Essays on Human Capital Formation, Living Standards and Selective Migration

Dissertation zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Dipl. Volkswirtin (int.) Yvonne Stolz
aus Eschenbach

Tübingen
2011

Tag der mündlichen Prüfung:	27.03.2012
Dekan:	Professor Dr. rer. soc. Josef Schmid
1. Gutachter:	Professor Dr. Joerg Baten
2. Gutachter:	Professor Dr. Martin Biewen

"I'd like to call the last part of the 20th century, and the beginning of the 21st century, the Age of Human Capital. Nowadays a primary determinant of a country's standard of living is how well it succeeds in utilizing the skills, knowledge and health of its people"'

Gary S. Becker: "Human Capital"

ACKNOWLEDGEMENTS

There are some people, I would like to particularly mention and thank for supporting me in writing this thesis:

I deeply appreciate the support, academic advice and guidance of my co-author, supervisor and mentor Professor Joerg Baten. His openness in sharing his experience and knowledge is truly acknowledged. He provided me with the opportunity to pursuit my own research questions, to present my work at international conferences, workshops and summer schools. I always enjoyed our co-authorships which all have been extremely productive. Furthermore, he gave me with the opportunity to teach – which was the most enriching experience for me – because those who teach actually learn the most (in this respect, I also want to thank all my students for providing me with this positive experience with their motivation and eagerness to learn).

Professor Martin Biewen's willingness to supervise this thesis as a second referee is gratefully acknowledged. He provided valuable econometric advice.

Special thanks also go to Professor Jaime Reis, co-author and deeply respected colleague, for stimulating joint research on Portugal, my "mother"-land. He enriched our work with his broad knowledge and experience on Portuguese economic history. I also thank him for his welcoming way in receiving me as a visiting research fellow at the ICS in Lisbon (University of Lisbon) in 2009.

Further, I would like to thank Professor Tarcísio Botelho, (Universidade Federal de Minas Gerais) for productive co-authorship on the Brazilian immigration experience and his welcoming way in receiving us in Rio de Janeiro.

I have received valuable comments and have had insightful conversations and discussions with Professor Pedro Lains, Dr. Nikola Koepke and Professor Claude Diebolt.

For discussions, comments and inspiration, I want to thank Kerstin Manzel and Valeria Prayon (also for joint work, both of you!), Linda Twrdek, Matthias Blum and Dominic Behle – it was great working with such great and talented people! Thanks also to the new generation of Tuebingen PhD students. Kristina Kohlmann and Eoin McLaughlin helped to improve the English.

Financial Support of the European Science Foundation, HIPOD and the German Science Foundation is gratefully acknowledged.

Finally, I would like to thank my family and friends for being there for me, whenever – no matter what. This thesis would not have been possible without your constant support.

Despite of all the support, it is me alone, who is responsible for the content of this thesis.

TABLE OF CONTENTS

ABBREVIATIONS

DFG	German Science Foundation
ESF	European Science Foundation
FE	Fixed Effects
GDP	Gross Domestic Product
GMM	Generalized Method of Moments
IPUMS	Integrated Public Use Microdata Series
IV	Instrumental Variable
LOWESS	Locally Weighted Scatterplot Smoothing
LSDV	Least Squares Dummy Variable
MEN	Middle Eastern Nations
NAPP	North Atlantic Population Project
OLS	Ordinary Least Squares
RE	Random Effects
UN	United Nations
WLS	Weighted Least Squares

Country abbreviations according to DIN ISO 3166

INTRODUCTION

1.1 A brief history of "Human Capital" in Economics

In 2004, the Association for the German Language[1] elected the term *human capital* ugliest word of the year, because it was perceived as degrading human beings to mere economically measurable factors.

The debate on the term "human capital" goes back to the 17th century, a time, when the discipline of economics was still placed within the wider range of philosophy. Sir William Petty (1623-1687) is said to have been the first to consider the value of human beings when estimating the wealth of a nation (Kiker 1966). While he was mainly interested in Public Finance, others have tried to estimate the value of human beings for purposes such as quantifying the power of a nation, the loss of human lives who die in war or the impact of international migration (ibid.). Around a century later, the founding father of economics, Adam Smith defined citizen's skills, talents and abilities as a form of capital in his seminal work *An Inquiry into the Nature and Causes of the Wealth of Nations* in 1776 (Laroche, Merette and Ruggeri, 1999). For Smith, being a liberal thinker and philosopher, the accumulation of human capital was a means that enhanced both individual productivity and enabled people to lead worthwhile lives in dimensions other than the economic sphere (Sen, 1997). John Stuart Mill, a 19th century economist and philosopher, who also tried to quantify the economic value of a human being, however, stressed that humans also were the "purpose, for which wealth exists" (Kiker, 1966, p. 486). The discussion until the early 20th century was then not so much about the ethical implications of the term but rather about how to define it in an economic sense. Should all human beings be included in the concept of human capital (as Léon Walras proposed) or only the productive ones (as suggested by Henry D. Macleod) (both cited in Kiker 1966)? Should the skills of an individual be defined as a value additional to the human himself (ibid.)? Alfred Marshall put an end to the discussion by defining the concept as "unrealistic" because human capital is not a marketable good (ibid.). Only in the 1960s of the last century the rational choice economists Theodore W. Schultz, Gary S. Becker and Jacob Mincer revived the application of the concept in economic theory. They provided *microeconomic foundations* to individual decisions such as human capital investments (Schultz 1961) or parental decisions of human capital investments in children (Becker 1960) or decisions on the marriage market (Becker

[1] In German: „Gesellschaft für Deutsche Sprache", http://www.unwortdesjahres.uni-frankfurt.de/, last accessed March, 14th, 2011.

1972). Differences in individual income were attributed to differences in human capital acquisition (Mincer 1958). It was only in the late 80s and the early 90s, that human capital became a very prominent concept in *macroeconomic theory*, as well. *Endogenous growth theory* states a positive relationship between economic performance and human capital (Lucas 1988, Romer 1994). These models endogenize technological progress which had, until this point, been treated as exogenous in *neoclassical models* (Solow 1956). Human capital accumulation was recognized as one of the major drivers of technological progress (Romer 1994). About that time, economists also began to search for models that were appropriate to explain the very long run, now explicitly including a historical dimension. Scholars working on *Unified Growth Theory* searched for models that were able to explain both, the long epoch of Malthusian stagnation and the transition to the process of modern economic growth which was only initiated by the industrial revolution at some point in the 19th century (Mokyr and Voth, 2010). Around the new millennium, Galor and Weil (2000) proposed a model that was to explain these processes – human capital at the centre of the mechanisms of the model - building largely on the microeconomic foundations of Theodore W. Schultz and Gary S. Becker. They depart from the idea that human capital is increasingly important over time when people have to adapt more and more rapidly to new technologies. Formal schooling and education are more appropriate to enhance people's capability to get accustomed to these processes, they argue, as opposed to informal training and learning by doing. Furthermore, the parental choice between education of children (child-quality) and family size (child-quantity) is one of the crucial mechanisms in explaining the demographic aspects of the transition from Malthusian stagnation to modern economic growth.

