifer & Vishny). http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
Scandinavian Commercial Code	Dummy variable: 1 if the country has scandinavian legal origin; 0 otherwise	QOG The Quality of Government Institute (Original source: La Porta, López-de- Silanes, Shleifer & Vishny). http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
colonia_spain		QOG The Quality of Government Institute. http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/
Blood pressure	The mean SBP (Systolic Blood Pressure) of the male population, counted in mm-Hg; this mean is calculated as if each country has the same age composition as the world population.	School of Public Health, Imperial College London.http://www1.imperial.ac.uk/publichealth/departments/ebs/projects/eresh/majidezzati/healthmetrics/metabolicriskfactors/
Colestherol	The mean SBP (Systolic Blood Pressure) of the male population, counted in mm-Hg; this mean is calculated as if each country has the same age composition as the world population.	School of Public Health, Imperial College London.http://www1.imperial.ac.uk/publichealth/departments/ebs/projects/eresh/majidezzati/healthmetrics/metabolicriskfactors/

Table A.3 Summary of main variables

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
% of people involved in									
Nascent firms	4.32	3.68	4.22	3.15	3.58	3.65	3.72	4.21	3.37
Young firms	2.75	2.97	3.17	2.77	3.21	3.53	3.58	3.63	3.09
Established firms	4.57	5.54	5.75	5.50	6.64	5.98	6.33	7.92	7.05
Closed firms		2.83	2.99	2.32	2.73	2.50	2.61	3.14	2.63
% of people									
Education high school or more	63.26	59.94	72.91	55.87	56.93	66.82	64.75	69.13	71.65
Provided credit to the network	0.78	1.05	0.98	0.81	0.95	0.98	1.14	1.16	0.96
Firm's size by entrepreneurial									
Nascent firms		2	7	3	4	3	4	11	11
Young firms		8	6	5	7	6	6	5	7
Established firms	8	13	15	9	11	9	10	10	10
Sector of activity									
Extractive sector	9.15	7.98	8.99	9.93	6.24	8.75	7.23	8.56	9.97
Transforming sector	29.19	28.80	27.36	30.48	26.86	31.74	28.83	28.16	24.12
Business services	21.23	22.00	22.74	21.24	21.42	17.24	21.57	19.07	15.19
Consumer oriented	40.42	41.23	40.91	38.35	45.49	42.26	42.37	44.20	50.71
Medium/high technology intensity	7.78	7.09	7.09	7.07	7.06	4.91	5.61	5.12	3.10
Obs.	62,598	115,418	92,228	140,537	110,870	171,465	153,657	133,793	156,825

Table A.4 Summary of main variables grouping by country GDP per capita (World Bank classification)

High-income countries

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
% of people involved in									
Nascent firms	3.4	3.1	3.5	2.5	3.0	2.8	3.0	3.1	2.4
Young firms	2.5	2.6	2.6	2.4	2.8	2.6	2.8	3.1	2.3
Established firms	4.5	5.6	5.5	5.4	6.9	5.1	5.9	8.0	6.7
Closed firms		2.3	2.2	1.8	1.7	1.7	1.7	1.9	1.9
% of people									
Education high school or more	70.3	65.5	76.6	58.7	61.5	70.7	63.7	72.5	0.8
Provided credit to the network	0.8	1.2	1.0	0.8	1.0	0.8	0.9	0.9	7.5
Firm's size by entrepreneurial stage									
Nascent firms	-	2	5	3	4	5	5	3	5
Young firms	-	6	5	4	10	8	8	6	4
Established firms	3	3	5	3	4	5	6	4	7.6
Sector of activity									
Extractive sector	10.73	8.97	10.71	11.16	6.98	9.17	7.9	9.26	25.2
Transforming sector	29.3	29.0	27.3	30.9	27.8	30.4	28.9	27.6	19.1
Business services	25.2	26.8	26.9	24.4	26.0	24.7	25.9	23.2	46.0
Consumer oriented	34.8	35.2	35.1	33.6	39.2	35.8	37.3	39.9	3.12
Medium/high technology intensity	9.42	9.02	7.01	7.6	7.97	5.58	6.52	5.83	
Obs.	48,754	87,073	79,610	118,375	84,489	125,443	113,242	79,718	104,391

Low income countries

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
% of people involved in									
Nascent firms	7.1	5.3	8.9	7.2	5.5	6.3	6.0	6.4	6.3
Young firms	3.4	3.9	6.6	4.9	4.5	6.4	6.2	4.6	5.4
Established firms	4.6	5.5	7.4	6.3	5.8	8.7	7.8	7.7	8.0
Closed firms		4.5	8.0	5.1	6.4	5.3	5.6	5.7	4.6
% of people									
Education high school or more	41.3	44.8	49.8	39.3	41.2	54.4	68.1	62.2	65.5
Provided credit to the network	0.8	0.8	1.0	0.8	0.9	1.7	1.9	1.6	1.5
Firm's size by entrepreneurial stage									
Nascent firms	-	1	9	2	3	3	4	8	3
Young firms	-	8	7	5	6	5	6	5	8
Established firms	8	7	7	6	6	7	7	5	7
Sector of activity									
Extractive sector	5.73	5.64	4.5	5.99	4.25	8.29	5.91	7.58	10.35
Transforming sector	28.9	28.3	27.6	29.1	24.3	33.3	28.7	28.9	22.8
Business services	12.7	10.7	11.9	11.2	9.3	8.9	13.0	13.2	10.1
Consumer oriented	52.7	55.4	56.0	53.7	62.2	49.5	52.4	50.3	56.8
Medium/high technology intensity	4.24	3.11	7.37	5.31	4.52	3.82	3.91	4.04	3.07
Obs.	13,844	28,345	12,618	22,162	26,381	46,022	40,415	54,075	52,434

Table A.5 Number of Observations per Cohort

Cohort	Freq.	Percent
<29male	118,663	11.85
>28male	87,396	8.73
>38male	82,135	8.2
>48male	70,088	7.0
>58male	107,228	10.71
<29female	121,738	12.16
>28female	106,129	10.6
>38female	98,491	9.83
>48female	82,431	8.23
>58female	127,159	12.7
Total	1,001,458	100

Table A.6 Firms' characteristics by entrepreneurial stages by region

	Africa	Asia	Western Europe	Latin America	North America	Oceania	Eastern Europe
Nascent firms							
Extractive sector	3.76	6.15	4.89	4.49	5.34	10.3	8.24
Transforming sector	23.94	22.25	24.2	23.85	22.43	24.83	30.29
Business services	11.38	9.21	26.8	10.63	27.77	27.24	21.42
Consumer oriented	60.92	62.39	44.11	61.03	44.46	37.62	40.05
Medium/High technology	4.16	5.3	7.79	5.83	7.6	11.2	6.35
Firm size	2	4	3	5	2	4	3
Young firms							
extractive sector	3.72	5.2	6.25	3.49	5.15	10.41	9.03
transforming sector	25.41	25.46	26.83	28.9	25.56	26.07	31.33
business services	14.26	9.08	26.93	11.8	29.87	30.36	22.83
consumer oriented	56.61	60.26	40	55.81	39.42	33.15	36.81
High tech	5.2	4.06	7.37	5.22	6.56	9	5.09
Firm size	4	11	5	5	7	4	10
Established firms							
extractive sector	5	9.79	12.37	5.33	11.36	18.35	12.99
transforming sector	31.6	27.41	30.34	32.09	24.76	30.43	34.82
business services	13.4	8.07	22.1	12.51	31.82	25.85	19.72
consumer oriented	50	54.72	35.18	50.07	32.07	25.37	32.47
High tech	2.73	3	5.87	3.89	5.38	7.5	6.24
Firm size	8	11	10	5	19	11	13

Table A.7 Pairwise correlations of main variables

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Nascent firms	1.00											
(2)	Young firms	0.76	1.00										
(3)	Established firms	0.37	0.48	1.00									
(4)	Firms closed	0.57	0.53	0.44	1.00								
(5)	Jobs nascent firms	-0.01	0.09	0.19	0.03	1.00							
(6)	Jobs young firms	0.17	0.15	0.13	0.13	0.07	1.00						
(7)	Jobs established firms	0.05	0.05	0.17	0.05	0.07	0.21	1.00					
(8)	Informal credit	0.47	0.42	0.33	0.48	0.17	0.26	0.15	1.00				
(9)	Ratio 90/10	-0.05	-0.14	-0.17	-0.06	-0.12	-0.11	-0.03	-0.23	1.00			
(10)	Credit	-0.21	-0.24	-0.13	-0.27	-0.08	0.01	0.25	-0.08	0.19	1.00		
(11)	GDP pc 1800	-0.37	-0.35	-0.17	-0.46	0.00	0.02	0.29	-0.26	0.17	0.59	1.00	
(12)	High school or more	0.09	0.04	-0.05	-0.04	-0.09	0.03	0.12	0.09	0.15	0.20	0.27	1.00

Table A.8 IV First Stage Pseudo-Panel Regression: Impact of inequality on firm's life cycle

	Nascent firms Log(Index CreditProtection)		**Young firms** Log(Index CreditProtection)		**Established firms** Log(Index CreditProtection)		**Closed firms** Log(Index CreditProtection)	
Initial conditions								
Log (Ratio 90/10)	-0.229***	(0.002)	-0.224***	(0.002)	-0.221***	(0.002)	-0.239***	(0.002)
Log (GDPpc1800)	0.284***	(0.001)	0.294***	(0.001)	0.291***	(0.001)	0.284***	(0.002)
Region (reference group: Africa)								
Asia	-0.240***	(0.005)	-0.249***	(0.005)	-0.232***	(0.005)	-0.244***	(0.006)
Western Europe	-0.136***	(0.005)	-0.146***	(0.005)	-0.145***	(0.005)	-0.111***	(0.005)
Latin America	-0.392***	(0.005)	-0.401***	(0.005)	-0.389***	(0.005)	-0.378***	(0.005)
North America	-0.325***	(0.005)	-0.332***	(0.005)	-0.325***	(0.005)	-0.300***	(0.006)
Oceania	0.155***	(0.004)	0.158***	(0.004)	0.163***	(0.004)	0.171***	(0.004)
Eastern Europe	-0.031***	(0.005)	-0.046***	(0.005)	-0.035***	(0.005)	-0.011**	(0.005)
Year (reference: 2001)								
2002	-0.013***	(0.002)	-0.021***	(0.002)	-0.012***	(0.002)	-0.006*	(0.003)
2003	-0.105***	(0.002)	-0.119***	(0.002)	-0.106***	(0.002)	-0.092***	(0.004)
2004	-0.059***	(0.002)	-0.053***	(0.002)	-0.046***	(0.002)	-0.046***	(0.004)
2005	-0.056***	(0.002)	-0.057***	(0.002)	-0.045***	(0.002)	-0.040***	(0.004)
2006	-0.055***	(0.002)	-0.056***	(0.002)	-0.045***	(0.002)	-0.050***	(0.007)
2007	-0.057***	(0.002)	-0.069***	(0.002)	-0.043***	(0.002)	-0.036***	(0.005)
2008	0.036***	(0.002)	0.024***	(0.002)	0.052***	(0.002)	0.063***	(0.004)
2009	-0.036***	(0.002)	-0.039***	(0.003)	-0.033***	(0.003)	0.469***	(0.001)
Individual characteristics								
% of individuals with high school or more (at cohort level)	0.143***	(0.002)	0.159***	(0.002)	0.156***	(0.002)	0.135***	(0.005)
Cohort (Male aged 16-28 reference group)								
Male 29-38	0.003	(0.002)	0.003	(0.002)	0.006***	(0.002)	0.003*	(0.002)
Male 39-48	0.017***	(0.002)	0.020***	(0.002)	0.016***	(0.002)	0.015***	(0.002)
Male 49-58	0.031***	(0.002)	0.040***	(0.002)	0.033***	(0.002)	0.030***	(0.002)
Male 59-64	0.058***	(0.002)	0.059***	(0.002)	0.064***	(0.002)	0.036***	(0.007)
Female 16-28	0.002	(0.002)	0.002	(0.002)	0.001	(0.002)	0.004***	(0.002)
Female 29-38	0.002	(0.002)	0.004**	(0.002)	0.006***	(0.002)	0.001	(0.002)
Female 39-48	0.020***	(0.002)	0.024***	(0.002)	0.017***	(0.002)	0.016***	(0.002)
Female 49-58	0.034***	(0.002)	0.045***	(0.002)	0.037***	(0.002)	0.038***	(0.002)
Female 59-64	0.068***	(0.002)	0.063***	(0.002)	0.077***	(0.002)	0.048***	(0.003)
Legal origin (reference: other legal origin)								
English	0.458***	(0.001)	0.448***	(0.001)	0.454***	(0.001)	0.469***	(0.001)
Constant	0.229***	(0.010)	0.153***	(0.010)	0.143***	(0.010)	0.213***	(0.015)
No. Observations	959,199		942,535		973,873		914,094	
R-squared	0.526		0.535		0.518		0.536	
Shea Partial R2	0.1458		0.1404		0.1441		0.1507	
Partial R2	0.1458		0.1404		0.1441		0.1507	
F statistic test excluded instruments	150,000***		130,000***		140,000***		150,000***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.9.1 IV First Stage Pseudo-Panel Regression: Impact of inequality on job creation in nascent firms

	Log(IndexCredit Protection)		Asia* Log(IndexCredit Protection)		Western Europe* Log(IndexCredit Protection)		Latin America* Log(IndexCredit Protection)		North America* Log(IndexCredit Protection)		Oceania* Log(IndexCredit Protection)		Eastern Europe* Log(IndexCredit Protection)	
Initial conditions														
Log (Ratio 90/10)	-0.114***	(0.022)	-0.619***	(0.024)	0.206***	(0.027)	-0.101***	(0.016)	-0.192***	(0.017)	0.062***	(0.021)	0.339***	(0.027)
Log (GDPpc1800)	0.107***	(0.012)	-0.059***	(0.018)	1.102***	(0.017)	0.005	(0.006)	0.260***	(0.018)	-1.191***	(0.026)	-0.126***	(0.015)
Individual characteristics														
% of individuals with high school or more	0.176***	(0.020)	0.123***	(0.023)	-0.409***	(0.025)	0.180***	(0.020)	0.094***	(0.014)	-0.074***	(0.024)	0.138***	(0.019)
Cohort (Male aged 16-28 reference group)														
Male 29-38	0.017	(0.011)	0.098***	(0.016)	0.100***	(0.017)	-0.011	(0.007)	-0.082***	(0.013)	-0.036**	(0.017)	-0.036***	(0.014)
Male 39-48	-0.004	(0.011)	0.085***	(0.018)	0.093***	(0.020)	-0.043***	(0.009)	-0.050***	(0.014)	-0.025	(0.015)	0.013	(0.020)
Male 49-58	0.045**	(0.018)	0.072***	(0.023)	0.101***	(0.022)	0.044***	(0.011)	-0.083***	(0.014)	0.035	(0.025)	-0.051***	(0.017)
Male 59-64	0.027	(0.026)	0.037	(0.029)	0.054	(0.039)	-0.001	(0.013)	-0.014	(0.016)	-0.021	(0.027)	0.057**	(0.028)
Female 16-28	0.018	(0.017)	-0.025	(0.017)	0.192***	(0.023)	0.082***	(0.010)	-0.087***	(0.013)	-0.004	(0.018)	-0.120***	(0.025)
Female 29-38	0.007	(0.012)	0.146***	(0.018)	0.178***	(0.019)	0.032***	(0.006)	-0.100***	(0.019)	-0.107***	(0.022)	-0.062***	(0.016)
Female 39-48	0.043***	(0.011)	0.040**	(0.020)	0.032	(0.020)	0.050***	(0.011)	-0.051***	(0.013)	0.039*	(0.021)	0.037**	(0.018)
Female 49-58	-0.067***	(0.022)	0.003	(0.040)	-0.034	(0.045)	-0.029***	(0.011)	0.174***	(0.052)	-0.087	(0.053)	-0.026	(0.022)
Female 59-64	0.018	(0.028)	0.166***	(0.047)	0.189***	(0.059)	-0.047**	(0.021)	-0.151***	(0.022)	0.049*	(0.029)	-0.113***	(0.033)
Technology sector (reference: No/ Low technology sector)														
Medium or high	-0.028**	(0.014)	-0.002	(0.016)	-0.002	(0.019)	-0.011	(0.010)	0.016	(0.015)	-0.012	(0.017)	-0.011	(0.017)
Sector (reference: Extractive sector)														
Transforming sector	-0.033**	(0.015)	-0.059***	(0.022)	0.008	(0.024)	-0.010	(0.011)	-0.018	(0.016)	-0.042*	(0.024)	0.066***	(0.019)
Business services	-0.021	(0.015)	-0.031	(0.022)	0.014	(0.025)	-0.014	(0.011)	-0.025	(0.017)	-0.008	(0.024)	0.030	(0.020)
Consumer oriented	-0.028**	(0.014)	-0.028	(0.021)	0.016	(0.023)	-0.009	(0.010)	-0.021	(0.015)	-0.044**	(0.022)	0.043**	(0.018)
Year (reference: 2001)														
2002	-0.042	(0.028)	-0.214***	(0.036)	-0.505***	(0.043)	-0.011	(0.015)	0.132***	(0.020)	-0.264***	(0.032)	0.568***	(0.032)
2003	-0.198***	(0.029)	-0.248***	(0.035)	-0.326***	(0.044)	-0.179***	(0.021)	0.193***	(0.023)	-0.142***	(0.031)	0.404***	(0.032)
2004	0.002	(0.027)	-0.262***	(0.037)	-0.425***	(0.045)	-0.035**	(0.016)	0.087***	(0.018)	-0.060**	(0.027)	0.503***	(0.032)
2005	0.001	(0.027)	-0.357***	(0.035)	-0.312***	(0.043)	-0.039***	(0.015)	0.137***	(0.020)	-0.194***	(0.031)	0.484***	(0.031)
2006	-0.000	(0.027)	-0.220***	(0.038)	-0.434***	(0.050)	-0.053***	(0.016)	0.147***	(0.021)	-0.161***	(0.031)	0.572***	(0.038)
2007	-0.004	(0.028)	-0.104***	(0.033)	-0.218***	(0.042)	0.029*	(0.015)	0.076***	(0.018)	-0.119***	(0.030)	0.386***	(0.035)
2008	0.311***	(0.030)	-0.235***	(0.036)	-0.085*	(0.044)	0.206***	(0.020)	0.203***	(0.022)	-0.240***	(0.029)	0.511***	(0.037)
Legal origin (reference: English)														
French	-0.673***	(0.012)	-0.030	(0.020)	0.302***	(0.021)	0.009	(0.010)	-0.247***	(0.018)	-0.515***	(0.022)	0.078***	(0.015)
Socialist/Communist	-0.322***	(0.023)	-0.009	(0.024)	-0.955***	(0.025)	0.010	(0.008)	-0.139***	(0.015)	-0.004	(0.023)	1.311***	(0.032)
German	-0.267***	(0.016)	0.422***	(0.036)	0.412***	(0.031)	-0.043***	(0.010)	-0.213***	(0.017)	-0.644***	(0.022)	-0.100***	(0.019)
Scandinavian	-0.569***	(0.012)	-0.103***	(0.020)	0.731***	(0.021)	-0.021**	(0.009)	-0.137***	(0.013)	-0.760***	(0.021)	-0.094***	(0.015)
Colonial origin (reference: other colonial origins or never colonized by a western oversea)														
Spain	-0.041***	(0.011)	-0.663***	(0.015)	-0.684***	(0.017)	1.351***	(0.013)	0.138***	(0.012)	-0.150***	(0.012)	0.030**	
Blood pressure	-0.017***	(0.002)	-0.051***	(0.003)	0.057***	(0.003)	-0.016***	(0.001)	-0.038***	(0.002)	-0.055***	(0.002)	0.027***	(0.001)
Colestherol	0.425***	(0.030)	-1.094***	(0.036)	0.315***	(0.029)	0.178***	(0.017)	-0.015	(0.014)	1.743***	(0.036)	0.321***	(0.025)
Constant	1.568***	(0.171)	14.398***	(0.266)	-15.835***	(0.289)	1.253***	(0.115)	3.710***	(0.251)	7.206***	(0.313)	-5.585***	(0.203)
No. Observations	6,952		6,952		6,952		6,952		6,952		6,952		6,952	
R-squared	0.722		0.799		0.893		0.914		0.301		0.646		0.772	
Partial R2 of excluded instruments	0.63		0.7021		0.7708		0.8788		0.2216		0.5490		0.7344	
Shea R2	0.1942		0.1268		0.2161		0.1522		0.1107		0.2703		0.2599	
F statistic test excluded instruments	1117.91***		1474.49***		1931.21***		7744.93***		35.89***		262.46***		361.21***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.9.2 IV First Stage Pseudo-Panel Regression: Impact of inequality on job creation in young firms