This short review highlights that, over the last some 50 years, human capital has become increasingly important in explaining both microeconomic and macroeconomic phenomena. Therefore, it is not surprising that renowned scholars refer to the 20th century and the beginning of the 21st century as the "age of human capital" (Becker 2002) or the "human capital century" (Goldin 2001). Other scholars take into account longer time periods, namely the period from 1870 until 2010 and refer to this extended 20th century as the "century of education" (Morrison and Murtin 2009). The roots for the global disparities in human capital formation, however, often go back well into historical times. Already at the very beginning of the 19th century, for example, differences in adult literacy across countries were huge. For Hungary, Tóth (2000), finds an adult literacy rate for people born around 1800 of somewhat over 22 percent, whereas in the US, census data shows, that already more than 80 percent of the population were literate (Ruggles et al., IPUMS). There-

fore, human capital studies with a historical perspective can shed light on global human capital disparities that exist today.

Empirically, establishing the macroeconomic relationship between human capital and growth has proven a challenging task. Among others, Capolupo (2008), Cohen and Soto (2007), and Woessmann (2003) show that the impact of human capital might have even been underestimated in previous research due to a lack of consistent and long-run data on human capital. Therefore, new data and innovative approaches to measure human capital are important. Also, providing insights on geographic regions and time periods where evidence on human capital has been scarce, so far, is a clear desideratum in human capital research. This thesis is a step towards filling the gap.

1.2 Human capital and international migration

The 19th century saw migration flows that were unprecedented in human history and shaped the face of the world as it is today (Hatton and Williamson 1998).[2] Immigration policies did not yet shape the direction and magnitude or the skill selectivity of the migration flows, as it is the case nowadays. Therefore, looking at the first global era can provide valuable insights for the migration phenomena which we observe around the globe, today. The first global era allows a glance at the fundamental mechanisms that underlie international migration. Understanding migration patterns will become increasingly important, as more and more people worldwide are living outside their countries of birth. Not only due to economic disequilibria, but also global phenomena like the climate change or the recent events in northern Africa and the Middle East are push factors that provide increasingly strong incentives for people to migrate. In 2010, some three percent of the world population were migrants. All in all, this is around 213.9 million people. More than half of them lived in only ten states. The US is still the "most wanted" destination country, Germany ranks on number three with over ten million migrants residing here.[3]

Hatton and Williamson (1998) found that 19th century migrants responded overwhelmingly to economic incentives. But while human capital selectivity of international migrants attracts a lot scholarly attention when it comes to current migration pattern (Michel A. R. Beine, Frédéric Docquier and Hillel Rapoport, to mention but a few scholars who work in the field), research on the 19th century is scarce.

[2] This dissertation deals with voluntary migration, only.

[3] All figures see: United Nations, Department of Economics and Social Affairs (UN/DESA): International Migrant Stock: the 2008 Revision. (Stand: Mai 2010). Development Prospects Group, World Bank.

Measuring human capital with respect to international migration was one of the major reasons the "human capital" concept was invented in the first place. Already in the 17th century, Sir William Petty pointed to the growth effects of immigration due to human capital gains. He, however, did not assert certain abilities and skills to the migrant but instead he calculated pure manpower (Vinokur 2006). By the beginning of the critical phase of the first global era – steamships had substituted sailing boats thus transport costs had decreased substantially- Friedrich Kapp predicted a massive gain for the United States from human capital (Kapp 1870, as cited in Kiker 1966). In the United States, however, a report of the immigration commission in 1911 concluded that immigration was impacting negatively on wages and working conditions (Hatton and Williamson 1998). In the economies that tried to adapt the industrial revolution pioneered in Great Britain, not crude manpower, but increasingly advanced skills like literacy and numeracy were the crucial factors. At the same time, mass schooling was globally on the rise and people in the 19th century were more literate and numerate than it had been the case 100 years before (Benavot and Riddle, 1988, Kaestle 1985). Human capital, embodied in migrants, adds to the total human capital stock in receiving nations. This has implications for sending countries' human capital endowment, as well. Only few economic history studies exist that explore human capital selectivity of international migrants quantitatively:

Looking at a *receiving nation*, Farbice Murtin and Martina Viarengo (2009) show why the US population was so reluctant to keep the doors open to new immigrants. They compare average years of schooling of the immigrants versus US natives for the period 1870 – 1930. They find a gap in schooling in comparison to the population of the destination country which even widened over time. In contrast, the selectivity pattern for the seventeenth and eighteenth centuries seems to be different: Abramitzky and Braggion (2006) present evidence that indentured servants to the West Indies were negatively selected, whereas those who went to North American colonies were positively selected in terms of their human capital. The authors used the length of servitude contracts to proxy unobservable characteristics of human capital like experience, education, health, and strength.

Looking at a prominent *source country* of 19th century migrants, Simone A. Wegge (2002) studied occupational self-selection of German emigrants to the US during the late 19th century. Comparing more than 10,000 emigrants to the native population of the source region, she found a positive selection in terms of the emigrants' occupations. Ray Cohn (2009) also studied selectivity of US immigrants during the first half of the 19th century using self-reported occupations of migrants. He states a decreasing positive selection in comparison to the source countries over the course of time.

The cited studies show that there exist several research gaps:

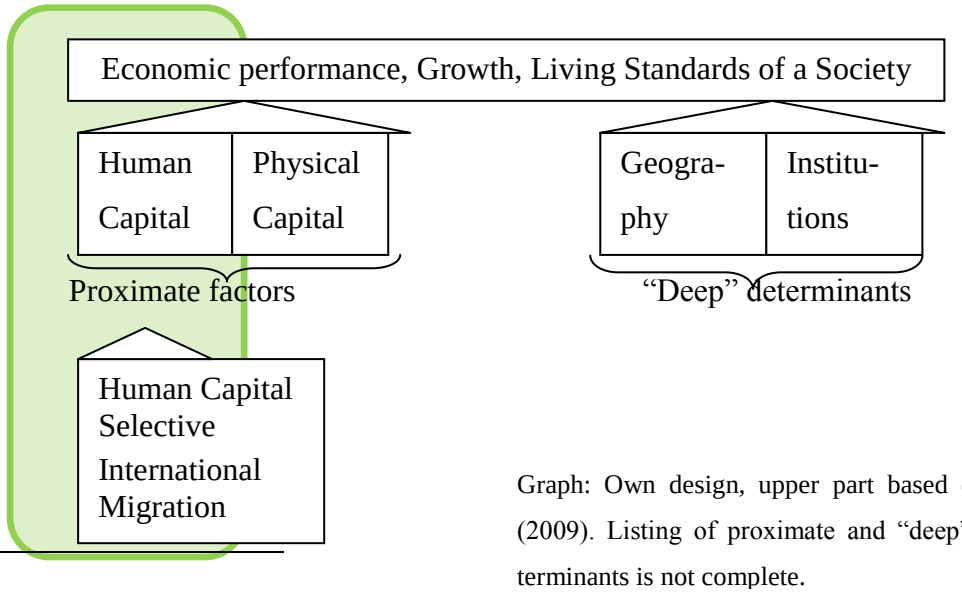
First, notwithstanding the fact that a lot of scholarly attention is focused on selective migration of the globalized world today; studies on skill selective international migration from a historical perspective are scarce.[4]

Second, so far, there exists no study on the topic that draws conclusions by comparing evidence on human capital selectivity from a variety of destination countries, taking a long-run perspective and an international view.

Third, the applied indicators in the existing literature differ across studies and exhibit severe limitations. For instance, average years of schooling (Murtin and Viarengo 2009) are an input measure: this proxy does not take into account different qualities of schooling institutions across source countries. As another example, taking the length of servitude contracts (Abramitzky and Braggion 2006) as a proxy for unobserved human capital characteristics does not consider that the merchant might have an incentive to keep a highly productive individual for a longer time period because it would maximize his profits. The use of self-reported occupations (Wegge 2002, Cohn 2009) has limitations with regard to the credibility of these statements: migrants might have exaggerated their occupational status.

As a conclusion, there is a demand for studies employing a comprehensive approach to capture human capital selectivity of international migrants in a long-run perspective, from an international point-of-view by using an output measure to proxy human capital. This thesis provides supply on this matter.

1.3 Aim and Outline of the thesis



Graph: Own design, upper part based on Capolupo (2009). Listing of proximate and "deep" growth determinants is not complete.

[4] However, there exist numerous studies on the evolution, magnitude and determinants of migration flows and the implications for sending and receiving nations, see for example Sanchez-Alonso (2000), Hatton and Williamson (1998), Balderas and Greenwood (2010).

The graph illustrates where the topics of the thesis can be located in a broader economic framework. Human capital is one of the proximate determinants of economic growth and living standards in general. Selective migration can impact on this important factor. Thus the thesis deals with the nexus of living standards, human capital and selective migration, framed by the shaded box. The aim is therefore

- To provide new data on human capital trends and human capital of migrants applying the age-heaping method as a proxy. Empirical evidence on geographical regions and periods, where not much is known, until now, is generated (namely Latin America and Portugal).
- To establish the empirical relationship between biological living standards and human capital in the very long run.
- To estimate the determinants of human capital selectivity of international migrants and assessing the impact of selective migration on source and destination countries.

To accomplish the goal, this thesis consists of six chapters of which five are intended, or, in one case, already accepted for publication. Three of the remaining articles are currently under review in refereed journals. One is to be submitted in the near future. The outline of the thesis is the following:

Chapter two focuses on human capital as an important determinant of living standards. Human capital is proxied with the age-heaping technique. Additionally in this article, literacy evidence is provided to cross-check the results. As an indicator of the biological standard of living, human stature is used. The chapter provides a large data set on Portuguese living standards from the early 18^{th} to the 20^{th} centuries. It answers the question: When and why did the Portuguese become the shortest Europeans? In order to find the answer to this question, the trend in Portuguese living standards from the 1720s until recent times is estimated with Maximum Likelihood and OLS techniques. The data shows that during the early nineteenth century average height in Portugal did not differ significantly from average height in most other European countries. But when around 1850, their anthropometric values began to climb sharply, Portugal's, however, did not. In OLS and IV panel estimations, delays in both real-wage convergence and human-capital formation in Portugal are found to be the main factors hindering any improvement there in the biological standard of living.

Chapter three provides data on human capital development for seven Latin American countries from the 17th to the 20th centuries, a geographical region and time period, where data was scarce until now. It is found that Latin America was on a path of conver-

gence with Western Europe during the early 18th century. During the early 19th century, not only did numeracy development stagnate in some Latin American countries but differences among some of them increased. While numeracy rates in Argentina, Uruguay, and to a lesser extent Brazil, underwent, along with Europe, a significant increase in the late 19th century, they declined in Mexico, Ecuador, and Colombia. It is found that mass immigration contributed to human capital formation in countries like Argentina, Uruguay and Brazil.

Chapter four analyzes the impact of human capital selective migration on destination countries. By providing new evidence on Brazilian human capital formation in the very long run and data on the human capital endowment of Brazilian immigrants during the age of mass migration, it is shown that human capital of international migrants can induce spillover effects to overall human capital accumulation that persist until today. It is argued that human capital formation is a highly path dependent and persistent process. In a panel of Brazilian regions, increases in numeracy are significantly positively related with the increase in absolute immigrant population per state. Thus the states that received most migrants also developed most quickly in terms of human capital.

Chapter five explores the question of whether relative inequality in source and destination countries matters for the brain-drain phenomenon. Human capital selectivity during the 1820s to the 1900s is analyzed. In a sample of 52 source and five destination countries selective migration is found to be determined by relative anthropometric inequality in source and destination countries. The results remain robust in OLS, IV and GMM approaches. These results confirm the Roy model of migrant self-selection (Roy 1951, Borjas 1987). Moreover, the evidence shows that countries like Germany and UK experienced a small positive effect, because the less educated emigrated in larger numbers.

Chapter six provides a new data set on human capital selectivity of female migrants during the first global era. Census data for five major immigrant countries show that during the 19th century migration, women made up above 40 percent of migrants, whose experience has been somewhat neglected by scholars, so far. The determinants of female human capital selectivity are estimated. In the empirical analysis, it is shown that skill selectivity of married women are driven by relative inequality, whereas single women's migration decision is determined by proxies for migration costs such as distance and common language. The results shed light on the female experience of the era of mass migrations.

The thesis ends with a summary and provides some directions for further research.

1.4 Definition of Human Capital

In this thesis, human capital is defined as proposed by Gary S. Becker (2002):

> "They are called human capital because people cannot be separated from their knowledge, skills, health, or values in the way they can be separated from their financial and physical assets."[5]

Although Amartya Sen (1997) has argued to clearly distinguish between *human capital* and *human capability*, I argue that human capital in all its facets is an enabling and empowering factor for people. Therefore, the concepts today overlap to a great extent. Because first, human capital influences economic output as a proximate determinant, second, it is, "directly relevant to the well-being and freedom of people" (Sen 1997, p. 1960), because education allows people to make more informed decisions, to participate in social life, it enhances opportunities not only on the job market but also for example, on the marriage market, and human capital also enables people to simply enjoy literature or the fine arts. Third it also influences social change because more informed and educated citizens tend to participate more actively in social life. These aspects show that human capital produces *positive external effects*, a fact, which has been stressed by human capital researchers (Laroche, Mérette and Ruggeri 1999, Davies 2002).

However, when measuring human capital, it must be acknowledged that empirical outcomes depend largely on the proxy that is used. Different indicators capture different aspects of human capital, which forces the user to interpret the results accordingly. The author is well aware, that human capital is a multifaceted concept, including all sorts of talents, skills and abilities embodied in people. The ability to have new ideas, being motivated or being risk-friendly can be seen as part of human capital. These abilities are innate abilities. When using indicators to measure human capital, however, economists have to turn to aspects of human capital that are observable which is why we use indicators.

1.5 Methodological aspects: age-heaping as an output measure of human capital

This thesis applies the concept of age-heaping as a proxy of human capital (Mokyr 1983, Fitzpatrick 1986, A'Hearn, Crayen and Baten 2009, Humphries and Leunig 2009). This method exploits the fact that people living in societies with low human capital typically tend to round their age to convenient multiples of five instead of reporting their exact age

[5] http://www.econlib.org/library/Enc/HumanCapital.html, last accessed April, 15th, 2011.

(see Crayen and Baten 2009). The share of multiples of five to the other numbers in the age distribution is expressed by the Whipple Index (A'Hearn, Baten, Crayen 2009).

(1) 

For a more intuitive interpretation, the Whipple-Index is transformed here into the ABCC-Index, which is a linear transformation of the former (ibid.).

(2)  if $Wh \geq 100$; else $ABCCI = 0$.

It ranges between 0 and 100, where zero stands for an age distribution where everybody states a rounded age and 100 represents no heaping at all. Its coherent construction allows comparability over space and time (A'Hearn Baten Crayen 2009). Furthermore, the indicator has proven robust when applied to different data sources (Crayen and Baten 2009, Manzel, Baten and Stolz 2011). One of the major determinants of numeracy is schooling investments as a major input to mass education (Crayen & Baten 2009). The availability of age statements in historical documents allows drawing evidence from a wide range of sources thereby providing data on regions and time periods where empirical evidence was scarce until now.