	Log(IndexCredit Protection)		Asia* Log(IndexCredit Protection)		Western Europe* Log(IndexCredit Protection)		Latin America* Log(IndexCredit Protection)		North America* Log(IndexCredit Protection)		Oceania* Log(IndexCredit Protection)		Eastern Europe* Log(IndexCredit Protection)	
Initial conditions														
Log (Ratio 90/10)	-0.028**	(0.012)	-0.559***	(0.014)	0.119***	(0.019)	0.172***	(0.010)	0.038***	(0.010)	-0.215***	(0.011)	0.398***	(0.017)
Log (GDPpc1800)	0.193***	(0.008)	-0.211***	(0.011)	0.956***	(0.012)	-0.139***	(0.005)	0.329***	(0.012)	-0.677***	(0.020)	-0.208***	(0.007)
Individual characteristics														
% of individuals with high school or more (at cohort level)	0.145***	(0.012)	0.065***	(0.015)	-0.337***	(0.017)	-0.009	(0.012)	0.201***	(0.013)	-0.015	(0.017)	0.163***	(0.012)
Cohort (Male aged 16-28 reference group)														
Male 29-38	0.012	(0.008)	0.001	(0.011)	-0.002	(0.013)	0.002	(0.007)	-0.006	(0.010)	0.012	(0.012)	0.001	(0.009)
Male 39-48	0.017*	(0.009)	-0.039***	(0.013)	0.002	(0.014)	0.010	(0.008)	0.014	(0.013)	0.019	(0.015)	0.005	(0.011)
Male 49-58	0.042***	(0.011)	-0.049***	(0.017)	0.040**	(0.020)	-0.020**	(0.009)	-0.005	(0.016)	0.061***	(0.018)	-0.002	(0.014)
Male 59-64	0.037	(0.031)	-0.089***	(0.022)	-0.122***	(0.037)	0.026	(0.026)	0.120***	(0.040)	0.064	(0.045)	0.085***	(0.030)
Female 16-28	-0.001	(0.009)	-0.026*	(0.014)	-0.008	(0.016)	0.012	(0.008)	0.015	(0.014)	0.012	(0.018)	-0.005	(0.010)
Female 29-38	0.023***	(0.008)	-0.037***	(0.011)	0.030**	(0.014)	0.026***	(0.008)	0.007	(0.012)	0.022	(0.014)	-0.029***	(0.011)
Female 39-48	0.042***	(0.032)	-0.028**	(0.023)	-0.021	(0.042)	0.025**	(0.036)	0.029*	(0.038)	0.040**	(0.045)	0.009	(0.030)
Female 49-58	0.033	(0.028)	-0.056***	(0.020)	-0.080***	(0.027)	-0.035**	(0.016)	0.079***	(0.021)	0.024	(0.019)	0.044	(0.028)
Female 59-64	0.041	(0.032)	-0.143***	(0.024)	-0.108***	(0.041)	0.044	(0.035)	0.178***	(0.041)	0.063	(0.044)	0.064**	(0.030)
Technology sector (reference: No/ Low technology sector)														
Medium or high	-0.017*	(0.010)	0.019	(0.013)	0.041**	(0.017)	-0.010	(0.009)	-0.019	(0.015)	0.002	(0.016)	-0.042***	(0.012)
Sector (reference: Extractive sector)														
Transforming sector	-0.030***	(0.011)	0.001	(0.013)	-0.007	(0.016)	0.040***	(0.008)	-0.007	(0.013)	-0.044***	(0.016)	-0.032**	(0.013)
Business services	-0.030***	(0.011)	-0.009	(0.014)	-0.066***	(0.017)	0.013*	(0.008)	0.036**	(0.015)	-0.005	(0.018)	-0.020	(0.013)
Consumer oriented	-0.043***	(0.011)	0.049***	(0.013)	0.011	(0.016)	0.024***	(0.008)	-0.016	(0.012)	-0.086***	(0.016)	-0.037***	(0.012)
Year (reference: 2001)														
2002	-0.104***	(0.032)	-0.110***	(0.024)	-0.378***	(0.039)	0.022	(0.028)	0.251***	(0.045)	0.057	(0.046)	0.069**	(0.030)
2003	-0.197***	(0.032)	-0.072***	(0.024)	-0.139***	(0.040)	-0.131***	(0.035)	0.130***	(0.042)	0.118***	(0.045)	-0.018	(0.029)
2004	-0.054*	(0.032)	-0.101***	(0.023)	-0.309***	(0.039)	0.027	(0.028)	0.086**	(0.039)	0.151***	(0.045)	0.124***	(0.029)
2005	-0.090***	(0.032)	-0.136***	(0.024)	-0.092**	(0.039)	-0.054*	(0.029)	0.089**	(0.040)	0.084*	(0.045)	0.051*	(0.029)
2006	0.012	(0.031)	0.018	(0.023)	-0.181***	(0.039)	-0.031	(0.029)	0.103**	(0.041)	-0.009	(0.044)	0.190***	(0.031)
2007	-0.050	(0.032)	0.081***	(0.024)	-0.010	(0.039)	0.009	(0.030)	-0.042	(0.041)	-0.053	(0.045)	0.073**	(0.030)
2008	0.094***	(0.033)	-0.210***	(0.023)	-0.125***	(0.039)	0.004	(0.029)	0.153***	(0.040)	0.017	(0.043)	0.245***	(0.032)
Legal origin (reference: English)														
French	-0.584***	(0.008)	-0.234***	(0.012)	0.367***	(0.013)	0.205***	(0.006)	-0.342***	(0.011)	-0.417***	(0.015)	0.071***	(0.007)
Socialist/Communist	-0.297***	(0.016)	0.090***	(0.014)	-0.689***	(0.020)	0.057***	(0.007)	-0.253***	(0.012)	-0.160***	(0.011)	1.236***	(0.018)
German	-0.196***	(0.010)	0.356***	(0.022)	0.409***	(0.020)	0.068***	(0.006)	-0.389***	(0.014)	-0.577***	(0.018)	0.090***	(0.010)
Scandinavian	-0.316***	(0.010)	-0.099***	(0.013)	1.005***	(0.016)	-0.008	(0.006)	-0.214***	(0.011)	-0.848***	(0.025)	-0.139***	(0.010)
Colonial origin (reference: other colonial origins or never colonized by a western oversea)														
Spain	-0.028***	(0.008)	-0.425***	(0.008)	-0.541***	(0.010)	1.059***	(0.012)	0.122***	(0.007)	-0.099***	(0.006)	0.004	(0.006)
Blood pressure	-0.012***	(0.001)	-0.049***	(0.001)	0.058***	(0.002)	-0.009***	(0.001)	-0.060***	(0.002)	-0.011***	(0.001)	0.020***	(0.001)
Colestherol	0.436***	(0.019)	-0.702***	(0.019)	0.500***	(0.019)	0.184***	(0.012)	-0.060***	(0.012)	0.873***	(0.026)	0.477***	(0.015)
Constant	0.089	(0.104)	13.001***	(0.160)	-15.860***	(0.189)	0.761***	(0.078)	5.837***	(0.172)	2.515***	(0.177)	-4.628***	(0.127)
No. Observations	22,119		22,119		22,119		22,119		22,119		22,119		22,119	
R2	0.654		0.675		0.828		0.700		0.417		0.447		0.776	
Partial R2 of excluded instrument	0.4664		0.5409		0.6715		0.6359		0.3438		0.3585		0.7125	
Shea R2	0.1596		0.1368		0.2216		0.2294		0.1577		0.1835		0.3205	
F statistic test excluded instruments	1654.21***		2501.37***		4193.59***		6516.38***		240.8***		193.72***		1118.16***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.9.3 IV First Stage Pseudo-Panel Regression: Impact of inequality on job creation in established firms

	Log(IndexCredit Protection)		Asia* Log(IndexCredit Protection)		Western Europe* Log(IndexCredit Protection)		Latin America* Log(IndexCredit Protection)		North America* Log(IndexCredit Protection)		Oceania* Log(IndexCredit Protection)		Eastern Europe* Log(IndexCredit Protection)	
Initial conditions														
Log (Ratio 90/10)	-0.030***	(0.007)	-0.487***	(0.010)	0.096***	(0.012)	0.140***	(0.006)	0.045***	(0.007)	-0.161***	(0.007)	0.287***	(0.011)
Log (GDPpc1800)	0.184***	(0.005)	-0.165***	(0.007)	0.940***	(0.009)	-0.121***	(0.004)	0.305***	(0.008)	-0.697***	(0.012)	-0.193***	(0.006)
Individual characteristics														
% of individuals with high school or more (at cohort level)	0.105***	(0.008)	0.094***	(0.010)	-0.414***	(0.011)	-0.058***	(0.007)	0.254***	(0.009)	-0.078***	(0.011)	0.217***	(0.008)
Cohort (Male aged 16-28 reference group)														
Male 29-38	0.009	(0.007)	-0.010	(0.010)	0.003	(0.012)	0.004	(0.006)	0.005	(0.010)	0.014	(0.010)	-0.009	(0.008)
Male 39-48	0.016**	(0.007)	0.001	(0.010)	-0.021*	(0.012)	0.007	(0.006)	0.017*	(0.010)	0.009	(0.010)	0.008	(0.008)
Male 49-58	0.025***	(0.007)	0.007	(0.011)	-0.053***	(0.012)	-0.009	(0.006)	0.028***	(0.010)	0.013	(0.010)	0.030***	(0.008)
Male 59-64	0.017*	(0.010)	0.026*	(0.016)	-0.155***	(0.017)	-0.020**	(0.008)	0.055***	(0.014)	0.059***	(0.015)	0.067***	(0.010)
Female 16-28	0.009	(0.009)	-0.013	(0.013)	0.017	(0.017)	0.009	(0.008)	0.016	(0.013)	0.017	(0.015)	-0.021*	(0.012)
Female 29-38	0.006	(0.008)	-0.026**	(0.011)	0.014	(0.013)	0.006	(0.007)	0.016	(0.011)	0.020*	(0.011)	-0.013	(0.009)
Female 39-48	0.013	(0.008)	-0.030***	(0.010)	-0.025*	(0.013)	-0.005	(0.007)	0.058***	(0.011)	0.023**	(0.011)	-0.004	(0.009)
Female 49-58	0.015	(0.011)	-0.023*	(0.012)	-0.063***	(0.015)	-0.010	(0.008)	0.068***	(0.012)	0.021*	(0.012)	0.013	(0.013)
Female 59-64	0.026**	(0.011)	0.011	(0.017)	-0.186***	(0.019)	-0.018*	(0.009)	0.083***	(0.015)	0.096***	(0.016)	0.048***	(0.011)
Technology sector (reference: No/ Low technology sector)														
Medium or high	-0.018**	(0.007)	-0.007	(0.012)	0.046***	(0.013)	-0.021***	(0.006)	-0.016	(0.011)	-0.005	(0.011)	-0.003	(0.009)
Sector (reference: Extractive sector)														
Transforming sector	-0.045***	(0.005)	0.003	(0.007)	-0.063***	(0.009)	0.028***	(0.004)	0.002	(0.007)	-0.048***	(0.008)	0.011*	(0.006)
Business services	-0.052***	(0.006)	-0.027***	(0.008)	-0.115***	(0.010)	0.015***	(0.004)	0.042***	(0.008)	-0.012	(0.009)	0.021***	(0.007)
Consumer oriented	-0.059***	(0.005)	0.031***	(0.007)	-0.033***	(0.008)	0.006	(0.004)	-0.006	(0.007)	-0.088***	(0.008)	0.017***	(0.006)
Year (reference: 2001)														
2002	-0.070***	(0.008)	-0.136***	(0.009)	0.047***	(0.013)	-0.004	(0.005)	0.010	(0.015)	0.053***	(0.013)	-0.078***	(0.009)
2003	-0.138***	(0.009)	-0.107***	(0.010)	0.290***	(0.015)	-0.130***	(0.010)	-0.106***	(0.014)	0.134***	(0.014)	-0.200***	(0.011)
2004	-0.041***	(0.007)	-0.126***	(0.009)	0.050***	(0.012)	-0.002	(0.006)	-0.099***	(0.012)	0.124***	(0.013)	-0.011	(0.008)
2005	-0.064***	(0.008)	-0.182***	(0.010)	0.296***	(0.013)	-0.083***	(0.007)	-0.110***	(0.012)	0.086***	(0.013)	-0.096***	(0.008)
2006	0.021***	(0.007)	-0.027***	(0.010)	0.104***	(0.012)	-0.024***	(0.005)	-0.091***	(0.013)	0.019	(0.012)	0.038***	(0.010)
2007	-0.012	(0.009)	-0.001	(0.011)	0.339***	(0.013)	-0.009	(0.007)	-0.208***	(0.012)	-0.014	(0.012)	-0.077***	(0.008)
2008	0.134***	(0.008)	-0.238***	(0.011)	0.268***	(0.013)	0.011**	(0.006)	-0.059***	(0.013)	0.017	(0.010)	0.071***	(0.009)
2009	0.060***	(0.010)	-0.171***	(0.014)	0.462***	(0.017)	0.029***	(0.008)	-0.120***	(0.015)	0.001	(0.014)	-0.148***	(0.010)
Legal origin (reference: English)														
French	-0.629***	(0.005)	-0.210***	(0.009)	0.277***	(0.009)	0.165***	(0.003)	-0.329***	(0.008)	-0.413***	(0.009)	0.051***	(0.004)
Socialist/Communist	-0.298***	(0.011)	0.014	(0.012)	-0.859***	(0.014)	0.045***	(0.005)	-0.231***	(0.008)	-0.166***	(0.008)	1.328***	(0.013)
German	-0.194***	(0.006)	0.563***	(0.014)	0.202***	(0.013)	0.053***	(0.003)	-0.375***	(0.009)	-0.565***	(0.011)	0.050***	(0.006)
Scandinavian	-0.283***	(0.005)	-0.060***	(0.008)	0.966***	(0.010)	0.006*	(0.003)	-0.202***	(0.007)	-0.798***	(0.014)	-0.140***	(0.006)
Colonial origin (reference: other colonial origins or never colonized by a western oversea)														
Spain	-0.009	(0.007)	-0.477***	(0.006)	-0.557***	(0.007)	1.098***	(0.009)	0.096***	(0.005)	-0.098***	(0.004)	0.030***	(0.003)
Blood pressure	-0.015***	(0.001)	-0.048***	(0.001)	0.056***	(0.001)	-0.007***	(0.001)	-0.052***	(0.001)	-0.014***	(0.001)	0.022***	(0.001)
Colestherol	0.417***	(0.013)	-0.859***	(0.016)	0.451***	(0.014)	0.157***	(0.008)	-0.051***	(0.009)	0.944***	(0.018)	0.403***	(0.012)
Constant	0.650***	(0.071)	13.199***	(0.103)	-15.395***	(0.126)	0.725***	(0.057)	5.085***	(0.111)	2.690***	(0.103)	-4.236***	(0.099)
No. Observations	53,067		53,067		53,067		53,067		53,067		53,067		53,067	
R2	0.645		0.685		0.817		0.713		0.387		0.463		0.815	
Partial R2 of excluded instruments	0.5023		0.5688		0.6787		0.6563		0.3249		0.3823		0.7674	
Shea R2	0.1616		0.1533		0.1763		0.2154		0.1406		0.1959		0.2178	
F statistic test excluded instruments	4580.40***		4786.64***		11390.30***		12558.59***		441.32***		511.07***		4015.29***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.10 IV Second Stage Pseudo-Panel Regression: Impact of inequality on firm's life cycle excluding self-employed