In recent studies, the indicator has proven to be valuable in a wide range of topics in human capital research (see, for example Humphries and Leunig 2009, Cinnirella 2008, Clark 2007). For a society at an early stage of human capital formation, the age heaping approach is a suitable indicator to proxy human capital. Usually, societies reach higher numeracy values before they reach better levels of literacy, as knowing one's exact age is a basic skill compared to reading and writing. Nevertheless, the indicator is highly correlated with literacy and other human capital indicators (Crayen and Baten 2009, Manzel Baten and Stolz 2011)

1.6 References

A'Hearn, Brian, Baten, Joerg and Crayen, Dorothee (2009): "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital" *Journal of Economic History* 68, no. 3, pp. 783-808.

Abramitzky, Ran and Braggion, Fabio (2006): "Migration and Human Capital: Self-Selection of Indentured Servants to the Americas" *Journal of Economic History* 66, no. 4, pp. 882-905.

Balderas Ulysses J. and Michael J. Greenwood (2010): "From Europe to the Americas: a comparative panel-data analysis of migration to Argentina, Brazil, and the United States, 1870–1910", *Journal of Population Economics* 23, No. 4, pp. 1301-1318.

Becker, Gary S. (1960): "An Economic Analysis of Fertility." In: *Demographic and Economic Change in Developed Countries*. Universities-National Bureau Conference Series, no. 11. Princeton, N.J.: Princeton Univ. Press (for NBER).

Becker, Gary S. (2002): "The Age of Human Capital", in E. P. Lazear (ed.): *Education in the Twenty-First Century.* Palo Alto: Hoover Institution Press, pp. 3-8.

Benavot, Aaron and Riddle, Phyllis (1988): "The Expansion of Primary Education, 1870-1940: Trends and Issues" *Sociology of Education* 61, no. 3, pp. 191-210.

Borjas, G.J. (1987): "Self-Selection and the Earnings of Immigrants" *The American Economic Review* 77, no. 4, pp. 531-553.

Capolupo, Rosa (2009): "The New Growth Theories and Their Empirics after Twenty Years", Economics: The Open-Access, Open-Assessment E-Journal 3.

Cinnirella, Francesco "Optimists or Pessimists? A Reconsideration of Nutritional Status in Britain" *European Review of Economic History* 12, no. 3, (2008): 325–354.

Clark, Gregory. *A Farewell to Alms. A Brief Economic History of the World*. Princeton and Oxford: Princeton University Press, 2007.

Crayen, Dorothee (2008): *Quantitative Analyzes of Long-run human capital development: age-heaping as an indicator for numeracy*. Inaugural-Dissertation zur Erlangung des Doktorgrades der Wirtschaftswissenschaftlichen Fakultät, Universität Tübingen.

Crayen, D., Baten, J. (2009): "Global Trends in Numeracy 1820–1949 and its Implications for Long-term Growth" *Explorations in Economic History* 47, no. 1, pp. 82-99.

Cohen, Daniel and Soto, Marcelo (2007): "Growth and Human Capital: Good Data, Good Results" *Journal of Economic Growth* 12, No. 1, pp. 51-76.

Cohn, R.L. (2009): *Mass Migration under Sail: European Immigration to the Antebellum United States*. Cambridge University Press. Cambridge.

Davies, Jim (2002): "Empirical Evidence on Human Capital Externalities" Working Paper, Department of Economics, University of Western Ontario, Canada.

Fitzpatrick, David (1986):"'A share of the honeycomb': Education, emigration and Irishwomen" *Continuity and Change* 1, no. 2, pp. 217 – 234.

Galor, Oded and Weil, David N. (2000):"Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and beyond" *The American Economic Review* 90, no. 4, pp. 806-828.

Goldin, Claudia (2003): "The human capital century" *Education Next* 1, pp. 73-78.

Hatton, Tim J. and Williamson, Jeffrey G. (1998): *The Age of Mass Migration: Causes and Economic Impact*. Oxford University Press. New York.

Humphries, Jane and Leunig, Tim (2009): "Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London?" *Explorations in Economic History* 46, no.1, pp. 120-131.

Kaestle, Carl, F. (1985): "The history of Literacy and the History of Readers" Review of Research in Education 12, pp. 11-53.

Kiker, B.F. (1966): "The Historical Roots of the Concept of Human Capital" Journal of Political Economy Vol. 74, No. 5 (Oct., 1966), pp. 481-499

Laroche, Mireille, Mérette, Marcel and Ruggeri G.C. (1999): "On the concept and Dimensions of Human Capital in a Knowledge-Based Economy Context" *Canadian Public Policy – Analyze de Politiques* 25, no. 1, pp. 87-100.

Lucas, Robert E. Jr. (1988):"On the mechanics of Economic Development" *Journal of Monetary Economics* 22, no2., pp. 3-42.

Manzel, Kerstin, Baten, Joerg and Stolz, Yvonne (2011): "Convergence and Divergence of Numeracy: The Development of Age Heaping in Latin America, 17th to 20th Centuries", *Economic History Review*, forthcoming.

Mincer, Jacob (1958): "Investment in Human Capital and Personal Income Distribution" *Journal of Political Economy* 66, no. 4, pp. 281-302.

Mokyr, Joel, (1983): *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850*. London and Boston.

Mokyr, Joel and Voth, Joachim (2010): "The demographic transition and human capital", In: *The Cambridge history of Modern Europe Vol. 1: 1700 – 1870*, pp. 43-69.

Morrison, Christian and Murtin, Fabrice (2009): "The century of education", CEE discussion papers, CEEDP0109. Centre for the Economics of Education, London School of Economics and Political Science, London.

Murtin, Fabrice and Viarengo, Martina (2010):"American education in the age of mass migration", *Cliometrica* 4, no. 2, pp. 113-139.

Romer, Paul M. (1994): "The Origins of Endogenous Growth" *The Journal of Economic Perspectives* 8, no. 1, pp. 3-22.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3, pp. 135–46.

Ruggles,, S., J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010

Sánchez-Alonso, Blanca (2000): "Those Who Left and Those Who Stayed Behind: Explaining Emigration from the Regions of Spain, 1880–1914" *The Journal of Economic History* 60, pp 730-755.

Schultz, Theodore W. (1961): "Investment in Human Capital" *The American Economic Review* 51, no. 1, pp. 1-17.

Solow, Robert, M. (1956): "A Contribution to the Theory of Economic Growth" *The Quarterly Journal of Economics* 70, no. 1, pp. 65-94.

Sen, Amartya (1997): "Editoral: Human Capital and Human Capability" *World Development* 25, no. 12, pp. 1959-1961.

Tóth, Istvan, G. (2000): *Literacy and Written Culture in Early Modern Central Europe*. Budapest, CEU Verlag.

United Nations, Department of Economics and Social Affairs (UN/DESA): International Migrant Stock: the 2008 Revision. (Stand: Mai 2010). Development Prospects Group, World Bank

Vinokur, Annie (2006): "Brain drain migration revisited", *Globalization, Societies and Education* 4, pp. 7-24.

Wegge, S.A. (2002): "Occupational Self-selection of Nineteenth-Century German Emigrants: Evidence from the Principality of Hesse-Cassel" *European Review of Economic History* 6, no. 3, pp. 365-394.

Woessmann, Ludger (2003):"Specifying Human Capital" *Journal of Economic Surveys* 17, no. 3, pp. 239-270.