	(1) Nascent IV		(2) Young IV		(3) Established IV		(4) Closed IV	
Initial conditions								
Log (Ratio 90/10)	-0.175***	(0.005)	-0.151***	(0.005)	-0.091***	(0.004)	-0.176***	(0.011)
Log (GDPpc1800)	-0.747***	(0.006)	-0.746***	(0.007)	-0.498***	(0.006)	-0.681***	(0.009)
Institutional environment								
Log (IndexCreditProtection)	1.066***	(0.011)	0.829***	(0.012)	0.224***	(0.011)	0.734***	(0.011)
Region (reference group: Africa)								
Asia	0.197***	(0.011)	1.141***	(0.012)	1.616***	(0.010)	0.749***	(0.021)
Western Europe	0.163***	(0.010)	0.722***	(0.011)	1.323***	(0.010)	-0.002	(0.011)
Latin America	1.497***	(0.012)	1.623***	(0.012)	1.474***	(0.011)	1.348***	(0.012)
North America	0.859***	(0.010)	0.998***	(0.011)	1.440***	(0.010)	0.429***	(0.016)
Oceania	0.119***	(0.010)	0.537***	(0.010)	1.384***	(0.009)	-0.212***	(0.010)
Eastern Europe	0.292***	(0.011)	0.515***	(0.011)	0.879***	(0.010)	0.013	(0.013)
Individual characteristics								
% of individuals with high school or more (at cohort level)	-0.095***	(0.006)	-0.319***	(0.006)	-0.307***	(0.005)	-0.126***	(0.027)
Cohort (Male aged 16-28 reference group)								
Male 29-38	0.151***	(0.004)	0.199***	(0.004)	1.142***	(0.004)	0.317***	(0.004)
Male 39-48	-0.082***	(0.004)	-0.094***	(0.004)	1.354***	(0.004)	0.288***	(0.004)
Male 49-58	-0.551***	(0.005)	-0.477***	(0.005)	1.234***	(0.005)	0.248***	(0.006)
Male 59-64	-1.440***	(0.007)	-1.414***	(0.007)	0.451***	(0.005)	0.095**	(0.038)
Female 16-28	-0.621***	(0.004)	-0.623***	(0.004)	-0.637***	(0.004)	-0.392***	(0.004)
Female 29-38	-0.477***	(0.004)	-0.339***	(0.004)	0.367***	(0.004)	-0.122***	(0.004)
Female 39-48	-0.679***	(0.004)	-0.682***	(0.004)	0.594***	(0.004)	-0.190***	(0.005)
Female 49-58	-1.148***	(0.005)	-1.157***	(0.005)	0.368***	(0.005)	-0.383***	(0.007)
Female 59-64	-2.181***	(0.009)	-2.116***	(0.009)	-0.333***	(0.005)	-0.551***	(0.013)
Year (reference: 2001)								
2002	-0.214***	(0.005)	-0.092***	(0.005)	0.106***	(0.005)	-0.312***	(0.016)
2003	0.008	(0.006)	0.129***	(0.006)	0.257***	(0.005)	-0.085***	(0.021)
2004	-0.466***	(0.006)	-0.115***	(0.005)	0.250***	(0.005)	-0.212***	(0.018)
2005	-0.157***	(0.005)	-0.033***	(0.005)	0.352***	(0.005)	-0.290***	(0.019)
2006	-0.112***	(0.005)	0.106***	(0.006)	0.415***	(0.005)	-0.211***	(0.036)
2007	-0.112***	(0.005)	0.055***	(0.005)	0.436***	(0.005)	-0.215***	(0.023)
2008	-0.123***	(0.005)	0.114***	(0.005)	0.688***	(0.005)	-0.095***	(0.022)
2009	1.127***	(0.007)	1.378***	(0.008)	1.106***	(0.005)		
Constant	0.571***	(0.029)	0.252***	(0.029)	-1.651***	(0.027)	0.152**	(0.069)
No. Observations	929,305		913,781		937,281		889,726	
R-squared	0.497		0.506		0.594		0.464	
F-test	30162.72***		30154.75***		29253.65***		27052.57***	
K-P rk F statistic (weak identification test)	140,000***		130,000***		120,000***		150,000***	
Endogeneity test	5636.82***		3675.33***		779.85***		2451.33***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.11 IV Second Stage Pseudo-Panel Regression: Impact of inequality on job creation excluding self-employed

	Nascent IV		Young IV		Established IV	
Initial conditions						
Log (Ratio 90/10)	0.511***	(0.130)	-0.532***	(0.072)	-0.341***	(0.021)
Log (GDPpc1800)	-1.270***	(0.157)	0.076	(0.066)	-0.039**	(0.019)
Institutional environment						
Log(IndexCreditProtection) Total effect [1]	6.252***	(0.606)	1.984***	(0.284)	1.883***	(0.143)
*Ommited: Log(IndexCreditProtection)*Africa*						
Log(IndexCreditProtection)	0.158	(0.175)	0.318***	(0.080)	0.314***	(0.039)
Log(IndexCreditProtection)*Asia	1.278***	(0.165)	0.396***	(0.094)	0.127***	(0.033)
Log(IndexCreditProtection)*Western Europe	1.142***	(0.107)	0.163***	(0.059)	0.306***	(0.031)
Log(IndexCreditProtection)*Latin America	1.258***	(0.151)	0.333***	(0.070)	0.192***	(0.027)
Log(IndexCreditProtection)*North America	0.888**	(0.430)	0.248***	(0.093)	0.275***	(0.045)
Log(IndexCreditProtection)*Oceania	0.099	(0.125)	-0.063	(0.084)	0.198***	(0.032)
Log(IndexCreditProtection)*Eastern Europe	1.430***	(0.164)	0.589***	(0.055)	0.472***	(0.030)
Individual characteristics						
% of individuals with high school or more (at cohort level)	-0.327**	(0.134)	-0.109*	(0.056)	-0.110***	(0.016)
Male 29-38	0.457***	(0.089)	0.024	(0.040)	0.128***	(0.012)
Male 39-48	0.257**	(0.125)	-0.400***	(0.058)	0.011	(0.016)
Male 49-58	0.178	(0.228)	-0.155	(0.101)	-0.149***	(0.031)
Male 59-64	-0.280**	(0.116)	-0.467***	(0.039)	-0.301***	(0.016)
Female 16-28	-0.490***	(0.083)	-0.270***	(0.032)	-0.332***	(0.012)
Female 29-38	-0.614***	(0.080)	-0.466***	(0.043)	-0.300***	(0.015)
Female 39-48	-0.034	(0.280)	-0.654***	(0.056)	-0.465***	(0.018)
Female 49-58	-0.028	(0.380)	-0.752***	(0.106)	-0.450***	(0.031)
Female 59-64						
Technology sector (reference: No/ Low technology sector)						
Medium or high	-0.001	(0.070)	0.053	(0.037)	0.014	(0.014)
Sector (reference: Extractive sector)						
Transforming sector	0.055	(0.100)	0.082*	(0.045)	0.037***	(0.013)
Business services	0.227**	(0.106)	0.071	(0.047)	0.048***	(0.014)
Consumer oriented	0.138	(0.098)	0.041	(0.044)	0.002	(0.012)
Year (reference: 2001)						
2002	-0.529*	(0.306)	-0.358***	(0.109)	0.085***	(0.019)
2003	0.418	(0.326)	-0.214**	(0.102)	0.292***	(0.020)
2004	0.343	(0.300)	-0.605***	(0.102)	0.035**	(0.018)
2005	0.067	(0.317)	-0.298***	(0.105)	0.128***	(0.018)
2006	-0.175	(0.329)	-0.411***	(0.105)	0.260***	(0.018)
2007	-0.250	(0.274)	-0.205**	(0.104)	0.266***	(0.018)
2008	-0.836***	(0.312)	-0.403***	(0.106)	0.121***	(0.018)
2009					0.440***	(0.031)
Constant	7.128***	(1.277)	1.694***	(0.573)	1.841***	(0.155)
No. Observations	5,432		19,691		85,057	
R-squared	0.63		0.78		0.89	
F test	587.4***		1581.86***		5222.46***	
K-P Wald rk F statistic (weak identification test)	18.08***		92.39***		563.33***	
Endogeneity test	24.02***		44.04***		420.48***	

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.12 IV Second Stage Pseudo-Panel Regression: Impact of inequality on firm's life cycle using alternative inequality indicators

	(1) Nascent IV	(2) Young IV	(3) Established IV	(4) Closed IV
Initial conditions				
Log (Gini)	-0.495***	0.036***	-0.401***	-0.237***
	(0.017)	(0.016)	(0.015)	(0.085)
Log (Top90)	-0.678***	-0.941***	-0.385***	-0.870***
	(0.010)	(0.010)	(0.009)	(0.019)
Log (Middle 50)	0.686***	0.358***	1.169***	1.052***
	(0.025)	(0.025)	(0.022)	(0.097)
Log (Bottom 10)	0.052***	0.014***	-0.008***	-0.012
	(0.004)	(0.004)	(0.004)	(0.008)
Log (Top20/Bottom20)	-0.083***	-0.018***	-0.061***	-0.033***
	(0.003)	(0.003)	(0.003)	(0.008)
Log(Middle)	1.830***	0.699***	1.746***	1.116***
	(0.034)	(0.035)	(0.031)	(0.153)
No. Observations	959,199	942,535	973,873	914,094

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01
Top90 is the income share of the 9th decile relative to the income share of the 1st decile
Middle 50 is the income share of the 5th decile relative to the mean income
Bottom 10 is the income share of the 1st decile relative to the mean income
Top20/Bottom20 is the income share of the 8th decile relative to the 2nd decile
Middle is the income share of the middle class, defined as the income share of the 2nd to 4th quintiles.
Control variables as in Table 1.

Table A.13 IV Second Stage Pseudo-Panel Regression: Impact of inequality on job creation using alternative inequality indicators

	(1) Nascent IV	(2) Young IV	(3) Established IV
Initial conditions			
Log (Gini)	-1.385***	-1.402***	-0.989***
	(0.322)	(0.165)	(0.069)
Log (Top90)	3.441***	0.731***	0.703***
	(0.228)	(0.136)	(0.042)
Log (Middle 50)	-3.288***	2.688***	1.956***
	(0.601)	(0.267)	(0.136)
Log (Bottom 10)	0.382***	0.383***	0.321***
	(0.097)	(0.040)	(0.017)
Log (Top20/Bottom20)	0.105	-0.310***	-0.211***
	(0.082)	(0.039)	(0.016)
Log(Middle)	2.059***	2.687***	1.955***
	(0.001)	(0.267)	(0.135)
No. Observations	6,952	22,119	53,067

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01
Top90 is the income share of the 9th decile relative to the income share of the 1st decile
Middle 50 is the income share of the 5th decile relative to the mean income
Bottom 10 is the income share of the 1st decile relative to the mean income
Top20/Bottom20 is the income share of the 8th decile relative to the 2nd decile
Middle is the income share of the middle class, defined as the income share of the 2nd to 4th quintiles
Control variables as in Table 1.

Table A.14 Second Stage Pseudo-Panel Regression: Firm's life cycle using alternative instrumental variables

	(1) Nascent	(2) Young	(3) Established	(4) Closed
Panel a) IV: Language				
Key independent variables				
Log (Ratio 90/10)	-0.258***	-0.292***	-0.263***	-0.178***
Log (IndexCreditProtection)	2.028***	2.334 ***	2.486***	0.676***
First stage summary results				
K-P Wald rk F statistic (weak identification test)	905.36***	971.81***	980.37***	582.48***
Endogeneity test	1072.696***	1412.251***	1965.14***	15.701***
Shea partial R2	0.0041	0.0045	0.0043	0.0057
Partial R2	0.0041	0.0045	0.0043	0.0057
Panel b) IV: Religion				
Key independent variables				
Log (Ratio 90/10)	-0.072***	-0.871***	-0.0797***	-0.135***
Log (IndexCreditProtection)	0.684***	0.871***	0.164***	0.0218***
First stage summary results				
K-P Wald rk F statistic (weak identification test)	8005.066***	8019.8***	7955***	4872.42***
Endogeneity test	1952.341***	1233.588***	143.8***	27.58***
Shea partial R2	0.0257	0.0268	0.0254	0.0246
Partial R2	0.0257	0.0268	0.0254	0.0246

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01
Control variables as in Table 1.

Table A.15 Second Stage Pseudo-Panel Regression: Impact of Inequality on job creation using alternative instrumental variables

	(1) Nascent	(2) Young	(3) Established
IV Language			
Log (Ratio 90/10)	-12.07	-11.9	-1.32***
Log (IndexCreditProtection)	1.88	5.92	2.45***
IV Religion			
Log (Ratio 90/10)	-0.160	-2.13***	-1.25***
Log (IndexCreditProtection)	4.8***	2.61***	2.43***

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01
Control variables as in Table 2.

Table A.16 Summary results instrumental variable: Language. Job creation

IV- Language	Log(IndexCredit Protection)	Asia* Log(IndexCredit Protection)	Western Europe* Log(IndexCredit Protection)	Latin America* Log(IndexCredit Protection)	North America* Log(IndexCredit Protection)	Oceania* Log(IndexCredit Protection)	Eastern Europe* Log(IndexCredit Protection)
Nascent firms							
Shea partial R2	0.0001	0.0027	0.0012	0.0001	0.0005	0.0003	0.0004
Partial R2	0.6535	0.5708	0.7054	0.443	0.2106	0.5451	0.7644
F test excluded instruments	1259.43	782.57	1524.8	581.64	34.86	250.78	411.78
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**0.013**						
Endogeneity test	47.05***						
Young firms							
Shea partial R2	0.0002	0.0009	0.0014	0.0009	0.0633	0.0004	0.0532
Partial R2	0.486	0.4802	0.631	0.3229	0.3374	0.3564	0.7437
F test excluded instruments	1547.47	2268.34	3513.56	907.55	236.86	206.85	1241.32
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**0.12**						
Endogeneity test	89.67***						
Established firms							
Shea partial R2	0.0036	0.0199	0.0265	0.0179	0.0653	0.0092	0.1675
Partial R2	0.515	0.5096	0.6474	0.2788	0.3206	0.3804	0.7856
F test excluded instruments	4321.54	4282.36	8815.39	1401.52	444.34	524.48	4546.76
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**5.607**						
Endogeneity test	226.05***						

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.17 Summary results instrumental variable: Religion. Job creation

	Log(IndexCredit Protection)	Asia* Log(IndexCredit Protection)	Western Europe* Log(IndexCredit Protection)	Latin America* Log(IndexCredit Protection)	North America* Log(IndexCredit Protection)	Oceania* Log(IndexCredit Protection)	Eastern Europe* Log(IndexCredit Protection)
Nascent firms							
Shea partial R2	0.0226	0.2565	0.5256	0.0806	0.0501	0.0509	0.2264
Partial R2	0.6616	0.6466	0.707	0.3622	0.2272	0.7039	0.7465
F test excluded instruments	1356.44	1634.28	1501.46	402.36	36.8	524.48	386.8
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**8.39**						
Endogeneity test	112.553***						
Young firms							
Shea partial R2	0.0056	0.0488	0.0636	0.0199	0.1554	0.0131	0.2274
Partial R2	0.4981	0.5336	0.6312	0.2493	0.3696	0.4997	0.7233
F test excluded instruments	2032.3	2670.9	3471.06	869.56	264.88	370.33	1047.77
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**5.88**						
Endogeneity test	68.16***						
Established firms							
Shea partial R2	0.0041	0.037	0.0452	0.0116	0.0906	0.0113	0.2621
Partial R2	0.5256	0.5571	0.6396	0.2281	0.3493	0.5185	0.777
F test excluded instruments	4597.68	4558.24	9061.98	1668.55	489.86	932.3	4128.77
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K-P Wald rk F statistic (weak identification test)	**7.52**						
Endogeneity test	189.66***						

Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Essay 3
Schooling progression in Uruguay: Why some children are left behind?

Schooling progression in Uruguay: Why some children are left behind?

Abstract

This study examines the factors that differently affect children's educational path in Uruguay. Specifically, I focus on the effects of long-term parental income crystallized by cognitive and non-cognitive abilities, parental educational background and race, and short-term family income proxied by the opportunity cost of education, on children's schooling progression in Uruguay.

For this purpose, I use a sequential probability model which allows me to analyze the factors affecting the dynamics of the children's educational path. The results show that long-term parental income is the main factor influencing schooling attainment while short-term family income has decreasing effects over the children's education path. Specifically, parental educational background, race, cognitive and non-cognitive abilities have effects of diverse magnitude across stages of schooling progression. I find that cognitive ability, measured by repetition, has long-lasting effects on children's education attainment. Motivation and risky behavior measuring non-cognitive ability also influence children's schooling completion at early stages of education.

These findings call for public intervention focused on improving cognitive and non-cognitive abilities to enable children attaining higher education, particularly those from disadvantaged parental backgrounds.

3.1 Introduction

It is well known in the literature that children's parental background plays a major role in explaining educational inequality. Several studies have shown that children of well-off parents generally receive more and better schooling and benefit from material, cultural and genetic inheritances (Checchi, 2006). Heckman and coauthors refer to the long-term family factors reflected by: parental educational background, children scholastic ability, motivation, self-esteem, as important sources of disparities across individuals' educational attainment. In turn, these sources of disparities in education, may well translate into other individuals' economic outcomes, such as earnings. As long as large differences exist in educational opportunities, individuals will have different chances of success in life.

In turn, attaining a level of education is something that happens over a long period of time and it is split into different schooling stages, like finishing primary education, completing secondary level, and so on. Therefore, knowing the influence of parental background variables at each stage of the schooling transition can give a more complete picture of how inequality of education attainment came about. Each of the alternative sources of inequality pointed by the literature call for specific policy prescriptions at different stages of the schooling progression, which may well have different effects on equity and efficiency of the education system and subsequent labor market outcomes.

The objective of this paper is to analyze to what extent intergenerational transmission of parental traits takes place for children's educational attainment in Uruguay. Specifically, this paper aims to study whether parental education, race, child's scholastic ability, motivation and risky behavior as measures of socio-emotional endowments, and short-term family income proxied by the opportunity cost of education, are key determinants of individuals' educational path decisions and, if they are, at what stage in the schooling process they take on their importance.[1]

Uruguay is a particularly interesting country to analyze this issue for many reasons. First, it stands out in the Latin American region because of a large tradition of publicly provided education and social inclusion. For instance, primary school was made compulsory in 1877, universal primary schooling was achieved in the 1950s

[1] In this study cognitive ability, scholastic ability and performance in different educational levels are used as synonyms, while socio-emotional endowments and non-cognitive ability are used interchangeably.

(Manacorda, 2008). In addition, the system provides free access to educational institutions in all schooling levels; in postsecondary education university is publicly provided, students do not need to pay any fee or perform any entrance test; one feature that distinguish Uruguay from others countries of the region. Also, the country ranks among the highest in the region in terms of its socioeconomic indicators, presenting the lowest poverty rate and income inequality indicators in the region (Panorama Social de America Latina, Cepal, 2012).

However, the Uruguayan education system shows major shortcomings. In the Latin American context, while the proportion of population aged 18 to 29 living in urban areas with complete secondary in 2000 is less than 20% in Uruguay; this rate is 40% in Chile and 30% in Paraguay (SITEAL, 2005). Chile presents one of the highest indicators of income inequality and is characterized by a private education system especially at the university level; while Paraguay ranks below Uruguay in terms of the Human Development Index. In this line, several studies stress that the Uruguayan educational system is unable to retain a large share of students in lower high school (Furtado, 2003; da Silveira and Queirolo, 1998), picture that worsens when educational attainment across afro and non-afro descendants is analyzed.[2] Therefore, a relevant question is why despite the great offer of public education, children living in Uruguay do not attain higher levels of education. This is what makes Uruguay an interesting case study.

The contributions of this paper are twofold. First, it contributes to the recent literature developed by Bowles and Gintis (2001, 2002) and Heckman and co-authors by addressing the importance of cognitive and non-cognitive abilities, parental educational background, and race, on young people' (or their parents) educational choices in a middle income country such as Uruguay. Indeed, empirical studies exploring the impacts of multiple abilities on education attainment are scarce and mainly focused on developed countries, while less usual for developing countries mainly because of data availability. In this sense, the rich dataset used in this paper enables me to exploit information on motivation (measured as motives reported for secondary enrollment) and risky behavior such as adolescence use of marijuana, two factors pointed out in the literature as important ones reflecting socio-emotional factors,

[2] See Table 1.

and in turn affecting education attainment (see for instance Heckman et al., 2006; Heckman et al., 2014; Gullone and Moore, 2000).