2. PORTUGUESE LIVING STANDARDS 1720-1980: HEIGHTS, INCOME, AND HUMAN CAPITAL

2.1 Abstract

When and why did the Portuguese become the shortest Europeans? In order to find the answer to this question, we trace the trend in Portuguese living standards from the 1720s until recent times. We find that during the early nineteenth century average height in Portugal did not differ significantly from average height in most other European countries but that when, around 1850, their anthropometric values began to climb sharply, Portugal's did not. We conclude that delays in both real-wage convergence and human-capital formation in Portugal were the chief factors hindering any improvement there in the biological standard of living.

This chapter is based on a paper written jointly with Jaime Reis (ICS, University of Lisbon) and Joerg Baten (University of Tuebingen). It is accepted by the *Economic History Review*. The concept of the paper was developed jointly, analyzes and writing equally shared.

2.2 Introduction

It was Portugal's fifteenth- and sixteenth-century seafaring scientists and explorers who placed this small country at the cutting edge of maritime science and navigation and thereby transformed the global status of the rest of Europe as well. Their expeditions along the African west coast and to Latin America, not to mention Vasco da Gama's discovery, on his first voyage (1497-1499), of the sea route to India, transformed their contemporaries' consciousness of the world that lay beyond the horizon. However since those glory days the Portuguese have come to feel that their country lags behind the rest of Europe -- and they are right, if anthropometry is taken as welfare indicator: recent data indicate that the Portuguese are on average shorter than any other European nationality.[6] When and why did the Portuguese slip to last place in the European height race?

To tackle this question, we adopt a two-step approach. First, we reconstruct height development over almost three centuries, from the 1720s to the 1980s, and find that at the beginning of that time span Portuguese heights were in line with European heights generally. Second, we turn our attention to the possible causes of this evolution, focusing on real-wage development and human-capital formation as possible factors and find that it was the sluggish pace in Portugal of both real-wage convergence and human-capital formation that was the culprit.

More specifically, we have assembled a number of previously unexploited resources: a data set spanning eleven birth decades, from the 1720s to the 1830s, providing the first anthropometric documentation on eighteenth-century Portugal, and new regional data that complement those previously available for heights up until 1910.[7] No new sources are needed for the period 1910-1980, thanks to the comprehensive height statistics recorded by Sobral[8] and Padez.[9]

Taken together, these sources enable us to compare Portuguese stature with that of a sample from other European countries over the very long run and thus determine the timing and extent of the divergences among their respective biological living standards. It turns out that the Portuguese standard of living stagnated after 1840, at a time when other

[6] The height of Portugal's 1980s birth cohort (today in their 20s) is 172.1 cm, whereas that of Europe as a whole is 176.9 cm. Moldovans are in second-to-last place, at 174.2 cm (Baten and Blum, "Height"; and Sobral, "Secular Changes," pp. 491-504; for Portuguese data see Padez and Johnston, "Male Height," pp. 287-298; Padez, "Secular Trend," pp. 15-22; and www.undp.org); Reis, "Crescimento Económico," pp. 153-169; and Reis, "Urban Premium," pp. 69-94.

[7] Reis, "Urban Premium," pp. 69-94; and Reis, "Crescimento Económico," pp. 153-69.

[8] Sobral, "Secular Changes," pp. 491-504.

[9] Padez, "Secular Trend," pp. 15-22.

European regions developed sustained upward trends. It was thus in the 1890s that the Portuguese came to be the shortest population in Europe.

In addition, we present a major extension, back to 1720, of the existing series for Portuguese real wages and compile new information on relative food prices for a sample of European countries, including Portugal. Original estimates for Portuguese numeracy during the early eighteenth century, based on Inquisition records, are provided, as well as fresh data on nineteenth-century infant mortality.

Our study is composed of six sections. In the first we introduce the height data; we then go on to estimate the Portuguese height trend for the period 1720-1910. In the third section we compare this estimate with a sample from other European countries in order to determine whether the Portuguese have always been shorter than other Europeans: a determination that contributes to our knowledge of the biological standard of living throughout Europe. In the next two sections we scrutinize possible determinants of these differences and test our hypothesis with panel OLS and IV regressions. In the sixth and last section we offer our conclusions.

2.3 Data

Our data set spans nearly three centuries of Portuguese height development, if we include the twentieth-century data already studied by other authors.[10] While there have been height-development studies based on archaeological evidence which have covered even longer periods, this is the first time that such a long one has been treated in a single article based on archival height records -- a treatment possible only because military records preserved in various archives throughout Portugal contain high-quality information dating as far back as the eighteenth century (notes to Table 1). Drawing on the *Arquivo Histórico Militar* (Lisbon), we have collected large samples for the recruitment years 1763 and 1791, and a few smaller ones for the intervening years, as well as substantial samples for the 1820s-1840s (Table 1). We thus have a sufficient number of observations for all of the birth decades under consideration, with the exception of the number, slightly below our minimum, for the first of those decades, that of the 1720s.[11] Height information from the birth decades of the 1830s on being more abundant, we were able to collect a large sample for each of these later periods. We extended the already well documented post-1857 meas-

[10] Sobral, "Secular Changes," pp. 491-504; and Padez, "Secular Trend," pp. 15-22.

[11] With the exception of the 1720s (N=47), we include only those birth decades for which at least 50 cases per recruitment regime (see below for the definition of recruitment regimes in Portugal) are available; in most cases, N is far higher.

urement period data set[12] by 3,545 observations from the north of Portugal (Viana do Castelo, Braga, and Porto) preserved in the *Arquivo Geral do Exercíto* (Lisbon), to which we have added observations from the district archives of Évora, Faro, and Porto.

This data set, based on conscription records (the universal military service having been introduced in 1857), has been standardized to six Portuguese districts as they existed in the eighteenth century.[13] The fact that the sample spanning the second half of the nineteenth century provides thorough coverage, being representative of the sampled districts, means that when estimating the time trends for the entire period under study we can control for geographical composition.

We exclude extreme heights (below 140 and above 200 cm)[14] and analyze only those recruits from 18 to 50 years of age, thereby excluding all those who were still growing or already shrinking. We include dummy variables for ages 18-22, to control for late growth. Despite all of these precautions, our data feature several internal discrepancies, the most important of which is due to modifications over the years in the Portuguese military's minimum-height requirement. The following section explains how we solved this problem.

Minimum height requirement (MHR)

Prior to the introduction of the military draft in 1857, the only men whose heights were recorded in military records were those who not only passed the military's minimum height requirement (MHR) but actually served; hence the height evidence is truncated below the MHR. How does this truncation affect the selectivity of our data? The fact that the distribution falls short sharply below 62 *polegadas,* or inches, in 1763 and 1774 suggests that this was the cutoff point (Figure 1, Panel A). From 1776 to 1807 the height evidence is truncated at 60 *polegadas* but reveals, on the right side of the distribution, the presence of a grenadier contingent (N=169) composed of men chosen for their exceptional stature (Figure 1, Panel B). The period 1820-56 features two jumps, from 56 to 57 *polegadas* and then from 57 to 58 *polegadas* (Figure 1, Panel C). Thus it seems that prior to 1857 three successive MHRs were applied: between 1763 and 1774, 62 *polegadas*; between 1776 and 1807, 60 *polegadas*; and after 1820, 57 *polegadas.*[15] A review of the main regulations governing

[12] Reis, "Urban Premium," pp. 69-94; and Reis, "Crescimento Económico," pp. 153-69.
[13] Minho, Trás-os-Montes, Beira, Estremadura, Alentejo, and Algarve.
[14] The latter extreme was not reached.
[15] We performed truncated regressions with Stata 10, using 62, 60, and 57 inches (*polegadas*) as truncation points.

this issue indicates that this was indeed the case.[16] Finally, in 1857, with the establishment of general conscription, an MHR of 155 cm (56.8 *polegadas*) was established, and later, in 1887, reduced to 154 cm (56.6 *polegadas*).