Second, by exploiting the sequential process of education attainment, it is possible to identify different impacts of the key variables over the individual's educational path. Specifically, by analyzing the effect of parental educational background, multiple abilities and race; and the opportunity cost of education at different decision points in the schooling transition process, it is possible to distinguish between long and short-term family income affecting schooling; to disentangle a direct effect of these key variables on the educational level attained, but also an indirect effect to the extent that parental background affects previous educational choices.

Therefore, this study goes beyond previous analyses on education focused on developing countries by saying that measures of cognition are important predictors of child's outcomes, and by recognizing the different effects of diverse abilities across the individual's schooling transition in a middle-income country such as Uruguay.

This paper uses a unique micro-dataset elaborated by the Uruguayan Statistics Institute: the Youth National Survey (ENAJ: *Encuesta Nacional de Adolescencia y Juventud*), a cross-sectional national representative survey on adolescence and youth conducted in 2008. The sample is based on the same households interviewed in the Continuous Household Survey (ECH: *Encuesta Continua de Hogares*) for 2008, thus being possible to merge the information from both surveys. Detailed information on socio-demographic characteristics, migration trajectories, educational history, risky behaviors, parental education, among others, is provided. In addition, the retrospective information contained in this dataset allows me to construct educational trajectories, as well as early behaviors of interest for theoretical ages of participation in the education system.

The empirical strategy considers a dynamic educational model developed by Cameron and Heckman (1998, 2001) in which schooling attainment is modeled as the outcomes of sequential choices made at each educational level using probability models and conditional on previous educational choices. In turn, the model accounts for individual unobserved heterogeneity, such as ability or motivation, which may affect individuals' schooling progression.

The results suggest that long-term family factors greatly influences child's schooling transitions. Students with more favorable parental educational backgrounds and with better performance in the educational system are more likely to survive higher

schooling stages. Race is an important factor preventing schooling progression for girls and, to a lesser extent for boys. Less motivated individuals and with risky behaviors are less likely to survive early schooling stages and therefore, to attain higher education. In addition, short-term family income, measured as the opportunity cost of education at each schooling level, has decreasing effects across the educational path; turning less important -in comparison to long-term family factors- the higher we move on the educational path.

These findings are in line with the literature, which suggests that early child's life cycle is a sensitive period for the formation of cognitive skills and has persistent effects on higher stages of the schooling transition. Also, non-cognitive ability, despite data limitations for its measurement, is seen to be an important factor affecting schooling progression. Thus, our results call for public interventions focused on cognitive and non-cognitive abilities at different stages of the life cycle in order to compensate children from disadvantaged parental backgrounds.

The remainder of this paper is organized as follows. The next section presents an overview of the literature on education, specially focusing on the literature of cognitive and non-cognitive abilities. Section 3 describes the Uruguayan educational system. Section 4 introduces the data and presents descriptive analysis. Section 5 describes the econometric methods. Section 6 presents and discusses the main findings of the study. Finally, Section 7 concludes.

3.2 Education inequality, cognitive and non-cognitive abilities

This paper is broadly related to the literature analyzing the different factors that influence individual's education attainment. It is well known in the literature that higher education is positively correlated with higher income. Thus, a pertinent question is why individuals choose different levels of education. Checchi (2006) stresses that families are often unaware of the economic benefit of education or are prevented from sending their children to further education by their financial needs.

The literature on education has focused on the factors preventing full access to education and individuals' educational choices, such as individual's unobservable abilities, family cultural background, family financial resources, public resources, residential choices and social capital.[3] Most of these factors exhibit intertemporal and

[3] See Checchi (2006) for an exhaustive overview of the literature on Economics of Education.

intergenerational persistence. Therefore, alternative intergenerational transmission channels are identified in the literature, which in turn calls for specific policy recommendations.

Within this line of research, the literature of inequality of opportunity analyzes the different factors influencing education attainment. The most accepted concept of inequality of opportunity refers to the notion that inequalities which are brought about by individual' circumstances, like gender, ethnicity and race, place of birth, family background, which are beyond the individual's control, are considered ethically unacceptable, while inequality resulting from individual's effort and choice are ethically accepted (Roemer, 1998). This definition requires that any inequality attributed to the influence of exogenous circumstances should be reduced, compensated by public interventions.

Based on this framework, several empirical studies address the alternative mechanisms through which intergenerational transmission may operate by estimating the relationship between one individual's educational attainment and her parental education, income, or occupation; controlling for other child's circumstances like race and gender, among others (as in Bourgignon et al., 2003; Ferreira and Gignoux, 2008; Peragine and Serlenga, 2007, among others). Therefore, the coefficient relating parental background and a child's outcome measures the intergenerational transmission of an attribute from one generation to the other. For the Uruguayan case, González and Sanromán (2010) find persistent effects of parental educational background on education attainment for afro and non afro-descendants. In turn, Porzecanski (2008) studies the determinants of the educational gap between afro and non afro-descendants in Uruguay analyzing the impact of family background on repetition in primary level, and dropouts of adolescents in the educational system.

In this study, I follow an alternative framework developed by Heckman and coauthors (Cameron and Heckman, 2001; Heckman and Carneiro, 2003; Cuhna and Heckman, 2007), which considers the total effect of family background on education attainment. Specifically, these authors refer to long-term family factors including long-term levels of family income, reflected by parental education, scholastic ability, motivation, time preferences, risk aversion and self-esteem, as important factors shaping later success in life, which in turn may explain sources of disparities across individuals' education attainment. Also, short-term family income influences individual's education attainment.

Specifically, Cameron and Heckman (2001) find that short-term family income effects are weakened most in the later schooling transitions, playing no role in college entry decisions. To the extent that the influence of long-term family income measured at a point in time is diminished by the inclusion of cognitive abilities or family background variables, the authors conclude that long-term family factors crystallized in these variables are the driving forces behind schooling attainment, and not short-term credit constraints experienced in the late adolescent years.[4]

In turn, these authors analyze the educational level attained by one individual as a sequential process, in which the individual chooses the educational level conditional on having completed the previous educational level. By doing so, it is possible to examine the different effects of variables of interest on individual's educational attainment, and to do so at different stages of the educational path.

Previous studies have followed this empirical strategy, mainly focused on developed countries for which adolescent and youth panel datasets with information on individuals' educational path and past performance are largely available (Cameron and Heckman, 2001 for US; Holm and Jaeger, 2011, and Blanden et al., 2002 for the UK; multiple tracks choices in the educational path for the Danish case in Karlson, 2011, and for Germany in Dustmann et al., 2004). Also, cross sectional data containing information on past performance in the educational system allows Bernardi (2012) to analyze schooling transitions in Spain.[5] The one exception for Latin American countries within this framework is found in Pal (2004) for the Peruvian case in which, using 1994 Peru Living Standards Measurement Study data, analyses the impact of parental background and individual's ability on individuals' schooling transitions.

In general, these empirical studies measure ability with previous performance in the educational system, such as repetition or test scores. These measures have been criticized by recent literature. Indeed, the literature has recognized that abilities are multiple in nature and that previous studies using IQ or previous performance (repetition, test scores) measures does not properly account for ability.

For instance, Bowles and Gintis (2002) stress that *"inheritance process operating through superior cognitive performance and educational attainments of those*

[4] Note that this framework differentiates from the Inequality of Opportunity framework since it does not distinguish between individuals' circumstances and efforts. Specifically, Cameron and coauthors only point out that abilities reflect long-term parental income.

[5] This framework is also extended in Cappellari (2004) for the analysis of individual's transitions between the type of high school chosen (private or public) and university enrollment and school-to-work transition using a cross sectional sample of high school leavers in Italy.

well-off parents, while important, explain at most half of the intergenerational transmission of economic status. Moreover, while genetic transmission of earnings-enhancing traits appears to play a role, the genetic transmission of IQ appears to be relatively unimportant". These authors conclude that empirical studies on intergenerational transmission of economic status have over-studied education and cognitive abilities, while other individual characteristics such as wealth, race and non-cognitive behavioral traits have been under-studied.

Unlike other personal traits such as height or weight, personality traits cannot be directly measured. Non-cognitive abilities, such as perseverance, motivation, risk aversion, self-esteem, self-control, have direct effects on wages (after controlling for schooling), schooling, performance on achievement tests, and other aspects of social and economic life. The most widely accepted taxonomy of personality traits is the Big Five model defined as: Conscientiousness ("the tendency to be organized, responsible and hardworking"), Openness to Experience ("the tendency to be open to new aesthetic, cultural, or intellectual experiences"), Extraversion ("an orientation of one interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability"), Agreeableness ("the tendency to act in a cooperative, unselfish manner"), and Neuroticism/Emotional Stability (Emotional stability is "predictability and consistency in emotional reactions, with the absence of rapid mood changes"; Neuroticism is "a chronic level of emotional instability and proneness to psychological distress").[6] Overall, observed productivities, efforts, and actions are used to infer traits using conventional factor analysis in which the tests are measures of different domains of personality based on observer reports or self-report.[7]

Although the relationship between personality traits and education has not received much attention, mainly due to data availability, a certain consensus emerges in the literature. Perseverance and preferences related to an interest in learning, two traits which are related to Consciousness and Openness to Experience, increase the likelihood of individuals' attaining more years of schooling (Lundberg, 2013; Almlund et al., 2011). In turn, Heckman, Stixrud and Urzua (2006) find that locus of control and self-

[6] See Table 1.3 (p45) in Almlund et al. (2011) for a comprehensive definition of the Big Five Domains, facets and related traits.

[7] The Big Five model is not without its critics. The main ones stress that the model is atheoretical; omits individual's motivation (what people value or desire), while other psychologists suggest that the categories are too crude to be useful; or the luck of consensus among researchers about identifying and organizing lower order facets of the Big Five factors (Almlund et al., 2011).

esteem (traits related to Neuroticism) play an important role for adolescents schooling decisions, having different effects across schooling levels. Nonetheless, data availability often determines which measure of non-cognitive skills is used in empirical analysis (Brunello and Schlotter, 2011). One possibility for overcoming data limitations or surveys without good questions on personality type is found within the psychological literature on personality traits and adolescent risk-behavior. Gullone and Moore (2000) identify different categories of risky behaviors traits, two of them -rebellious and reckless risk-taking- were found to be negatively correlated with Consciousness.[8] Following the psychological literature, Heckman et al. (2011) and Heckman et al. (2014) propose to use behaviors that have proved to be strongly correlated with Consciousness and Agreeableness, namely: violent behavior such as fighting at school or work and hitting or threatening to hit someone, tried marijuana, daily smoking, regular drinking, and any intercourse before age 15; measures of socio-emotional factors that affect schooling progression.

Overall, this framework stresses those both cognitive and non-cognitive abilities, as part of long-term parental background, jointly with parental education, race/ethnicity, and other family characteristics, play an important role on the multiple periods in the individuals' life cycle. The existence of critical and sensitive periods of childhood in skill formation and different roles played by cognitive abilities and socio-emotional factors across an individual's life cycle calls for different policies in time (Heckman and Mosso, 2014). For instance, parental inputs have different effects at different stages of the child's life cycle with cognitive skills affecting more at early ages and non-cognitive skills affecting more at later ages (Cunha and Heckman, 2008). In turn, both cognitive and non-cognitive skills can be shaped by interventions and that there are effective margins for social policy (see Heckman and Mosso, 2014; Heckman, Pinto, and Savelyev, 2013).

[8] Examples of rebellious risk-taking are drinking, smoking, and staying out at night. Examples of reckless risk-taking are drinking and driving, having unprotected sex, and speeding.

3.3. The Uruguayan Educational System

The educational system is organized in four levels: pre-school, primary education (grades 1-6, with theoretical ages 6 to 11), secondary level which includes lower high school (*Ciclo básico*, grades 7 to 9, theoretical ages 12-14) and upper high school (*Bachillerato*, grades 10 to 12, theoretical ages 15-17); and tertiary level (university and teaching training institutes). Primary and lower high school levels are compulsory.[9] Lower and upper high school are offered in both *liceos* (non-vocational secondary schools), and in vocational schools (*UTUs*). The different schooling stages are both public and privately provided (see Figure 1).

Table 1 presenting schooling progression by gender and race for the population aged 20 to 29 shows one of the major caveats of the educational system. While enrollment in primary is timely and completion of primary education almost universal, the system fails in retaining a large share of students at different schooling stages.

It is worth noting the great fall in the proportion of people completing each level across the educational system. In particular, low enrollment rates in postsecondary (20.5% for the total sample, in Table 1) may be explained by the low proportion of people completing previous education levels. Note for instance the low proportion of young people with complete lower high school or complete upper high school (64.5% and 29.2% respectively for the final sample, in Table 1). Differences between afro and non afro-descendants are also striking. In particular, 5% of afro-descendant males and 13.7% of afro descendant women have complete secondary education, compared to 28.8% and 36.4% for non afro-descendant men and women respectively (Table 1).

Some main features that characterize the educational system in Uruguay are provided in Table A.1 in the Appendix. In particular, it is highlighted the great proportion of population aged 12 to 29 who is or was enrolled in a public institution at different levels of schooling stages. Nonetheless, notice that the proportion of students in a private institution increases for higher levels of education. Also, students largely choose general education institutions (*Liceos* or *Bachilleratos*).

An important feature which deserves to be highlighted is the low supply of tertiary education institution located in the Interior of the country.[10] The main University in Uruguay is the Universidad de la República (UdelaR), which is public and

[9] Since 2008 upper high school and pre-school are compulsory. Ley General de Educación No. 18.347

[10] *Interior* is commonly used to identify the regions of the country excluding Montevideo, the capital of Uruguay, and includes 18 Departments.

freely provided, meaning that students do not have to pay any fee or pass any entrance test. But the UdelaR is mainly located in Montevideo, the capital of Uruguay, so students wanting to enroll in college and not living in Montevideo need to migrate to the capital. Also private colleges are mainly located in Montevideo. This may prevent many students without financial family support to access college.[11]

3.4. Data and descriptive statistics

This study uses the National Youth Survey (*ENAJ, Encuesta Nacional de Adolescencia y Juventud*), a cross-sectional national representative survey on adolescents and youth conducted in 2008 by the Uruguayan National Statistics Institute (*Instituto Nacional de Estadística*). The survey universe consists on all adolescents and young Uruguayans between 12 and 29 years living in cities larger than 5,000 inhabitants. In total the original survey comprises 4,993 individuals. Since the sample is based on the same households interviewed in the Continuous Household Survey (ECH) for 2008, information from both surveys can be merged.

The ECH is one of the main sources of statistical information in the country providing socioeconomic information at the Department level, Montevideo and rest of the country (commonly named *Interior*, containing 18 Departments).[12] It offers detailed information on socio-demographic characteristics, migration trajectories, labor market participation, experience, and conditions, among others. The ENAJ incorporates rich information about other aspects of the respondents lives, such as education attainment, fertility, participation in crime, self-assessed health status, family background (mother and father education), alcohol and substance use, migration, first job characteristics, participation in social and political organizations.

Although the survey is not longitudinal, it contains retrospective information allowing me to construct educational trajectories, as well as past performance in the educational system (repetition in primary and secondary level), motives for attending secondary level, and risky behaviors of interest (such as drug consumption). Moreover, the ENAJ complements ECH in the sense that the latter suffers from major limitations of special importance to the analyses of the role of family background on college

[11] It is worth mentioning that since 2007 the UdelaR has being making great efforts in terms of territorial decentralization in order to give major opportunities to those students living in the Interior of the country. Also, some private universities are starting to locate in different regions of the country.
[12] A Department is a first-level political and administrative division of Uruguay.

enrollment decisions. For instance, the ECH contains information of the family background only for those individuals living in the origin household, while not information is provided for those who moved out. Then, studies based on the ECH may suffer from endogeneity issues, due to the possible sample selection of those individuals who left the household of origin (see Francesconi and Nicoletti, 2006). In addition, the ECH does not provide information on educational past history, such as repetition in primary and secondary level.[13] The ENAJ allows me not only to address the above mentioned issues, but also to take into account an individual's educational history and exploit information on motivation and risky behaviors.

The original sample is restricted to individuals aged 20 to 29, theoretical ages for which individuals are supposed to have completed at least secondary education. This restriction enables me to observe different educational transitions since the child enters the educational system until the higher attained level. After excluding observations with missing data on key interest variables, I obtain a final sample of 2,349 individuals.

Table 2 provides summary statistics for the final sample, and by gender and race. More than half of the sample is female (52%) while the proportion of afro-descendants is 11%.[14]

A first difference is observed between afro and non-afro descendants in terms of their parental educational backgrounds. For instance, the proportion of non-afro descendants with high educated parents (more than 12 years) doubles afro-descendants rate, while the proportion of afro-descendants with low educated parents is 20% higher than for non afro-descendants.

It is worth mentioning that pre-school enrollment, despite not being compulsory for the population considered covered a large proportion of the total sample (more than 80%). Primary education is almost universal (98% of the total sample completes this level), however a big concern refers to the high repetition rates observed for the whole sample (25% of children repeated at least once in primary), rate that worsens for afro-descendants (41%) almost doubling non afro-descendants' (22%).

[13] One exception is the Extended National Household Survey (ENHA: Encuesta Nacional de Hogares Ampliada) carried only on 2006, an extended survey with a specific module on education.

[14] Afro-descendance is captured in the ECH through the following question: *"Do you believe you have… (black or afro, Asian, white, native, other) descent?"*. The respondent can choose more than one option of racial descent. For this study, individuals reporting having black or afro descent are classified as afro-descendants. Non-afro descendants are all individuals reporting not having afro-descent (thus, including whites, Asian, native or other). It is worth noting that almost 90% declares *only* white descent, while less than 5% declares having native or other descent.

A second difference arises across genders when observing performance in primary level in which girls do better than boys (21% of girls repeated at least once versus 27% of the males). Tables 3 and 4 present summary statistics for different schooling levels for girls and boys respectively. Some observations can be made from these tables.

First, the proportion of children dropping out at each educational level is mainly from lower parental background (representing more than 70% in lower high school, and more than 40% in upper high school) In addition, it is observed that, while the proportion of students from disadvantaged parental educational background enrolled is lower at higher levels of schooling, the proportion of children from better-off parental educational background completing lower and upper high school and enrolled in postsecondary increases. The share of children from medium parental background enrolled and completing each level is stable across the educational path. These frequencies suggest that in Uruguay, transitions turn more selective for boys and girls from less advantaged parental educational background.

Second, afro-descendants are more likely to drop out in lower and upper high school than non afro-descendants. Especially for girls, the proportion of afro-descendants that drops out at each stage is more than twice the proportion of those enrolled at each level. Third, worse performance in primary and secondary level seems to prevent students from attaining higher levels of education. Note that the proportion of students who have never repeated primary level increases across schooling levels at the time that the proportion of repeaters decreases. A similar pattern is observed when focusing on repetition in secondary in which those more likely to survive higher schooling stages are those who performed better in secondary. Also, it is striking that the proportion of students enrolled in postsecondary education who have repeated primary is almost zero for both genders.

Differences across genders emerge across post-secondary enrollment for repeaters in secondary level. The proportion of repeater girls enrolled in post-secondary education is 13%, half of the rate observed for boys (24%).