The introduction, in 1852, of the metric system is another complication that has proved to be problematic for researchers. A thorough account of this problem of conversion and how it was overcome is found in Appendix A. Since the general-conscription regulations required all conscripts to undergo a physical examination that included having their height measured, the sample includes those who failed to meet the MHR or any other aspect of this examination. To control for the variations in the MHR prior to 1857, we use a truncated maximum likelihood method with varying lower limits, but we use simple OLS regression techniques for subsequent years because after 1857 the height evidence does not suffer from truncation

According to Dores Costa,[17] general conscription was extremely unpopular. among the Portuguese. Because both formal and informal protection networks enabled many ordinary persons as well as many of those in the upper social strata to avoid conscription altogether, data on these individuals, including but not limited to their heights, are slightly less frequent in the military records. Thus it is far more likely that those conscripts who actually served were (after one has factored in the MHR) a slightly negative sample of conscription-age males.

2.4 The Portuguese height trend 1720-1980

Given the complexity of the underlying data, several precautions are required in order to avoid introducing biases into an estimation of the Portuguese height trend for the period 1720-1980; our estimation takes into account these problems and their resolution (Table 2). The independent variables used include controls for army category, birth decade, age, and occupational and regional composition.

The trend is estimated by recruitment periods based on their different MHRs as defined in the previous section: 1763-1774, 1776-1807, 1820-1856 and 1857-1932. We aimed at running separate maximum likelihood truncated regressions for each recruitment

[16] The 1762 regulation of Count Lippe, the great Portuguese military reformer, stated that 62 *polegadas* was the MHR**.** In 1776, the Marquês de Pombal decreed a reform and an enlargement of the army (Selvagem, "Portugal Militar"). Although Selvagem does not refer to any change in the MHR at this time, it seems likely that this enlargement could have been achieved only by means of a broader recruitment base. A lower MHR, such as the 60 *polegadas* detected in the histogram of Figure 2, Panel B, therefore seems plausible. The next documented MHR, established by the Parliament in 1823, was 57.5 *polegadas,* also matches the available evidence. See Diario da Câmara dos Srs. Deputados da Nação Portuguesa.

[17] Dores Costa, *Insubmissão,* pp. 121-55.

regime. This would have been ideal, but it would have required a greater number of individual observations per birth decade than we had for the first of the four. Our solution was to combine the first two periods (Table 2, Column 1).[18] When we compared this combined regression with the one for the single recruitment period 1776-1807 (Table 2, Column 1), we found that the trends for the overlapping period were almost identical (Table 2, Columns 1 and 2; and Figure A.3, in the Appendix) and therefore concluded that the resulting time dummy coefficients could be used for the trend estimate (Table 2, Column 2).

Catch-up growth, for which we controlled with age dummies, correlates with changes over time in the biological standard of living in Portugal.[19] The constant and the birth-decade coefficients in combination with the coefficients of the age dummies revealed that the tallest recruits were found among the earliest group of recruits (Column 2); shorter ones were found in the next period (Column 3), and the shortest in the last one (Column 4). Column 2 also featured most of those who achieved their final height earliest; after the age of 18 their coefficients were no longer significant. The fact that in the last period (1857-1932) age dummies remained quite large and significant up to 21 and 22 years of age is in line with previous studies indicating that stunted individuals tend to undergo relatively late catch-up growth, during their early adult years.[20]

Birth-decade dummies and a constant were employed to generate an adjusted time trend that avoids the distortions associated with age, military rank, region, and occupation. For example, the constant in Column 2 yielded an estimate for a recruit from the district of Estremadura who belonged to the 1780s birth cohort (164.01 cm). Adding the coefficient for the birth decade of the 1750s (1.21), we obtain 165.22 cm, and so on for the other birth decades. We used the constants and time coefficients in Columns 3 and 4 to calculate estimates for the periods 1790-1830 and 1830-1910, respectively (that of 1830 is based on an average of Columns 3 and 4). We then moved beyond Estremadura, calculating the population shares of all six regions of Portugal during the eighteenth and nineteenth centuries, and added or subtracted the weighted coefficients for the five other regions. For example, the coefficient for the Minho region, -0.079, was multiplied by the Minho population

[18] In the truncated regression model implemented in the Stata software, several truncation points can be combined in one command.

[19] To control for catch-up growth, we experimented with age dummies for the 24-30-year age range and found no significant difference from adult height (the reference category). Controlling for early shrinking with an age dummy for the 45-50-year age range yielded no significant results either.

[20] Steckel, "Stature;" Komlos, "Shrinking."

share, 0.26 and added to the Estremadura height. The same calculation was performed on the other regions and resulted in regionally adjusted height estimates (Table 3).[21]

It is important in our next analysis that the data for the recruitment periods before and after 1857 – and also before and after 1820 -- are not severely sample-selected. Until we can determine whether or not the major developments in the height series were caused by the introduction of the general conscription in 1857, we cannot reject the hypothesis that the decline in heights from 1740 to 1820 was due to a changing selection. We begin with a consideration of the regional and occupational composition of the sample across all of the various recruitment regimes, move on to statistical tests of overlapping periods, and conclude with an examination of the height series by recruitment period.

Our occupation data tell us something about the socioeconomic composition of the army (even if many of the recruits were too young to know with any certainty how they would earn their living after military service) and therefore of the adult male Portuguese population in general. We use the Armstrong scheme of five occupational groups, which was designed to classify occupations in nineteenth-century censuses according to skill level,[22] and differentiate among the four recruitment periods (Figure 2). Each group of bars represents one occupational category, the individual bars within each group standing for each of the four recruitment periods. The first group, for example, represents those recruits whose occupation was recorded as unskilled worker, farmer, or unknown.[23] The share of recruits in this category during the first three recruitment periods was roughly 70 percent (that of the 1820-1856 period was slightly higher) and then declined sharply in the last period (1857-1936), by about ten percent: a decline that reflects the fact that during the late nineteenth century manufacturing composed an increasingly large portion of Portugal's

[21] The sources of our regional-population data are Guardado Moreira and Rodrigues Veiga: "Evolução da População," p. 40, and *Censo da População 1890* for the eighteenth and nineteenth centuries, respectively. We have decided not to apply an occupational adjustment with which we experimented because the censuses do not provide us with the occupational structure (groupings according to skill level, etc.) for the eighteenth and early nineteenth centuries. The best that we could do was calculate the occupational structure for the conscripts who were examined under the general-conscription law: a sample representative of the entire Portuguese population. Using the data for the first decade (1857-67), we arrived at an occupational structure of 82 percent unskilled workers, farmers, and unknown; three percent semi-killed workers/soldiers; eight percent skilled workers; four percent semi-professional, and (based on regional weighted averages) four percent professional. Given that the group comprising unskilled workers, farmers, and unknown is our constant and represents such a large percentage of the total, the upward adjustment would never be greater than one millimeter. This is too insignificant to justify the "cost" in terms of uncertainty about whether a similar occupational structure applied to the eighteenth century. Moreover, this means that we are comparing similar occupational sets, since many of the estimates from the comparison countries (Austria, Russia, the UK, etc.) are also based on samples representing the unskilled workers and farmers.