Regarding to our proxies of non-cognitive ability, it is observed that the proportion of boys who drop out lower high school with a risky behavior (tried marijuana before age 15) is almost twice that of girls (9.5% for boys and 4.9% for girls). For both genders it is found that the proportion of students that highly value education,

those more motivated to participate in secondary level, increases across the schooling stages.

Finally, a great proportion of students dropping out from the educational system are mainly those who attended all grade years of each stage in a public institution.

Overall, differences found across genders and among the educational path justify a disaggregated analysis by gender, and through a sequential model, in the sense that the educational system seems to turn more selective in boys' and girls' parental educational background, past performance in schooling stages, motivation for enrollment, afro-descendant girls especially between the first and second schooling stages, and those receiving public education.

3.5. Methodological framework

This analysis attempts to answer the following questions: i) to what extent parental education, cognitive and non-cognitive abilities and individual's race influence education attainment? and, ii) is there any differential effect of parental educational background, individual's cognitive and non-cognitive abilities and race at different stages of schooling transition?

In order to answer these questions, the estimation strategy follows Cameron and Heckman (2001) in which education attainment is analyzed through a dynamic discrete choice model of schooling progression. This strategy recognizes that schooling attainment is the outcome of previous educational choices, which in turn depend on individuals' observable characteristics, like gender, race/ethnicity, and family background; but are also influenced by her unobservable characteristics, such as motivation or ability. The probability that an individual enters post-secondary education depends on upper high school graduation, which in turn depends on completing lower high school, making the model fundamentally recursive.

Therefore, individuals moving from one educational level to the next one may differ on their unobservable characteristics, in which the less able or motivated individuals are less likely to succeed in the transitions to higher education stages. Cameron and Heckman (2001) stress that at each stage the opportunity cost of schooling attendance is different, for instance labor market opportunities and earnings may vary according to individuals' education. Hence, it is possible that the opportunity cost of education increases for higher schooling stages when the system turns more selective

across the educational path. The intuition behind the model is that if the student population is divided between high and low ability individuals; and in turn between those coming from wealthier households and poorer ones; then it is expected that (i) more able individuals are more likely to succeed in higher educational stages in comparison to less able ones; and (ii) individuals coming from poorer households, ceteris paribus, may be prevented to move to the next educational level because of the household financial restrictions. Therefore, the ones surviving higher schooling stages are a selected sample of those more able individuals and with wealthier or better-off parental background, making important to control for the effects of such educational selection in order to isolate the causal effects of family background variables on education attainment.

Overall, in a dynamic framework, two factors induce biased estimations of the effects of family background on schooling progression. The first one refers to omitted variables (that is, not accounting for individuals' ability or motivation), while the second one refers to the selection taking place at different stages of the schooling transitions.

3.5.1 A sequential model of schooling progression

Following Cameron and Heckman (2001) the model assumes that each individual make schooling decision based on a sequential choice model. The choices available to the individual are limited by their earlier schooling choices.

The expected utility derived from each educational level is modeled as a latent utility indexy_{is}^{*}:

$$y_{is}^{*} = X_{is}'\beta_s + \alpha_s\theta_i + u_{is} \;\; i = 1, \dots, N; \; s = 1, \dots, S \qquad (1)$$

where X_{is} is a vector of observed constraint and expectation variables relevant to schooling decision *s*, θ_i are unobserved factors for the econometrician but known to the agent. This is the source of the essential heterogeneity, which can reflect individual's ability, motivation, or preferences; while u_{is} represents an idiosyncratic error term which is assumed to be independent of the explanatory variables (X_{is} and θ_i) and is independent across individuals. Also, it is assumed that: $u_{is} \sim N(0,1)$

Then, I can define the binary outcome $y_{is} = \begin{cases} 1 & \text{if } y_{is}^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$ (2)

These assumptions allow writing down the probability of making choice *s* as a probit model. Conditioning on θ,

$$\Pr(y_{is} = 1 \mid X_{is}, \theta_i, y_{is-1}) = \Phi(X_{is}'\beta_s + \alpha_s\theta_i) \quad (3)$$

where y_{is-1} are the past decisions made by the individual *i* and $\Phi(.)$ is the standard normal cumulative distribution function.

The probability of any sequence of schooling choices made by the individual y_{is} given the observed variables and θ_i can be expressed as:

$$\prod_{s \,\epsilon\, C_i} [\, Pr(y_{is} = 1 \mid X_{is}, \theta_i, y_{is-1})]^{y_{is}} [Pr(y_{is} = 0 \mid X_{is}, \theta_i, y_{is-1})]^{1-y_{is}} \quad (4)$$

where C_i is the set of decision nodes that individual *i* has visited.

3.5.2 Empirical strategy

In this study, I consider three education levels for girls and boys separately: lower high school (y_{i1}), upper high school (y_{i2}) and postsecondary level (y_{i3}). Then, the sequential process for individual *i* consists on: first decide whether or not to complete lower high school based on the underlying and unobserved expected utility (y_{i1}^*). After completing lower high school, the individual decides whether or not to complete upper high school (y_2) conditional on the expected utility (y_{i2}^*). Finally, for those graduating from upper high school, the individual chooses whether or not to enroll in postsecondary education (y_3) conditional on the expected utility derived from this election (y_{i3}^*).

The sequence of life cycle schooling histories can be written as follows:

s={1, 2, 3}

$$y_{i1} = \begin{cases} 1 & \text{or completing lower high school if } y_{i1}^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

For those individuals completing lower high school,

$$y_{i2} = \begin{cases} 1 & \text{or completing upper high school if } y_{i2}^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Finally, for those individuals graduating for upper high school

$$y_{i3} = \begin{cases} 1 & \text{or enrolled in postsecondary if } y_{i3}^{*} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Given the two levels of selection and the outcomes we have four types of individuals:

- Those who choose not to complete lower high school $y_{i1} = 0$
- Those who complete lower high school but decide not to continue upper high school $y_{i1} = 1, y_{i2} = 0$
- Those who complete upper high school but decide not to enroll in postsecondary education: $y_{i1} = 1, y_{i2} = 1, y_{i3} = 0$
- Those who decide to enroll in postsecondary education: $y_{i1} = 1, y_{i2} = 1, y_{i3} = 1$

For each of the educational levels stated before, the conditional probabilities are:

Not completing lower high school:

$$\Pr(y_{i1} = 0 \mid X_{i1}, \theta_i) = 1 - \Phi(X_{i1}'\beta_1 + \alpha_1\theta_i) \quad (8)$$

Completing lower high school and not continuing:

$$\Pr(y_{i1} = 1 \mid X_{i1}, \theta_i) = \Phi(X_{i1}'\beta_1 + \alpha_1\theta_i) \quad (9)$$

Not completing upper high school:

$$\Pr(y_{i2} = 0 \mid X_{i2}, y_{i1}\theta_i) = \Phi(X_{i1}'\beta_1 + \alpha_1\theta_i) - \Phi_2(X_{i1}'\beta_1 + \alpha_1\theta_i, X_{i2}'\beta_2 + \alpha_2\theta_i, \rho_{12}) \quad (10)$$

Completing upper high school and dropping out:

$$\Pr(y_{i2} = 1 \mid X_{i2}, y_{i1}\theta_i) = \Phi_2(X_{i2}'\beta_2 + \alpha_2\theta_i, X_{i1}'\beta_1 + \alpha_1\theta_i, \rho_{12}) - \Phi_3(X_{i1}'\beta_1 + \alpha_1\theta_i, X_{i2}'\beta_2 + \alpha_2\theta_i, X_{i3}'\beta_3 + \alpha_3\theta_i, \rho_{12}, \rho_{13}, \rho_{23}) \quad (11)$$

Being enrolled in postsecondary education:

$$\Pr(y_{i3} = 1 \mid X_{i3}, y_{i2}, \theta_i) = \Phi_3(X_{i1}'\beta_1 + \alpha_1\theta_i,\ X_{i2}'\beta_2 + \alpha_2\theta_i,\ X_{i3}'\beta_3 + \alpha_3\theta_i, \rho_{12}, \rho_{13}, \rho_{23}) \quad (12)$$

where $\Phi(.)$ is the standard normal cumulative distribution function, $\Phi_2(.)$ is the bivariate standard normal cumulative distribution with correlation coefficient ρ_{12} and $\Phi_3(.)$ is the trivariate standard normal cumulative distribution with correlation coefficients $\rho_{12}, \rho_{13}, \rho_{23}$.

$$\rho_{12} = cov[u_1, u_2|X_1, X_2], \qquad \rho_{13} = cov[u_1, u_3|X_1, X_3], \qquad \rho_{23} = cov[u_2, u_3|X_2, X_3]$$

The model is estimated using maximum-likelihood technique of the joint trivariate sample selection model.[15]

3.5.2.1 Explanatory variables

The explanatory variables included in the analysis are: race, parental educational background (both mother and father highest educational level attained), age cohort fixed effects, regional fixed effects, dummy variables indicating whether the child attended or not to pre-school, and the type of institution attended in primary level.

In order to capture individuals' cognitive ability, grade repetition in educational level is included in the analysis. Specifically, I consider whether the child never repeated, repeated once or more than once primary level. When considering the probability of completing upper high school or the probability of enrollment in postsecondary education, I also consider whether the individual never repeated, repeated once or more than once secondary level.

Repetition takes place when the child does not obtain the knowledge and skills necessary to think critically and solve complex problems in different areas, needed to succeed in the society and economy where they live. These maturing mental abilities are thought to broadly underpin learning and cognitive skills. In the Uruguayan education system, it is the teacher's decision to assess whether the student meets at the end of the year the minimum requirements for promotion.[16] [17] Overall, non-repeater students are those who obtained the necessary knowledge and maturity to pass to the next grade.

For cognition, there is a fairly well-established set of terminologies and conventions in the literature, for which aptitude tests and achievement test are designed to capture the speed at which the individuals learn and the knowledge they acquire

[15] This technique ensures consistent estimators (Rosenman et al, 2010).
[16] Discretion may affect grading marks as teachers may have different preferences or expectations.
[17] In lower high school and first grade in upper high school, students are assigned a mark for each of the 12 taught subjects based on their performance during the year. Students pass a subject if they get a mark above a given threshold. Those who fail a subject must re-take it during subsequent exam sessions (Manacorda, 2008; p7). For grading promotion in second and third year in upper high school, exams in particular subjects are mandatory.

(Almlund, et al., 2011) .[18] Thus, for the aim of this paper, repetition seems to be a good proxy of cognition.[19]

In addition, two variables are used in order to proxy non-cognitive ability. First, I consider motivation for enrollment in secondary level. Although this variable is not explicitly recognized as a factor in the Big Five model, Almlund et al. (2011) stress that one of the main critics received by this model is that it is silent about motivation. However, as also pointed out in Almlund et al. (2011), some studies relate academic motivation to Openness to Experience (p136).

The ENAJ asks individuals about the motives for secondary enrollment. Based on the alternative responses given to this question, I categorize the enrollment motives as: high motivation (those individuals reporting high value of education), labor motives (individuals declaring enrollment while they find a job), and not motivated (individuals declaring enrollment because they were "pushed to"). I expect most motivated individuals to be more likely to complete lower and upper high school, as compared to those who are less motivated to acquire education. Table A.4 in the Appendix provides a detailed description of the construction of this variable.

Second, I consider a dummy variable equal to one if the individual has tried marijuana before age 15. As was outlined in Section 2, this risky behavior was found to be negatively related with Consciousness (Gullone and Moore, 2000) and to have a negative influence on schooling progression (Heckman et al., 2014).

Models of educational choices also include additional choice-specific covariates. First, I consider the type of institution attended at different levels of high school. Public institution (both in lower or upper high school) is a dummy variable equal to one if the individual completed all grades of the corresponding level in a public institution and zero otherwise (those with at least one grade attending in a private institution). In general, the choice of a school, e.g., a private (fee paying) school, may reflect parental motivation to produce children of better quality (i.e., with higher schooling). For instance, a private school is likely to be of a better quality than a public school in the sense that may provide better infrastructures, better teachers, better peers, lower ratio of students per class; possibly affecting the probability of completing a schooling level.[20]

[18] For a deeper discussion on intelligence, see Chapter 4 in Almlund et al., (2011).
[19] It is worth mentioning that cognitive ability is likely to be influenced by child's environment, such as parental education, issue that is controlled for in this analysis.
[20] See Checchi (2006) Chapters 4 and 5 for an extensive review of the literature on the influence of supply of education and education financing on education attainment.

Also, the track chosen in secondary level is considered. While in lower high school there are no significant differences in curricula between general education and vocational training education, for upper high school differences turn to be important. Vocational training education is more oriented toward job placement (but is also possible to continue to tertiary education) than general academic education. In addition, the track chosen may also reflect individuals' self-selection if more able individuals choose general education instead of vocational training.[21]

Finally, internal migration is considered for postsecondary enrollment. As was stated before, universities in Uruguay (both public and private ones) are mainly located in Montevideo, so those individuals with financial family support are more likely to migrate to Montevideo and to attend university than poorer ones. Motive for migration is a categorical variable that captures whether the individual did not migrate after completing secondary level, if migrated for study motives, or migrated for other motives.[22]

3.5.2.2 Exclusion restrictions

In this analysis, as is standard in the literature I use exclusion restriction variables in order to identify the model. The model requires a subset of variables influencing the probability of attaining a certain educational level and not directly affecting the probability of completing the next one. Also, the exclusion restriction variables are assumed to be independent of the model unobservables.

As in previous studies (Cameron and Heckman, 2001; Heckman et al., 2014; Bernardi, 2012; Pal, 2004) I consider as exclusion restriction variables reflecting labor market conditions at each schooling stage, at the time the relevant decisions are taken. In each schooling stage, the individual chooses between completing the education level and dropping out to participate in the labor market. The decision is made considering the current labor market conditions and expectations on future returns to education. These rates may account for the opportunity cost of education. Then, if the individual continues in the educational system, the decision to attain the next schooling stage will depend on the opportunity cost of education –on the labor market conditions– at the time the decision of completing the level is evaluated. In other words, labor market

[21] An interesting debate in the educational literature refers to the consequences of the time of tracking on equity and efficiency of educational outcomes. See for instance van Elk et al. (2011).
[22] Other motives for migration are mainly labor, health, and family motives declared for migration.

conditions at time t influences schooling choices at time t, and only indirectly affecting schooling decisions of completion of the next level taken in $t+1$. It is clear that if the individual decides to drop out from the system in lower high school he is indirectly deciding not to attain upper high school, because of the sequential process of education attainment, but the individual cannot decide completing upper high school if lower high school was not achieved. Also, these rates are exogenous to individuals' schooling decisions.

A priori, the role of local labor market conditions is unclear. On the one hand, a high probability of employment might convince students to quit school and enter the labor market. On the other hand, the higher expected education returns could definitively be a stimulus for acquiring further education (Moccetti, 2008).

Specifically, I consider unemployment and employment rates, which are calculated for young people (aged 24 or less years old), by gender and at the department level at theoretical ages in which the individual is supposed to be enrolled in each schooling stage. Employment rates considered at each stage of the schooling progression are the following: unskilled youth employment rate for those children deciding whether to complete lower high school, semi-skilled youth employment rate for those choosing to complete upper high school, and youth skilled employment rate for individuals considering post-secondary enrollment.

Detailed information on the elaboration and classification of the variables are provided in Table A.4, while a summary of the independent variables considered in this analysis is provided in Tables A.5 and A.6 in the Appendix.

3.6 Results

In this section I first focus on the results related to unobserved heterogeneity and its correlations. Next, I describe the implications of the estimates of the model by discussing in turn, (1) the determinants of the probability of the initial schooling stage, (2) the determinants of upper high school transitions for those who completed lower high school, (3) the postsecondary enrollment decision for those surviving previous schooling stages (subsection 6.1). Next, subsection 6.2 gives a more complete picture of the educational path for boys and girls living in Uruguay.

3.6.1 Unobserved heterogeneity and correlations

A trivariate probit model with sample selection is estimated separately for females and males. Before presenting the estimated results, a natural question that emerges in this type of model is whether it is necessary to control for unobserved heterogeneity. Estimates of the cross-equation correlations between unobservables provide insights of the endogenous selection processes. In other words, the significance of the correlations highlights the importance of estimating education attainment as a sequential process.

In Table 5 it is shown that for both genders, unobservables across the three schooling levels are negatively associated although differences exist in the statistical significance of the estimated correlations. For girls, statistical and significant association is detected between the first and second transitions, while for boys between the second and third transitions. Thus, results show that the three schooling stages are differently interlinked and differ for both genders. Unobserved factors that make girls more likely to succeed in lower high school reduce their likelihood of attaining upper high school. For boys, unobserved heterogeneity that makes them more likely to complete upper high reduces their chances to enroll in post-secondary education. Any interpretation of this result is difficult. Recall that cognitive skills, motivation and risky behavior as proxies of socio-emotional endowments, are controlled for in the model. Therefore, these negative correlations between the residuals are capturing other unobservables different from ability and motivation. It could be argued that cultural factors, social pressure or labor market conditions, may induce children to achieve the minimum educational credentials recognized by the society and, once these credentials are obtained, children drop out from the educational system. Also, institutional and organizational factors as well as differences in curricula and grading promotion, which are specific of each schooling stage, could be differently affecting individuals' decisions of schooling. It could be speculated that these factors may influence children's adaptation or integration into different academic schemes.[23]

Tests for the ignorability of each selection mechanism were based on a Wald test of whether every correlation connecting each equation of the model was equal to zero. The null hypothesis of sample selection ignorability is rejected for both genders (bottom panel of Table 5). Thus, the results provide strong evidence that not accounting for the

[23] See Rama (2004) for an extensive description of the particularities of the institutional and organizational factors in the Uruguayan educational system. Fernández-Aguerre (2010) summarizes different empirical studies analyzing individuals' drop out from different stages of schooling in Uruguay.

potential endogeneity resulting from unobserved heterogeneity would induce biased results. This is also in line with the descriptive analysis provided in Section 3.

Tables A.5 and A.6 in the Appendix present the estimates of simple probit models not accounting for sample selection, separately for girls and boys. The magnitude of the bias could be observed by comparing the estimated coefficients of the key independent variables between simple probit models and the ones obtained from the trivariate probit estimations. Overall, it can be concluded that not accounting for selection overestimates the effects of the key variables on education attainment.

3.6.2 Empirical findings

Next, I discuss the estimated effects of the explanatory variables from the trivariate probit model. Average marginal effects capturing direct effects of the key variables on the probability of different schooling stages separately for girls and boys are presented in Tables 6 and 7 respectively.[24]

First, in line with the literature, parental educational background is an important factor explaining schooling completion. For both genders it is shown that the likelihood of attaining lower high school level for a child with highly educated parents more than doubles the probability of those with medium educated parents (Column 1, Tables 6 and 7). Thus, at this schooling stage the system gives less opportunities of completion to children from worse-off parental educational backgrounds.

Specifically, girls with a high educated mother are almost 12 percentage points (pp) more likely to complete this level in comparison to a girl with a low educated mother. For boys the opportunity gap between those with a high educated father and a low educated father is almost 19pp. In turn, having a medium educated mother increases the likelihood of completing this level in 5.7pp in comparison to girls from lower parental education background. Boys with a medium educated mother or father are 5.7 and 4pp respectively more likely to attain this level than boys from a more disadvantaged parental education.

Cognitive ability proxied by repetition in primary level decreases the probability of graduating from lower high school. Boys repeating once or more than once are

[24] Alternative specifications were also estimated not showing significant differences with the coefficients presented in Tables 6 and 7. These estimations included interactions of: race and parental educational background; race and motivation; motivation and parental education; repetition in both secondary and primary with motivation; parental education and repetition; and repetition and race. None of these interaction were statistical significant.

10.5pp and 15.1 pp more likely to drop out at this level than non-repeaters. Similar effects of past performance on schooling attainment are observed for girls (10.3pp and 16.4pp respectively).

In line with what is expected in the literature, more motivated individuals are more likely to complete lower high school. Girls and boys reporting enrollment in secondary level because they were "pushed to" are less likely to complete this level in comparison to those declaring high value of education (13.9 and 10.3 pp respectively). Also, girls and boys reporting labor motives for enrollment in secondary are less likely to complete this level than those more motivated ones, possibly putting less effort in attaining this level because of the anticipated decision of dropping out from the system once a job is found (4.2 and 8.3 pp respectively).

It is worth noting that at this schooling stage, while cognitive ability has similar effects on the probability of schooling completion across genders, motives for enrollment do not. Other things being equal, not motivated girls are more likely to leave the system than not motivated boys.

In addition, the results point to lower opportunities for afro-descendant girls, who are 5.1pp less likely to complete this educational level than non afro-descendants. Conversely, race is not a significant factor preventing boys attaining this educational level.

Next, the type of institution attended during primary level and lower high school decreases the probability of successfully completing this level. Individuals attending all grades in a public institution have lower chances to complete this level than those with at least one year in a private institution (8.3pp and 11.1pp for girls, and 16 and 7.5pp for boys, respectively for school and lower high school). Despite the heterogeneity in quality across public and private institutions that could be found in Uruguay, the public ones are associated in the literature with lower quality, in terms of resources and infrastructure, number of students per teacher, peer effects, in comparison to private ones. An alternative explanation is that private schools (mainly religious ones) are more effective in producing more motivated students and self-disciplined students (Coleman and Hoffer, 1983).[25]

It is worth mentioning that persistent effects of pre-school attendance are observed for girls (5.3pp), while this effect vanishes for males. A possible explanation

[25] Quoted in Carneiro and Heckman (2003) p39.

of this result is given in Apps et al. (2013). These authors stress that this result is quite common in the international literature, and may be due to strong effect from improved language skills (usually higher in girls), combined with the lower impact of negative behaviors (like aggressiveness, and antisocial behaviors), which are more common in boys (p.194).

Labor market opportunities have opposite effects on the probability of exiting the education system across genders.[26] For girls, higher unemployment rate decreases the probability of completing lower high school. This effect could be reflecting girls' future labor market expectations. If girls perceive that the labor market does not provide great opportunities, then they are discouraged to invest in human capital, thus dropping out from the system. For men, higher opportunities for unskilled workers increase the probability of dropping out from the educational system. Both variables, which measure opportunity cost of education, could be also measuring short-run family resource constraints. When lack of resources in the household are observed, children are more likely to drop out from the educational system in order to complement family's income.

Next, I move on to analyze the determinants of upper high school attainment for those surviving previous schooling stage (Column 2 in Tables 6 and 7). It is observed that children with high and medium educated parents show higher probability of graduating from secondary level relative to children from lower parental background. Therefore, this educational stage is also found to be less supportive to children from worse-off parental educational background giving them fewer opportunities to attain this schooling level.

Specifically for girls, having a high educated mother or father increase the probability of completing upper high school in comparison to girls with a low educated parent (16.8 and 11.6pp respectively). Boys with a high educated father are 13.6pp more likely to complete this level than those with low educated fathers. Also, boys with medium educated fathers and girls with medium educated mothers are more likely to complete this level in comparison to those with less educated parents (7 and 5.5pp for boys and girls respectively).

Second, race is an important factor deterring girls' and boys' upper high school completion although the effect is greater for afro-descendant girls. This is observed

[26] Legal age for participating in the labor market is 14 years old in Uruguay for the period of analysis.

when comparing the statistical significance of both marginal effects, 9.4pp at 99% of significance for girls and 13.1pp at 90% for boys.

Past performance in secondary level is the most important factor in explaining students' probability of dropping out from the system. Having repeated this level once increases the probability of dropping out in 25pp and almost 30pp for girls and boys respectively; while students repeating more than once are 34.5pp and 37pp less likely to graduate from upper high school than non-repeaters (girls and boys respectively). Note also the persistent effect of past performance in primary on the next levels of the educational system, not only indirectly affecting the probability of dropping out the system in an early stage but also directly decreasing the likelihood of leaving upper high school (14.3 and 18.7pp for girls and boys respectively). Thus, consistent with Cameron and Heckman (2001), differences in cognitive ability appear at early ages and persist over time.

Socio-emotional factors proxied by risky behavior and motivation influence schooling progression but play different roles across genders. For instance, motivation for enrollment still explains girls' but not boys' success in attaining upper high school. Girls who reported having been "pushed to" attend secondary level are 15.6pp less likely to complete upper high school than more motivated ones, while non statistically significant effects of risky behavior on upper high school completion are observed for girls. Conversely, risky behavior has negative and statistically significant effect on boys' probability of completing upper high school (almost 18pp significant at 95%) whereas motivation for secondary enrollment is not statistically significant. This is consistent with the psychological literature stressing different adolescent personality traits and propensity to be engaged in risky behaviors between male and female adolescents (see Gullone and Moore, 2000).

Also, differences across genders are observed in relation to the effect of the type of institution enrolled on upper high school completion. Girls who attended all grades into a public institution are 6.2 pp less likely to complete this level than those with at least one year in a private institution, while no statistical and significant effect is observed for boys. In addition, students (or their parents) choosing a general academic track are more likely to survive this educational stage than those tracked in vocational training education or those with mixed tracks (those who have changed between tracks within upper high school).

Labor market conditions also influence children's decisions on schooling completion. For boys, higher semi-skilled employment rate when the child is aged 15 (the theoretical age for attaining first grade in upper high school) decreases the probability of completing this level in 5.3pp. In turn, higher unemployment rate when girls are aged 15, increases in 2.8pp the probability of dropping out from the system. Thus, favorable labor market conditions for semi-skilled workers increase the opportunity cost of education for boys, while less attractive labor market conditions decrease the opportunity cost of schooling for girls.

Finally, the determinants of postsecondary enrollment are analyzed for those students surviving previous schooling level (Column 3 in Tables 6 and 7). Two main variables explain participation in postsecondary education for boys and girls. First, different opportunities in postsecondary enrollment are still observed for students from different parental educational background. For instance, boys with medium educated fathers and high educated fathers are respectively 14.1pp and 35.6pp more likely to attain postsecondary education than those from low educated parental background. In turn, girls with a high educated father are more likely to be enrolled in postsecondary education in comparison to girls with low educated father (8.6pp at 10% of significance), while no statistical and significant effect is observed between girls with low and medium educated parents. Therefore, this level seems to be more unequal for boys than for girls, in the sense that parental educational background influences more in boys' enrollment.

Second, internal migration after finishing secondary level is an important variable influencing individuals' postsecondary enrollment. Those declaring study motives for internal migration are more prone to be enrolled in this educational stage in comparison to not migrating ones (13.1 and 20pp for girls and boys, respectively). Internal migration for study motives could be reflecting household permanent income on the understanding that, as far as postsecondary institutions, mainly the public University (UdelaR) and private universities are located in Montevideo, those students not living in the capital and wanting to continue college should move to the capital, assuming all the related costs of this decision, like housing, food, etc. In other words, wealthier families are more likely to invest in their children' postsecondary education in comparison to poorer families.

It is also worth mentioning that neither race, past performance in the educational system, adolescent risky behavior nor motivation for secondary enrollment are

important direct determinants of postsecondary enrollment for any gender. This is explained because a great proportion of afro-descendants, less able and motivated individuals did not "survive" the previous stages and that almost all who survive and can afford moving to Montevideo (if were living in the Interior of the country) are enrolled in postsecondary level. This is also consistent with the descriptive analysis presented in Section 5. Overall, this educational stage seems to be more homogeneous in terms of individuals' observables and unobservable characteristics, leaving aside afro-descendants, individuals from less advantaged parental educational backgrounds and from poorer households, and those who performed worse in previous schooling stages, less motivated and more risky behavior.

3.6.3 Interpretation of results

This subsection describes the main findings of this study characterizing the educational path for girls and boys separately.

In the first stage of girls' schooling transition, parental educational level, past performance in primary level, motivation reported for attending secondary level, race and parental choices in terms of pre-school enrollment and types of institution attended in primary and in lower high school, play a major role in explaining lower high school attainment. Also, short-term income restrictions affect girls' probability of completing this level, but are less statistically significant than long-term family factors.

In the next educational level, the system turns more unequal in terms of opportunity gap between girls with a high educated mother or father in comparison to girls from worse-off parental background. This is observed when comparing the marginal effects of different educational backgrounds on the probability of attaining this level. Note that girls with a high educated parent (mother or father) are more likely to graduate from upper high school than in the previous educational level, while the marginal effect of being from a medium parental educational background decreases its statistical significance. In the last transition, the estimated effect of parental education decreases its magnitude and looses statistical significance, interpreted as less unequal than previous stages in the sense that the opportunity educational gap between girls from different parental educational background decreases.

Second, it is observed that the opportunity gap of completion for girls with different scholastic abilities increases in comparison to previous stages. In turn, motivation is still an important factor fostering upper high school completion but is less

significant than in the previous stage. This could be due to less motivated individuals being less likely to survive the previous level and this stage is more "homogenous" in terms of motivated individuals. Nonetheless, the decreasing effects of non-cognitive abilities should be interpreted with caution. First, because we are measuring something that is unobservable for the researcher, and therefore the proxies used in this kind of studies are at best imperfect. Second, because as noticed by the psychological literature, socio-emotional factors could be influenced over the individual's life cycle, for instance by schooling.[27] Therefore, we can only state that those who declared enrolling in secondary level because at this time they highly valued education are more likely to complete this level than those who reported being "pushed to".

Finally, postsecondary level could be seen as the less unequal schooling stage for girls from different parental educational background, race, and abilities. It is observed a great homogeneity in terms of girls' characteristics in this level, mainly explained because afro-descendant, less able and motivated girls and from worse-off parental backgrounds are less likely to survive previous schooling stages.

As a consequence for the surviving girls, enrollment in postsecondary level is almost determined by the possibility to migrate and to less extent, for those having a high educated father. Therefore, the higher we move in the educational system, the more unequal the system becomes in terms of opportunities given to girls from different parental backgrounds. It is also observed that the opportunity cost of education has different effects across girl's educational path. While in the first stage of schooling progression, fewer opportunities in the labor market increase girls' likelihood of schooling drop out, in the second stage worse conditions in the labor market increase the probability of completing this level. In addition, the statistical significance of this coefficient decreases across the educational path, possibly reflecting that opportunity cost of education is less important at higher schooling stages.

Similar patterns of selection are observed in the schooling transitions for boys, in the sense that more we advance in the educational path, the fewer the boys from disadvantaged parental educational background, less motivated and with worse performance in primary and secondary, have a chance to attain higher educational levels. Overall, it is observed for both genders that cognitive abilities has persistent and increasing effects in the probability of attaining higher schooling levels. Socio-

[27] There is an interesting ongoing debate in the psychological literature on the permanent versus variability in personality traits across the individual's life cycle. See for instance Almlund et al. (2011).

emotional factors, while important decreases its impact across the schooling progression.

Some differences across genders are observed. For instance, upper high school becomes less unequal for boys from different parental educational background than in the previous stage since the estimated coefficient decreases (for high educated father) and looses statistical significance (medium educated father). In turn, postsecondary level turns to be the more unequal one for boys from low and medium parental educational background in comparison to the previous levels.

Second, race has a major role in preventing girls from graduating from lower and upper high school than for boys, for whom race is only statistically significant in the second stage. Since interactions between race and cognitive abilities; race and motivation for secondary enrollment; and race and parental educational background were not statistically significant (see footnote 18), we can rule out that the estimated negative effect of race on schooling progression is due to differences in terms of parental educational backgrounds, motivation or cognitive abilities. Different interpretations are given by the literature for this negative and statistically significant coefficient. For instance, Porzecanski (2008) stresses that this negative coefficient could be capturing different processes of discrimination. One the one hand, it may reflect discrimination within the educational system which in turn affect afro-descendants' schooling decisions. On the other hand, it could be associated to discrimination in the labor market where returns to education are lower for afro-descendants, then discouraging afro-descendants to acquire more education.

Third, motivation and risky behavior show different effects across genders. While motivation is an important factor deterring girls' schooling progression, for boys it is only important for completing lower high school. Moreover, risky behavior turns to be an important factor in explaining boys' upper high school graduation; but not significant in explaining girls' schooling attainment.

Fourth, children's (or their parents) decisions in terms of type of institution attended have negative and decreasing impact on boys' and girls' schooling completion, but is more significant for girls than for boys (for whom in the second stage it is not statistically significant).

The results summarized above are consistent with the recent literature that highlights the importance of individuals' multiple abilities across one individual's life cycle. This literature stresses that cognitive ability is determined early in life while non-

cognitive ability is more malleable later in life. Specifically, Heckman and Carneiro (2003) point out that cognitive ability is formed relatively early in life and becomes less malleable at later stages of child's development. According to these authors, by age 14, intelligence as measured by IQ tests seems to be fairly well set. Non-cognitive skills, in turn appear to be more malleable until the late adolescent years (Heckman and Mosso, 2014) thus allowing public interventions contribute to the formation of non-cognitive skills (Brunello and Schlotter, 2011).

Heckman and coauthors refer to long run family factors crystallized in parental educational background, in scholastic ability and socio-emotional factors, as the driving force behind schooling attainment, and not short-term credit constraints.

In this study, because of lack of data on family's income or wealth data at the time of schooling choices are made, the effect of short-term family income is reflected by the opportunity cost of education measured by labor market variables. In line with Cameron and Heckman (2001) and Carneiro and Heckman (2003) who show that short-term family income is more important for high school dropout and completion than for college enrollment decision, I find that the opportunity cost of education is significant in explaining educational level' attainment, but its effect is smaller in comparison to long-term family factors and decreases along the educational path.

Finally, as was mentioned before, the public University (UdelaR) has been making big efforts in terms of territorial decentralization since 2007. These actions could indeed have a positive effect in terms of access to postsecondary education for students from low and middle educational background in the Interior of Uruguay. The literature analyzing the impact of higher education supply expansion points that any reduction in the influence of at least one circumstance on individuals' educational choices can be considered as reducing inequality of opportunity in education (see for instance Bratti et al., 2008; Peragine and Serlenga, 2007). Expanding supply in postsecondary education institutions may be associated to a cost-reduction effect, related to the increased supply and the possibility of enrolling at a university without moving to a different city. Also, expansion of higher education institution is associated to a potential increase in the expected returns of a higher schooling due to the wider and more diverse available offer (Bratti et al., 2008). Then, if new entrants are children from less privileged families, the effect of expansion may be the one of inclusion and increasing equality of opportunity almost by definition. But also, this literature recognizes that if barriers of access exist, such as fee payment, credit markets

imperfections, or selection tests, the effects of the supply expansion on improving equality of educational opportunity is not so obvious.

Conversely to other educational systems, public university in Uruguay does not rely on scholastic ability and willingness to pay. Therefore, it could be expected that territorial decentralization may benefit students from lower family backgrounds if policy interventions aiming to correct the selection process operating in previous stages takes place. In other words, in order to take full advantages of this decentralization process and the system to be inclusive in terms of less advantaged children, public interventions in secondary level are mandatory. In particular, policies intended to improve the environment that shape child's multiple abilities at different levels of the educational path will be more effective in increasing schooling progression in the long run.

3.7 Conclusion

In this paper, I analyzed to what extent long-term family factors crystallized in parental educational background, race, cognitive and socio-emotional endowments, as well as short-term family income proxied by the opportunity cost of education influence child's schooling progression. By analyzing the impact of these key variables across different stages of the educational path, this analysis gives a more complete overview of the major caveats of the Uruguayan educational system and about the factors that differentially affect girls and boys' educational attainment and gives insights of the inequality of acquisition in education at each stage of schooling progression.

I use the National Youth Survey containing individual information on education achievement and performance across the educational path, risky behavior and motivation for secondary enrollment, internal migration and schooling choices in terms of type of institution attended, among others.

The empirical strategy considers a sequential probability model developed by Cameron and Heckman (1998, 2001) in which schooling attainment is modeled as the outcomes of sequential choices made at each educational level, individuals' unobserved heterogeneity and alternative schooling cost of attendance at different levels. By taking into account the selection on education attainment, we obtain unbiased estimated results. Also, this analysis provides information on the different roles played by the key variables at different stages of schooling progression.

The results of this study confirms previous analyses addressing the deficiencies of the secondary level education in Uruguay (Aristimuñ o, 2009; Manacorda, 2008; among others). Furthermore, it extends previous research by considering the effects of cognitive and non-cognitive abilities, jointly with parental educational background, race, and opportunity cost of education measuring short-term family income, on different stages of the educational path in Uruguay.

When measuring socio-emotional endowments we encounter multiple issues largely recognized by the literature, such as the difficulty in capturing multiple personality traits (due to its unobservable nature), data availability that limits the measures of non-cognitive skills that can be used; and the static dimension of our proxies.[28]

Despite these limitations in measuring non-cognitive ability, the presented results gives enough evidence on the importance of both types of abilities in schooling progression not only directly affecting each schooling stage, but also indirectly influencing later stages.

In particular, the estimated results identify as one major deficiency of the Uruguayan system, the inequality in the acquisition of education for children with less scholastic abilities, the less motivated and with riskier behaviors, afro-descendants and from worse-off parental educational background. Also, these variables have different impacts as the students progress to higher schooling stages. This selection is observed in both lower and upper high school thereby affecting individual's probability of enrollment in postsecondary education. As was noted above, Uruguay stands-out in the region because it provides public education at all levels of the educational path. However, our results indicate that free education does not fully guarantee that individuals from worse-off family backgrounds (understood as less able individuals, poorer parental educational backgrounds) have access to high levels of education. Then, public policies should be oriented to mitigate those factors affecting individuals' educational decisions, especially focusing on individuals' from lower parental educational background, less able and motivated individuals, and afro-descendants that because of lower expectations or discrimination in the labor market and the educational system are more likely to drop out the educational system.

[28] Recall that there is no agreement in the psychological literature regarding how changes in personality are affected over the individual's life cycle.

In addition, in light of the results of the analysis it can be stressed that if no actions are taken to correct the inequalities observed in lower and upper high school, the recent decentralization process carried out by the public university will not succeed in providing more opportunities to those students from less advantaged parental backgrounds.