[22] Armstrong, "Use of Information," pp. 191-310.

[23] The fact that fewer than than five percent of the recruits in this group identified themselves as farmers indicates that farming was not considered a distinct occupation. We have no way of determining how many of those recruits whose occupation was listed as "unknown" may, in fact, have been farmers. "Unknown" may have also included some unskilled workers.

GDP and that skill levels were rising.[24] During three of the four recruitment periods, the largest percentage of recruits came from Lisbon or elsewhere in the capital's province, Estremadura. The exception was the third period, 1820-1856, when a larger percentage came from the adjacent provinces of Alentejo and the Beira (Table 4), a regional shift indicating that in order to estimate the general Portuguese height trend one needs to use both regional dummies and regional weighting.

We also determined that differences in the height values of recruits belonging to one and the same ten-year birth cohort recruited under different regimes were insignificant (Appendix, Table A.4). Since the first two periods' samples were smaller than we would have wished, we checked to see whether or not there were implausible gaps between one of the resulting height series and the next, and found that all of the important height-trend changes -- most notably the eighteenth-century rise and fall, the recovery between the 1790s and the 1820s, and then the next decade's stagnation –take place within, not between, recruitment periods (Figure 3), disproving the hypothesis that the 1740-1820 decline in heights was due to nothing more than changes in recruitment regulations.

Finally, we found that both adjacent and overlapping birth cohorts were quite similar (Figure 3). More specifically, the 1780s birth cohort (in the 1776-1807 recruitment period) was similar to the 1790s birth cohort (in the 1820-56 recruitment period), and the 1830s birth cohort in the 1820-1856 period was almost identical to the same cohort in the 1857-1932 one. This last observation is key to this paper's premise that the height development associated with the 1830s birth cohort was not caused by the introduction of the military draft in 1857.

2.5 Portuguese versus European Biological Living Standards 1720-1980

We compare the standard of living in Portugal with the average of a sample of European countries comprising Austria, England, France, Germany, Hungary, Italy, Ireland, the Netherlands, Russia, Spain, and Sweden, organized into three groups according to their geographical situation and their height averages: Southern Europe (Italy and Spain), Center Europe (Austria, England, France, Germany, Ireland, the Netherlands, and Sweden), and Eastern Europe (Hungary and Russia).

During the first half of the eighteenth century not only was the range of height values among these countries rather narrow (Figure 4, values in Table 3) but Portugal's average was close to theirs. In fact, prior to the 1760s average height in Portugal increased

[24] Lains, *Progressos*.

steadily, by a total of 1.5 cm, although this left it somewhere between two and four centimeters behind the Center group. It then declined until the 1790s, the nadir of the biological standard of living not just in Portugal but in Eastern and Center Europe as well.

The early nineteenth century marked a modest recovery not just in Portugal but also elsewhere in Europe, despite the political chaos and economic dislocation of the Napoleonic era and its aftermath; Portugal maintained a standard of living somewhere between those of Eastern and Center Europe until the birth decades of the mid-nineteenth century, the point at which, when it comes to average height, Portugal diverged from the rest of Europe, with an alarming large proportion of the 1840s birth cohort measuring under 140 cm – a clear sign of severe stunting. Yet for Portugal, as opposed to the rest of Europe, no recovery was in sight.

While Center and Eastern Europe enjoyed a long and steady improvement, Portugal stagnated until the end of the century. In the 1890s it had dropped to last place behind Southern Europe; by the time that the 1910 birth cohort was measured, the gap had increased to a good 5 cm. Portugal had to wait until the second decade of the twentieth century to see a sustained height increase, and even then it was nothing like the European one. Only a modest improvement was achieved and it was not until the 1950s, that socioeconomic change began to have a stronger positive effect on height.[25] More specifically, both nutrition and the public-health system improved between the 1940s and the 1970s, with the result that during this period life expectancy soared, from just over 50 to 71 years.[26] Meanwhile, however, the rest of Europe was making the same sort of, and in some cases even greater, progress, with the result that by 1980, instead of seeing their standard of living converge with that of the rest of Europe, the Portuguese found that they were still 7 cm shorter than the core European populations. Even during the Golden Age, of the 1950s to the early 1970s, when it had one of the fastest rates of economic growth in Europe, Portugal was unable to catch up with Southern Europe.

This comparative account of Portugal's long-run anthropometric performance exposes two puzzling phenomena: Portugal's lagging behind not only the rest of Europe after about 1850 but also the rest of Southern Europe (Italy and Spain) after about 1890. In the next section we will try to explain the gap between Portugal and the rest of Europe by focussing on the period, ending in 1910, when it was at its widest before the period of world wars.

[25] Padez, "Secular Trend," pp. 15-22.

[26] In 1902 a Portuguese newborn's life expectancy was on average 44 years. See Costa Leite, "População."

2.6 Determinants of the biological standard of living in Portugal

Stature is a function of nutritional quantity and quality, health conditions, and medical care during the early years of life. These components in turn are determined by socio-economic factors such as income -- particularly in the case of the poor -- public health, relative prices of protein-rich foodstuffs relative to other goods, and other factors.27 Family income, on the one hand, and the cost of healthcare as well as food and shelter, on the other, during one's childhood are thus reliable predictors of one's adult stature. All of these elements are interrelated; for instance, without adequate family income, foods containing high-quality protein, vital to a child's physical development, may be prohibitively expensive. In addition to economic variables, one's parents' values are responsible for basic human-capital formation, which in turn has an impact on one's biological well-being.

Since the food, shelter, and other goods that largely determine childhood and adolescent growth are determined by the parents' purchasing power, we begin by looking at the evolution of real wages in our comparison-panel countries. Since real wages are closely correlated with GDP, they are often used, in the absence of other measures, to represent income.28 The wages considered, as a rule, are those earned by unskilled labor in a given country's principal city and deflated by local prices.29 A recent estimate of long-run GDP per capita for Portugal30 that modifies Maddison's controversial calculations31 treats fifty-year intervals, which is far too infrequent for our regressions. We use decadal real-wage estimates instead.

Series for Lisbon real wages (both skilled and unskilled) during the nineteenth century were established some time ago on the basis of data drawn from the accounts of the royal household and charitable institutions in Lisbon.[32] A similar study currently under way extends the series back to the early eighteenth century. We apply Robert Allen's "welfare ratio" procedure, [33] which calculates the number of basic consumption baskets that a

---

[27] Steckel, "Stature," pp. 1903-1940.

[28] Allen, "Great Divergence," pp. 411-447; Allen, "Progress and Poverty," pp. 403-443; and Williamson, "Labor Markets,", pp. 141-196.

[29] While a database with a wider geographic scope than that provided by a single city would be more nearly representative of national trends, the use of price and wage data from Lisbon may be justified on several grounds. Portugal is a small country (89,000 square km), with Lisbon located roughly at its center. During the period under study the regional market system was drawn together by its rivers providing access to its seaports. By European standards (Federico, "European Grain Invasion") its markets for basic foodstuffs were quite well integrated in the eighteenth century (Justino, *Formação*) and even more so in the nineteenth. In addition, qualitative evidence regarding internal labor migration -- unrestricted during the period under study -- is plentiful (Reis, "Trabalho").

[30] Valério, "Portuguese Economic Performance."

[31] Maddison, *World Economy*.

[32] Reis, "Trabalho," pp. 119-151.