The findings presented and discussed above gives support to policy interventions at different stages of schooling progression in order to level the playing field for children from different parental educational backgrounds, race, scholastic and non-cognitive abilities. In particular, policies intended to promote cognitive ability early in life and social and behavioral skills in adolescence and youth, mainly focused on children from more disadvantaged environments –who probably receive little encouragement and support at home– should be explored. Finally, girls and boys develop alternative socio-emotional abilities across their life cycle, which in turn influence differently schooling progression across genders. Also, race is an important factor preventing schooling transition for boys and girls. Thus, promoting cognitive and non-cognitive abilities from a gender perspective and taking into account ethnical/ racial diversity may have positive effects on child's achievement of higher education. Overall, improving educational opportunities for less advantaged children will not only have positive impacts on future labor market outcomes, but also on other social outcomes such as crime and health, among others.

Aknowledgements

I am particular grateful to Raúl Ramos for kindly reading an earlier version of this paper and for his insightful comments. Also, special thanks to the participants of the Lunch seminar organized by Regional Quantitative Analysis Group (AQR) at Universitat de Barcelona, and participants in the Lunch Seminar organized at the Departament d'Economia Aplicada at Universitat Autònoma de Barcelona. The comments and suggestions received were of great value for the elaboration of this paper.

References

Almlund, M., Duckworth, A.; Heckman, J, and Kautz, T. (2011) *"Personality psychology and economics"*, in E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), Handbook of the Economics of Education 4:1-181. Amsterdam: Elsevier.

Aristimuño, A. (2009) "El abandono de los estudios del nivel medio en Uruguay: un problema complejo y persistente", *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación* 7 (4).

Apps, P.; Mendolia, S., and Walker, I. (2013) "The impact of pre-school on adolescents' outcomes: Evidence from a recent English cohort", *Economics of Education Review* 37:183-199.

Barro, R. J.; and Lee, J.L. (2001) "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers* 3:541-563.

Bernardi, F. (2012) "Unequal transitions: Selection bias and the compensatory effect of social background in education careers", *Research in Social Stratification and Mobility* 30:159-174.

Betts, J., and Roemer, J. (2005) "Equalizing Opportunity for Racial and Socioeconomic Groups in the United States through Educational Finance Reform", University of California Economics WP Series.

Blanden, J.; Gregg, P.; and Machin, S. (2002) "Education and Family Income", http://personal.lse.ac.uk/machin/pdf/Education%20and%20Family%20Income%20Stoke%20Rochford%20Final%20Version.pdf

Bourgignon, F.; Ferreira, F.; and Menéndez, M. (2003) "Inequality of Outcomes and Inequality of Opportunities in Brazil", William Davidson Institute WP No. 630.

Bowles, S. and Gintis, H. (2001) "Schooling in Capitalist America Revisited", *Sociology of Education* 75 (1):1-18.

Bowles, S. and Gintis, H. (2002) "The Inheritance of Inequality", *Journal of Economic Perspectives* 16 (3):3-30.

Bratti, M.; Checchi, D.; and De Blasio, G. (2008) "Does the Expansion of Higher Education Increase the Equality of Educational Opportunities? Evidence from Italy", *LABOUR* 22 (Special Issue):53-88.

Brunello, G, and Schlotter, M. (2011) "Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems", IZA DP No. 5743.

Cameron, S., and Heckman, J. (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy* 106 (2):262-333.

Cameron, S., and Heckman, J. (2001) "The dynamics of educational attainment for black, Hyspanic and white males", *Journal of Political Economy* 109 (3):455-99.

Cappellari, L. (2004) "High School Types, Academic Performance and Early Labour Market Outcomes", IZA WP No. 1048.

Carneiro, P., Crawford, C. and Goodman, A. (2007) "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes", CEE DP No. 92.

Cepal (2013) Anuario estadístico de América Latina y el Caribe, Santiago de Chile.

Checchi, D. (2006) *"The Economics of Education, Human Capital, Family Background and Inequality"*, Cambridge University Press.

Cunha, F.; and Heckman, J. J. (2007) "The technology of skill formation", *American Economic Review* 97 (2):31-47.

Cunha, F. and J. J. Heckman (2008) "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation", *Journal of Human Resources* 43 (4):738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006) *"Interpreting the evidence on life cycle skill formation*", in E. A. Hanushek and F.Welch (Eds.), Handbook of the Economics of Education, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Da Silveira, P. and R. Queirolo (1998) "Son nuestras escuelas y *Liceos* capaces de enseñar?", CERES WP No.7.

Dustmann, C. (2004) "Parental background, secondary school track choice, and wages", *Oxford Economic Papers*, 56(2):209-230.

Fernández-Aguerre (2010) (coord. and ed.) *"La desafiliación en la Educación Media y Superior de Uruguay: Conceptos, estudios y políticas"*, Colección Art.2, Comisión Sectorial de Investigación Científica, Universidad de la República.

Ferreira, F., and Gignoux, J. (2008) "The Measurement of Inequality of Opportunity: Theory and an application to Latin America", The World Bank, Policy Research WP 4659.

Francesconi, M.; and Nicoletti, C. (2006) "Intergenerational mobility and sample selection in short term panels" *Journal of Applied Econometrics* 21:1265-1293.

Furtado, M. (2003), "Trayectorias Educativas de los Jóvenes: el problema de la deserción", Cuaderno de trabajo TEMS, No. 22, Montevideo.

González, C., and Sanromán, G. (2010) "Movilidad intergeneracional y raza en Uruguay", DT No.13/10, Departamento de Economía, Facultad de Ciencias Sociales Universidad de la República.

Gullone, E., and Moore, S. (2000) "Adolescent risky-taking and the five-factor model of personality", *Journal of Adolescence* 23:393-407.

Heckman, J., and Carneiro, P. (2003) "Human Capital Policy", IZA DP No. 821.

Heckman, J., Humphries, J., Veramendi, G; and Urzúa, S. (2011) "The Effects of Educational Choices on Labor Market, Health and Social Outcomes", University of Chicago WP No. 2011-002.

Heckman, J., Humphries, J., Veramendi, G; and Urzúa, S. (2014) "Education, Health and Wages", IZA DP No. 8027.

Heckman, J., and Mosso, S. (2014) "The Economics of Human Development and Social Mobility", IZA DP No. 8000.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013) "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes", *American Economic Review* 103(6):2052-286.

Heckman, J.; Stixrud, J.; and Urzúa, S. (2006) "The Effects of Cognitive and Noncognitive abilities on Labor Market Outcomes and Social Behaviour", NBER WP No. 12006.

Heckman, J. J., Urzúa, S., and E. J. Vytlacil (2006). "Understanding instrumental variables in models with essential heterogeneity", *Review of Economic Statistatistics* 88(3):389-432.

Holm, A.; and Jaeger, M. (2011) "Dealing with selection bias in educational transition models: The bivariate probit selection model", *Research in Social Stratification and Mobility*.

Karlson, K. (2011) "Multiple paths in educational transitions: A multinomial transition model with unobserved heterogeneity", *Research in Social Stratification and Mobility* 29:323-341.

Lundberg, S. (2013) "Educational Inequality and the Returns to Skills", IZA DP No. 7595.

Manacorda, M. (2008) "The Cost of Grade Retention", CEP Discussion Paper No 878.

Mare, R. (1980) "Social Background and School Continuation Decisions", *Journal of American Statistics Association* 75:295-305.

Moccetti, S. (2008) "Educational choices and the selection process before and after compulsory schooling", Temi di discussione series WP No. 691.

Pal, S. (2004) "Child schooling in Peru: Evidence from a sequential analysis of schooling progression", *Journal of Population Economics* 17:657-680.

Peragine, V. and Serlenga, L. (2007) "Higher education and equality of opportunity in Italy", ECINEQ WP 2007-79.

Porzecanski, R. (2008) "Raza y Desempeño Educativo en el Uruguay Contemporáneo: Un análisis de la brecha entre afro-descendientes y blancos", Paper presented in the "III Congreso de la Asociación Latinoamericana de la Población", Cordoba, Argentina.

Roemer, J. (1998) *"Equality of Opportunity"*, Cambridge MA: Harvard University Press.

Rama, G. (2004) "La evolución de la educación secundaria en Uruguay", *REICE-Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación* 2(1).

Roodman, D. (2010) "Estimating fully observed recursive mixed-process models with cmp", *Stata Journal* 11(2): 159-206-

Rosenman, R.; Mandal, B.; Tennekoon, V.; and Hill, L. (2010) "Estimating treatment effectiveness with sample selection". Washington State University http://faculty.ses.wsu.edu/WorkingPapers/Rosenman/WP2010-5.pdf

SITEAL (2005) "La educación superior en América Latina: acceso, permanencia y equidad" http://www.siteal.iipe-oei.org/sites/default/files/educacion_superior.pdf

van Elk, R.; van der Steeg, M; and Webbink, D. (2011) "Does the timing of tracking affect higher education completion?", *Economics of Education Review* 30:1009-1021.

TABLES AND FIGURES

Figure 1 The Educational System in Uruguay

School cycle	Grades	Theoretical ages	Compulsory	School type
Pre-school	0	4-5	No*	Centros CAIF, Guarderías
Educación primaria	1-6	6-11	Yes	Escuelas
Ciclo básico	7-9	12-14	yes	Liceos/UTU
Bachillerato	10-12	15-17	No	Liceos/UTU
Post secondary		18-23	No	University/ Teaching Training Institutes/ Tertiary education (vocational training) /Militar school/ Tertiary education Non University Institutes (private institutions)

* Since 2008 this level is compulsory for children aged 4-5 years. *Ley General de Educación N° 18.437*

Table 1 Schooling progression (%) by gender and ethnicity

	Total	Males	Females	Males		Females	
				Non- afro	Afro	Non- afro	Afro
Finished school	96.7	96.6	96.8	96.9	94.8	97.5	95.8
Finished CBU	64.5	63.3	65.8	65.9	47.7	69.5	46.2
Finished high school	29.2	25.4	32.8	28.8	5.1	36.4	13.7
Enrolled Tertiary level	20.5	16.0	24.8	18.2	2.7	27.2	12.0

ENAJ (2008)

Table 2 Summary statistics (%)

Variable	Total	Female	Male	No afro	Afro
Female	0.52			51.7	57.4
Afro-descendant	0.11	0.12	0.09		
Parents' background					
Mother's education					
Low level	0.48	0.48	0.47	0.46	0.65
Medium level	0.37	0.36	0.37	0.38	0.28
High level	0.15	0.15	0.15	0.16	0.07
Father's education					
Low level	0.52	0.54	0.51	0.50	0.69
Medium level	0.36	0.35	0.38	0.37	0.27
High level	0.11	0.11	0.12	0.12	0.04
Attended pre-school	0.83	0.83	0.84	0.84	0.75
Completed primary level	0.98	0.98	0.97	0.98	0.95
Public school (all years)	0.77	0.77	0.77	0.76	0.87
Performance in primary (Repeated)					
Never	0.76	0.78	0.73	0.78	0.59
Once	0.18	0.16	0.19	0.16	0.27
More than once	0.07	0.05	0.08	0.06	0.14
Obs.	2,349	1,228	1,121	2,100	249

Table 3 Summary statistics across the schooling progression for girls

	Lower highschool			Upper highschool			Post-secondary	
Variable	**Enrolled**	**Drop-out**	**Complete**	**Enrolled**	**Drop-out**	**Complete**	**Not enrolled**	**Enrolled**
Afro	0.098	0.179	0.082	0.074	0.116	0.050	0.047	0.051
Mother's edu level								
Low	0.445	0.761	0.382	0.342	0.497	0.257	0.378	0.220
Medium	0.384	0.217	0.417	0.434	0.425	0.438	0.472	0.428
High	0.171	0.022	0.201	0.225	0.079	0.304	0.150	0.352
Father's edu level								
Low	0.508	0.728	0.464	0.435	0.562	0.366	0.504	0.323
Medium	0.372	0.255	0.395	0.411	0.373	0.431	0.394	0.443
High	0.121	0.016	0.142	0.155	0.065	0.203	0.102	0.235
Attended pre-school	0.853	0.685	0.886	0.894	0.849	0.918	0.890	0.927
Public school (all years)	0.749	0.967	0.706	0.681	0.798	0.618	0.748	0.577
Performance in Primary								
Never repeated	0.839	0.533	0.901	0.923	0.839	0.968	0.929	0.980
Repeated once	0.136	0.370	0.090	0.074	0.151	0.032	0.071	0.020
Repeated 2+	0.024	0.098	0.010	0.004	0.010	0.000	0.000	0.000
Noncognitive ability								
Tried marijuana before 15yr	0.025	0.049	0.021	0.021	0.027	0.017	0.016	0.017
Motivation to enrollment								
Highly motivated	0.778	0.571	0.819	0.835	0.784	0.862	0.819	0.875
Labor motives	0.060	0.174	0.037	0.030	0.058	0.015	0.016	0.015
Not motivated	0.139	0.201	0.126	0.118	0.140	0.106	0.134	0.098
Other motives	0.023	0.054	0.017	0.017	0.017	0.017	0.031	0.012
Lower highschool vbles								
Public	0.790	0.989	0.750	0.723	0.887	0.634	0.756	0.597
Private	0.184	0.000	0.221	0.245	0.075	0.338	0.205	0.379
General education (all grades)	0.884	0.701	0.920	0.950	0.901	0.978	0.961	0.983
Vocational training (UTU all grades)	0.071	0.185	0.049	0.030	0.058	0.015	0.024	0.012
Upper highschool vbles								
Public institution (all yr)				0.728	0.853	0.660	0.780	0.623
General education (all grades)				0.870	0.750	0.935	0.835	0.966
Vocational training (UTU all grades)								
Performance in Secondary								
Never repeated				0.647	0.336	0.817	0.661	0.866
Repeated once				0.229	0.401	0.136	0.252	0.100
Repeated 2+				0.123	0.264	0.047	0.087	0.034
Migration motives (after highschool)								
Not migrated							0.638	0.616
Other motives							0.291	0.154
Study							0.071	0.230
Obs.	1,109	184	925	828	292	536	127	409

Table 4 Summary statistics across the schooling progression for boys

	Lower highschool			Upper highschool			Post-secondary	
Variable	**Enrolled**	**Drop-out**	**Complete**	**Enrolled**	**Drop-out**	**Complete**	**Not enrolled**	**Enrolled**
Afro	0.089	0.150	0.073	0.067	0.098	0.040	0.060	0.031
Mother's edu level								
Low	0.432	0.740	0.355	0.320	0.404	0.248	0.410	0.176
Medium	0.400	0.250	0.437	0.453	0.459	0.449	0.436	0.454
High	0.168	0.010	0.207	0.227	0.138	0.303	0.154	0.370
Father's edu level								
Low	0.475	0.745	0.407	0.374	0.474	0.288	0.504	0.191
Medium	0.395	0.250	0.431	0.445	0.428	0.459	0.444	0.466
High	0.130	0.005	0.161	0.181	0.098	0.253	0.051	0.344
Attended pre-school	0.870	0.760	0.897	0.904	0.872	0.931	0.855	0.966
Public school (all years)	0.747	0.965	0.693	0.667	0.771	0.578	0.675	0.534
Performance in Primary								
Never repeated	0.789	0.520	0.856	0.890	0.810	0.958	0.897	0.985
Repeated once	0.172	0.370	0.123	0.102	0.174	0.040	0.103	0.011
Repeated 2+	0.039	0.110	0.021	0.008	0.015	0.003	0.000	0.004
Noncognitive abilities								
Tried marijuana before 15yr	0.058	0.095	0.048	0.048	0.073	0.026	0.009	0.034
Motivation to enrollment								
Highly motivated	0.73	0.57	0.77	0.78	0.75	0.80	0.79	0.81
Labor motives	0.09	0.16	0.07	0.06	0.07	0.04	0.08	0.03
Not motivated	0.14	0.19	0.12	0.12	0.13	0.12	0.10	0.13
Other motives	0.05	0.09	0.04	0.04	0.05	0.03	0.03	0.03
Lower highschool vbles								
Public	0.775	0.960	0.729	0.701	0.829	0.591	0.778	0.508
Private	0.188	0.010	0.232	0.263	0.131	0.377	0.197	0.458
General education (Liceo all grades)	0.772	0.500	0.840	0.875	0.801	0.939	0.846	0.981
Vocational training (UTU all grades)	0.130	0.230	0.106	0.088	0.138	0.045	0.120	0.011
Upper highschool vbles								
Public institution (all yr)				0.705	0.798	0.625	0.795	0.550
General education (Liceo all grades)				0.761	0.664	0.844	0.667	0.924
Vocational training (UTU all grades)								
Performance in Secondary							0.795	0.550
Never repeated				0.545	0.324	0.736	0.675	0.763
Repeated once				0.252	0.346	0.172	0.205	0.156
Repeated 2+				0.203	0.330	0.092	0.120	0.080
Migration motives (after highschool)								
Not migrated							0.624	0.603
Other motives							0.085	0.248
Study							0.291	0.149
Obs.	1,005	200	805	706	327	379	117	262

Table 5 Estimated correlations of unobservables and test of ignorability

	Girls		***Boys***	
Correlations of unobservables	*Estimate*	*p-value*	*Estimate*	*p-value*
ρ_{12}(Complete Upper HS, Complete Lower HS)	-0.586	0.044	-0.314	0.485
ρ_{13} (Completing Lower HS, Postsec enrollment)	-0.347	0.459	-0.395	0.469
ρ_{23} (Completing Upper HS, Postsec enrollment)	-0.174	0.474	-0.591	0.028
	χ^2	*p-value*	χ^2	*p-value*
Wald test of ignorability				
$Ho: \rho_{12} = \rho_{13} = \rho_{23} = 0$	12.79	0.0017	26.41	0.0000

Ho: Sample selection is ignorable.

Table 6 Educational path (Girls) Average marginal effects

Variables	Lower high-school (1)		Upper high-school (2)		Post-secondary (3)	
Afro-descendants	-0.051**	(0.021)	-0.094**	(0.048)	0.104	(0.068)
Parental education (Ommited: low level of education)						
Mother's edu level medium	0.057***	(0.017)	0.055*	(0.032)	-0.029	(0.035)
Mother's edu level high	0.118***	(0.034)	0.168***	(0.045)	0.039	(0.045)
Father's edu level medium	0.032**	(0.016)	0.040	(0.030)	0.030	(0.033)
Father's edu level high	0.057	(0.045)	0.116**	(0.048)	0.086*	(0.049)
Multiple abilities						
Ommited variables in repetition (Never repeated)						
Repeated once school	-0.103***	(0.016)	-0.143***	(0.053)	-0.037	(0.084)
Repeated school 2+	-0.164***	(0.036)	.	.	.	.
Repeated once secondary	.	.	-0.251***	(0.027)	-0.047	(0.057)
Repeated secondary 2+	.	.	-0.345***	(0.036)	0.011	(0.080)
Motives for enrollment in secondary (Omitted: highly motivated)						
Not motivated	-0.139***	(0.025)	-0.156**	(0.063)	-0.013	(0.098)
Labor motives	-0.042**	(0.019)	-0.022	(0.040)	-0.027	(0.047)
Other motives	-0.082**	(0.040)	0.009	(0.102)	-0.153	(0.100)
Marijuana before 15			-0.065	(0.100)		
Stage- variant variables						
Lower high school						
Public institution	-0.111***	(0.043)	.	.	.	.
Unemployment rate	-0.126**	(0.058)	.	.	.	.
All years in public school	-0.083***	(0.029)	.	.		.
Attended pre-school	0.053***	(0.017)	.	.	.	.
Upper high school						
Public institution	.	.	-0.062*	(0.032)	.	.
General education	.	.	0.208***	(0.038)	.	.
Unemployment rate_age15	.	.	0.283*	(0.169)	.	.
Unemployment rate_age16	.	.	-0.146	(0.174)	.	.
Postsecondary education						
Migration motives (Omitted variable: not migrated)						
Motives for migration: studies	.	.	.	.	0.131**	(0.052)
Other motives for migration	.	.	.	.	-0.060*	(0.036)
Employment rate_skilled	.	.	.	.	0.546	(0.393)
Regional dummies					Yes all stages	
Cohort age dummies					Yes all stages	
Obs.	1109		825		536	

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 7 Educational path (Boys) Average marginal effects

Variables	Lower high-school (1)		Upper high-school (2)		Post-secondary (3)	
Afro-descendants	-0.020	(0.026)	-0.130*	(0.070)	-0.073	(0.094)
Parental education (Ommited: low level of education)						
Mother's edu level medium	0.057***	(0.017)	-0.014	(0.039)	0.010	(0.054)
Mother's edu level high	0.189***	(0.048)	0.051	(0.052)	0.084	(0.068)
Father's edu level medium	0.040**	(0.018)	0.070*	(0.037)	0.141***	(0.052)
Father's edu level high	0.187***	(0.065)	0.136**	(0.054)	0.356***	(0.078)
Multiple abilities						
Ommited variables in repetition (Never repeated)						
Repeated once school	-0.105***	(0.019)	-0.187***	(0.062)	-0.226	(0.137)
Repeated school 2+	-0.151***	(0.032)	-0.115	(0.237)	.	.
Repeated once secondary	.	.	-0.296***	(0.032)	0.004	(0.058)
Repeated secondary 2+	.	.	-0.373***	(0.036)	0.019	(0.088)
Motives for enrollment in secondary (Omitted: highly motivated)						
Not motivated	-0.103***	(0.025)	-0.004	(0.074)	.	.
Labor motives	-0.083***	(0.020)	0.003	(0.046)	.	.
Other motives	-0.085***	(0.031)	0.030	(0.082)	.	.
Marijuana before 15	.	.	-0.179**	(0.076)	0.198	(0.148)
Stage- variant variables						
Lower high school						
Public institution	-0.075*	(0.039)	.	.	.	.
Unskilled employment rate	-0.319***	(0.121)	.	.	.	.
All years in public school	-0.160***	(0.031)	.	.		.
Attended pre-school	0.032	(0.020)	.	.	.	.
Upper high school						
Public institution	.	.	-0.031	(0.037)	.	.
General education	.	.	0.192***	(0.038)	.	.
Semi-skilled Employment rate_age15	.	.	-0.527**	(0.257)	.	.
Semi-skilled Employment rate_age16	.	.	0.215	(0.261)	.	.
Postsecondary education						
Migration motives (Omitted variable: not migrated)						
Motives for migration: studies	.	.	.	.	0.199***	(0.065)
Other motives for migration	.	.	.	.	-0.078	(0.049)
Unemployment rate (postsec)	.	.	.	.	0.566*	(0.307)
Employment rate_skilled	.	.	.	.	-0.272	(0.477)
Regional dummies	Yes all stages					
Cohort age dummies	Yes all stages					
Obs.	994		706		378	

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

APPENDIX

Table A.1 Main descriptives of the Uruguayan Educational System

Variables	Total	Male	Female
Preschool			
Never attended	0.12	0.12	0.12
Public	0.54	0.55	0.54
Private	0.34	0.33	0.34
Type of school (all years)			
Public	0.78	0.77	0.78
Private	0.17	0.18	0.16
Mixed	0.05	0.04	0.06
Lower highschool			
Type of institution (attended al grades in)			
Public	0.78	0.77	0.79
Private	0.03	0.04	0.03
Mixed	0.18	0.19	0.18
Track			
General Education	0.86	0.81	0.90
Training institute	0.09	0.12	0.06
Mixed	0.06	0.08	0.04
Upper highschool			
Type of institution (attended al grades in)			
Public	0.72	0.70	0.74
Private	0.16	0.17	0.15
Mixed	0.12	0.13	0.11
Track			
General Education	0.82	0.76	0.88
Training institute	0.06	0.10	0.03
Mixed	0.11	0.14	0.09
Postsecondary			
Type of institution (last level attained in)			
Public	79.5	70.78	85.01
Private	20.5	29.22	14.99

ENAJ (2008), ECH (2008)

Table A.2 Simple probit Girls

Variables	Lower high-school		Upper high-school		Post-secondary	
Afro-descendants	-0.069***	(0.027)	-0.104**	(0.048)	0.098	(0.083)
Parental education (Ommited: low level of education)						
Mother's edu level medium	0.074***	(0.021)	0.065**	(0.031)	0.006	(0.041)
Mother's edu level high	0.159***	(0.045)	0.180***	(0.044)	0.118**	(0.050)
Father's edu level medium	0.043**	(0.021)	0.047	(0.030)	0.058	(0.038)
Father's edu level high	0.066	(0.057)	0.127***	(0.049)	0.143***	(0.053)
Multiple abilities						
Ommited variable in repetition (Never repeated)						
Repeated once school	-0.139***	(0.020)	-0.181***	(0.051)	-0.173*	(0.095)
Repeated school 2+	-0.214***	(0.046)	.	.	.	.
Repeated once secondary			-0.256***	(0.027)	-0.156***	(0.045)
Repeated secondary 2+			-0.342***	(0.037)	-0.144**	(0.072)
Motives for enrollment in secondary (Ommited: highly motivated)						
Not motivated	-0.180***	(0.032)	-0.165**	(0.069)	-0.055	(0.111)
Labor motives	-0.058**	(0.025)	-0.026	(0.041)	-0.043	(0.052)
Other motives	-0.113**	(0.052)	0.020	(0.102)	-0.189*	(0.108)
Marijuana before 15			-0.079	(0.098)		
Stage- variant variables						
Lower high school						
Public institution	-0.144***	(0.055)				
Unemployment rate	-0.164**	(0.076)				
All years in public school	-0.102***	(0.036)				
Attended pre-school	0.071***	(0.023)				
Upper high school						
Public institution			-0.066**	(0.032)		
General education			0.177***	(0.038)		
Unemployment rate_age15			0.149	(0.154)		
Unemployment rate_age16			-0.019	(0.152)		
Postsecondary education						
Migration motives (Ommited: not migrated)						
Motives for migration: studies					0.173***	(0.053)
Employment rate_skilled					0.661	(0.464)
Regional dummies	Yes all stages					
Cohort age dummies	Yes all stages					

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.3 Simple probit Boys

Variables	Lower high-school		Upper high-school		Post-secondary	
Afro-descendants	-0.030	(0.034)	-0.140**	(0.071)	-0.072	(0.087)
Parental education (Ommited: low level of education)						
Mother's edu level medium	0.078***	(0.023)	0.011	(0.038)	0.036	(0.050)
Mother's edu level high	0.265***	(0.066)	0.087*	(0.049)	0.147**	(0.060)
Father's edu level medium	0.054**	(0.024)	0.083**	(0.036)	0.156***	(0.045)
Father's edu level high	0.287***	(0.093)	0.156***	(0.052)	0.390***	(0.062)
Multiple abilities						
Ommited variable in repetition (Never repeated)						
Repeated once school	-0.146***	(0.025)	-0.237***	(0.054)	-0.281**	(0.113)
Repeated school 2+	-0.207***	(0.042)	-0.225	(0.251)	.	.
Repeated once secondary			-0.294***	(0.032)	-0.047	(0.052)
Repeated secondary 2+			-0.378***	(0.035)	-0.057	(0.075)
Motives for enrollment in secondary (Ommited: highly motivated)						
Not motivated	-0.121***	(0.034)	0.001	(0.065)	-0.216***	(0.084)
Labor motives	-0.101***	(0.028)	-0.016	(0.044)	-0.026	(0.065)
Other motives	-0.116***	(0.042)	0.014	(0.081)	0.022	(0.094)
Marijuana before 15			-0.176**	(0.079)	0.234*	(0.133)
Stage- variant variables						
Lower high school						
Public institution	-0.069	(0.043)				
Unskilled employment rate	-0.416**	(0.167)				
All years in public school	-0.210***	(0.040)				
Attended pre-school	0.034	(0.027)				
Upper high school						
Public institution			-0.039	(0.036)		
General education			0.182***	(0.036)		
Semi-skilled Employment rate_age15			-0.544**	(0.263)		
Semi-skilled Employment rate_age16			0.194	(0.266)		
Postsecondary education						
Migration motives (Ommited: not migrated)						
Motives for migration: studies					0.206***	(0.061)
Unemployment rate (postsec)					-0.264	(0.486)
Employment rate_skilled					0.526*	(0.310)
Regional dummies	Yes all stages					
Cohort age dummies	Yes all stages					

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A.4 Definition of independent variables

Variables	Description	Type of variable
Afro-descendant	1 if respondant declares afro descendance; 0 otherwise	Dummy
Age fixed effects	Age in years	Categorical
Mother's highest level attained	Classification according to years of education	Categorical
Low	Less than 9 years	
Medium	Between 9 to 12 years	
High	More than 12 years	
Father's highest level attained	Classification according to years of education	Categorical
Low	Less than 9 years	
Medium	Between 9 to 12 years	
High	More than 12 years	
Cognitive ability		
Performance in Primary (Repeated)		Categorical
Never	If the individual declares never repeated	
Once	if the individual declares repeated once	
More than once	if the individual declares repeated more than once	
Performance in Secondary (Repeated)		Categorical
Never	If the individual declares never repeated	
Once	if the individual declares repeated once	
More than once	if the individual declares repeated more than once	

Table A.4 Definition of independent variables (cont.)

Variables	Description	Type of variable
Noncognitive ability		
Motivation for secondary enrollment	If the individual declares as main reason for enrollment one of the alternatives:	Categorical
Highly motivated	Acquisition of education	
	Today is essential to study	
	You are interested on what you are studying	
	Expect to improve social status through education	
Labor motives	If the individual declares as main reason for enrollment one of the alternatives:	
	In order to quickly find a job	
	Studies while finding a job or start a family	
Not motivated	If the individual declares as main reason for enrollment one of the alternatives:	
	Oblished to	
Other motives	If the individual declares as main reason for enrollment one of the alternatives:	
	Receive subsidies	
	to meet other youths	
	others	
Tried marijuana before 15	Equal to one if the individual declares trying marijuana before age 15; 0 otherwise	Dummy

Table A.4 Definition of independent variables (cont.)

Variables	Description	Type of variable
Institutional variables		
Public school (all years)	Equal to one if the individual declares attending all grades of primary level in a public school; 0 otherwise	Dummy
Attended pre-school	Equal to one if the individual declares having attended pre-school; 0 otherwise	Dummy
Public in lower highschool	Equal to one if the individual declareshaving attended all grades of upper highschool in a public institution; 0 otherwise	Dummy
Public in upper highschool	Equal to one if the individual declares attending all grades of lower highschool in a public institution; 0 otherwise	Dummy
Vocational education	Equal to one if the individual declares having attended all grades of upper highschool in a General academic institution; 0 otherwise	Dummy
Labor market variables		
Youth unemployment rate	Unemployment rate of population aged less than 25 by gender, department of residence and different schooling stages*	Numerical
Employment rates	Employment rates calculated at the department of residence level and different schooling stages*	
Unskilled employment rate	Employment rate for workers with less than 9 years of education	Numerical
Semi-skilled employment rate	Employment rate for workers with 9 to 12 years of education	
Skilled employment rate	Employment rate for workers with more than 12 years of education	
Migration motives	If the individual declares as main motives for migration (after completing upper high school) Study Other (includes labor, health, family, and other motives Never moved	Categorical

*For example one girl living in Montevideo deciding whether or not to completing upper high school, the unemployment rates used in the model are 3 Female youth unemployment rates in Montevideo, one for each year when the girl was aged 15, 16 and 17; theoretical ages in which girl is supposed to be in upper high school. Similar strategy was used to the calculation of employment rates.

Table A.5 Independent variables

Observed personal characteristics	Race
Parental education level (mother and father)	Low (less than 9 yr) Medium (9 to 12 yr) High (More than 12 yr)
Institutional	Public school (all years) Attended pre-school
Cognitive ability	Performance in primary (Repeated) Never Once More than once
	Performance in Secondary (Repeated) Never Once More than once
Non-cognitive ability	Marijuana before age 15*
	Motivation to enrollment in secondary level Highly motivated Not motivated Labor motives Other motives

*Tried marijuana before 15 is only included in upper high school in order to avoid endogeneity issues in lower high school.

Table A.6 Independent variables. Stage-variant regressors

Lower highschool	***Upper highschool***	***Post-secondary enrollment***
Region of residence (departament)	Region of residence (departament)	*Motives for migration (at theoretical age of attendance)* Never migrated Study motives Other motives (family, labor, health, others)
	Performance in secondary level (Repeated) Never Once More than once	*Performance in secondary level (Repeated)* Never Once More than once
Labor opportunities		
Unemployment youth rate (by gender, region and for theoretical ages of attendance)	Unemployment youth rate (by gender, region and for theoretical ages of attendance)	Skilled Employment rate (by gender, region and for theoretical ages of attendance)
Unskilled Employment rate (by gender, region and for theoretical ages of attendance)	Semi-skilled Employment rate (by gender, region and for theoretical ages of attendance)	
Institution type (all years in public institution)	Institution type (all years in public institution)	
Unemployment youth rate (by gender, region and for theoretical ages of attendance)	Unemployment youth rate (by gender, region and for theoretical ages of attendance)	
Unskilled Employment rate (by gender, region and for theoretical ages of attendance)	Semi-skilled Employment rate (by gender, region and for theoretical ages of attendance)	
	Vocational education (all yr General educ.)	

4. Conclusions

The main aim of this thesis has been to contribute to the literature on economic development by providing empirical evidence on three channels suggested by the literature that may cause individuals and countries to be entrapped in poverty.

The **first essay** of this thesis studied the relationship between immigrants' social networks and their subsequent labor market outcomes in Spain for 1997-2007. For this purpose, I used the National Immigrant Survey carried on 2007 and conducted two empirical exercises. First, I analyzed the extent to which social networks affect immigrants' job match. Second, for immigrants keeping the first job in Spain, I studied to what extent social networks influence wages. The econometric technique followed a two-step type procedure similar to the one proposed by Heckman (1979) to control for endogeneity issues.

The main results of this essay showed a great reliance on immigrants' social networks for employment in the host country. Job mismatch is more likely to occur for those immigrants that upon arrival prefer to quickly being employed in a job provided by the network, even if it is not the most suitable one in terms of the immigrants' human capital and previous experience. In addition, the results confirmed a positive effect of the network size on the probability of job matching. For those keeping the first job, network size is found to penalize immigrants' wages. Also, despite we found differences across the wage distribution and gender, the strength of the network is found to penalize immigrants' wages.

These results may be reflecting that social capital accumulated by the network is restricted to a particular segment of the labor market and thus, limiting immigrant's job prospects to the network, and also depressing wages for those immigrants in segmented occupations or sectors of activity. From this analysis we suggested that policy interventions aiming to socially and economically integrate immigrants in Spain, should be focused on influencing immigrant's environment by for instance, promoting greater access to formal institutions in the labor market and reducing immigrant's dependence on the information transmitted by the network.

The aim of the **second essay** of this dissertation was to test the predictions of Banerjee and Newman's model, which s, suggests that development paths are determined by countries' initial conditions, notably wealth distribution and credit market institutions.

This model predicts that countries with high historical rate of credit to non-credit constrained people end up in a situation in which only a small share of the population might start-up new firms, but these firms do not grow over time. In this case, the process of development ends up in a situation of low wages, in which there is (almost) self-employment at small scale. Conversely, countries with a low proportion of credit constrained people will grow over time aided by a high share of people being able to start-up business, of these surviving over time and with an active labor market paying high salaries.

To empirically test these hypotheses, we built a pseudo-panel using data from the Global Entrepreneurship Monitor (GEM) for the period 2001-2009. The pseudo-panel was complemented with income distribution indicators prevailing in 1700s and 1800s, and credit protection indicators.

In order to address reverse causality between the proportion of people involved in entrepreneurship and current business regulation, the econometric technique used instrumental variable estimators.

The main findings of this essay support the predictions of Banerjee and Newman's (1993) model. We found negative and persistent effects of inequality prevailing in 1800s' on the likelihood of countries' developing a healthy entrepreneurial sector, understood as firms being created, surviving and creating jobs over time. Also, the more efficient credit markets proxied by the legal right index are, the more likely is that countries' have larger proportion of people involved in entrepreneurial activities, and to these developing firms over time. In this essay we proposed that to foster entrepreneurship to grow and create jobs over time, countries should focus on reducing their inequality levels and improve credit market institutions.

The **third essay** analyzed whether long-term parental background, crystallized by parental educational background, race, cognitive and non-cognitive abilities, and short-term family income measured by the non-monetary opportunity cost of education, affect child' schooling progression, and at what stage of the educational path they take on their importance.

To this end, I used a sequential probability model, in which education attainment is the outcome of the individual's previous schooling decisions. This methodology allowed me to control for potential endogeneity issues arising from individual's unobservable heterogeneity and non-random selection of the sample that may occur at different educational stages.

I used the National Youth Survey and National Household Surveys conducted in 2008 from which I constructed individuals' educational path trajectories.

The main findings of this essay showed that the Uruguayan educational system is highly stratified, only allowing those individuals with better parental educational background, more able and motivated individuals, and non afro-descendants to attain higher educational levels.

Short-term parental income and long-term parental factors both influence children's schooling progression in Uruguay although they have different impact across the educational path. Specifically, short-term family income decreases its importance as students progress to higher schooling stages, whereas long-term parental factors turn to be more important the higher we move on the educational system. In particular, persistent and increasing effects of cognitive abilities on schooling progression are found. Socio-emotional factors, proxied by motivation in secondary level and risky behavior also influence children' schooling progression.

This essay supports policy interventions at different schooling stages. Policies intended to promote cognitive ability early in life and social and behavioral skills in adolescence and youth from a gender perspective and taking into account ethnical/ racial diversity may have positive effects on child's education achievement.

The thesis has overall provided evidence that initial conditions, whether immigrant's networks, country's initial wealth distribution or children's family background, affect development in the short and long-run. The findings shown here thus contribute to the literature and suggest important policy interventions.