[33] Allen, "The Great Divergence," pp. 411-447.

family consisting of two adults and two children could purchase with the yearly income of a male adult laborer working a standard 250 days per year. This is based on an "average European" basket created for mid-eighteenth-century Strasbourg; we have adjusted it to account for variation in diets and in the availability of basic goods.[34] Our welfare ratio index, based on Lisbon prices and unskilled wages for the period 1720-1910, is in line with what is known about the broad lines of Portugal's economic evolution during this period (Figure 5).[35] If the living standard in the first half of the eighteenth century was stable, it was thanks to a combination of slow population growth with gold and diamond rushes in Minas Gerais and a thriving plantation economy on the coast of colonial Brazil. The steady decline in the welfare ratio over the course of the second half of the century was due to a more intense demographic pressure on resources coupled with the exhaustion of the earlier Brazilian boom and a weak productivity response in both the agricultural and manufacturing sectors. During the first two decades of the nineteenth century, the material conditions of the Portuguese recovered to their 1780s level, thanks to the stagnation of the population in tandem with a significant expansion in agricultural production, chiefly that of wine.

However, living conditions during the period from the 1840s to the 1860s were driven down to their lowest point since the century began by a series of natural disasters: the bad harvests of the "Hungry 1840s," the twenty-year-long oidium plague that slashed wine production by two-thirds, and a succession of catastrophic floods and coastal storms.[36] The shock to the agricultural sector, which accounted then for two-thirds of the total labor force, was severe. It was not until the 1870s that wages recovered, driven by a spurt in manufacturing, an increase in emigration to Brazil (yielding considerable home remittances), declining food prices due to grain imports, and a modest expansion in grain, wine, and olive-oil production. But the welfare ratio peaked in the 1880s, slipping back in the 1890s to where it had been nearly a century earlier. GDP per capita soon staged a mod-

[34] All of the main wage and price series used in this procedure were drawn from market situations and represent actual transactions. They were obtained from the accounts of religious and charitable institutions and the royal palaces. The wage series are derived from the building industry, as is the case in most of the literature, and represent the daily cash remuneration which was the only form of compensation. Prices were observed at different points throughout the year and collected for the following commodities: bread, wine, meat, olive oil, beans, eggs, hens, fuel, and linen cloth. For lack of data, housing was represented by a fixed proportion of total food consumption. In order to render our estimates of Lisbon "welfare ratios" comparable with those of other European regions, we substituted certain items in the so-called "Strasbourg consumption basket": e.g., wine and olive oil instead of beer and butter, respectively.

35 This analysis of Portugal's long-term economic evolution is based on several chapters in Lains and Silva, História Económica de Portuga, Vols. II and III.

[36] Joanaz, *Tempestades*.

est upturn, but the welfare ratio missed out on this owing to a decline in labor's share of total income.[37]

The next step is to compare this series with those for the principal cities of the set of countries used above for the height comparison in Figure 4. Until the 1740s, Lisbon's real-wage level is among the highest in Europe (Figure 6), but in the ensuing hundred years it stagnates and then declines, whereas those of both Center and Eastern Europe soar. By the end of the nineteenth century even those of Milan and Madrid, which had been at the back of the pack prior to the 1790s, are outperforming Lisbon's.

On the face of all this, real wages must be considered a potentially important determinant of heights. Previous studies in other countries however have detected important deviations in the income-height relationship, the best-known examples being the "early industrial growth puzzle", in England, and the "antebellum puzzle", in the U.S..[38] In Portugal there are also two observable deviations: the early eighteenth century, when the height increase outpaces that of wages; and a short period in the 1880s, when the reverse occurred. We therefore consider it necessary to put this hypothesis to the test in our regression analysis.

2.6.1 Human Capital

Gabriel Tortella and others have argued that underinvestment in human capital (that is, both education and practical skills) has hampered Southern European countries' long-term economic development.[39] Portugal, with its abysmal literacy and schooling rates, is a case in point. Around 1800, the rate of minimal literacy -- that is, the ability to sign one's name -- in rural regions was below 20 percent,[40] and in 1910 the situation was scarcely any better, both literacy and primary-school-enrollment rates being among the lowest in Europe. In the late nineteenth century the low standard of human capital in Portugal provoked frequent complaints on the part of entrepreneurs.[41] Progress was made in the twentieth century, but, as noted, the slowness of its evolution goes far towards explaining the fact that the biological standard of living in Portugal has long lagged behind that of other European countries.[42]

[37] Reis, "Trabalho," pp. 119-151.
[38] Komlos, "Shrinking"; and Komlos, "Anomalies"
[39] Tortella, "Economic Retardation," pp. 1-21.
[40] Reis, "Economic Growth."
[41] Reis, "Human Capital," pp. 22-48.
[42] Lains, *Progressos*, and Lains, "Catching-Up," pp. 369-386.

Investment in human capital manifests itself in two forms, education and income, both of which affect height trends. Education helps parents, and mothers in particular, to raise the standard of hygiene in their households and to provide their families with a nutritious diet, thereby improving their children's chances of achieving a healthy physical-growth rate. On the other hand, the parents' income level also determines the quality and quantity of food, shelter, and medical care received by the children. This effect of income will be captured by the real-wage variable in our regressions.

A database for literacy rates compiled for Portugal and the three regions that compose our European sample confirms Portugal's role as the laggard since at least as early as the 1830s (Figure 7).[43] Before this, however, data of this kind does not exist. The only human-capital indicator available for the years prior to the 1830s is of a far simpler sort than literacy: basic numeracy.

There has been a recent emergence of studies using data in which individuals report their age as an indicator of numeracy levels.[44] Brian A'Hearn, Dorothee Crayen, and Joerg Baten have shown that in societies characterized by a relatively low level of human capital, the frequency of errors on the part of individuals reporting their age is relatively high.[45] Their tendency to round off their age to the nearest multiple of five becomes evident in the frequency distribution. The ABCC index, which provides the percentage of a given population who are numerate -- that is, who do not round off -- is relevant to our study in that A'Hearn et al. have found a close correlation between this numeracy rate and the literacy rate. This correlation not only remains fairly constant over time but is robust when applied to different types of data as well.

Crayen and Baten provide evidence on numeracy in Portugal during the 1870s.[46] We have augmented this with the records of 108 individuals who were born between the 1700s and 1730s and were investigated by the Portuguese Inquisition.[47] Given that the Inquisition especially targeted Jews, whose educational level was, as a rule, above average, one might assume that the numeracy rate in these records would be above average as well, and thus introduce a bias into our calculations.[48] In fact, this may not be a problem since

[43] Eastern Europe is represented by Russia, Poland, and Hungary; Southern Europe by Italy and Spain (as above); and Center Europe by Austria, England, France, Germany, Ireland, the Netherlands, and Sweden.
[44] Mokyr, *Why Ireland*; Crayen and Baten, "Inequality"; A'Hearn, Baten, and Crayen, "Literacy"; Clark, *Farewell*; Cinnirella, "Nutritional Status"; and O'Grada, "Jewish Demography."
[45] A'Hearn, Baten, and Crayen, "Literacy."
[46] Crayen and Baten, "Numeracy Trends."
[47] We thank Kerstin Manzel and Rose Triebe for providing this evidence, obtained from the register of Inquisition records preserved in the Portuguese National Archive (*Torre do Tombo*).
[48] Botticini and Eckstein, "Human Capital," pp. 885-926.

only 20 percent of those in the sample were in more skilled occupations (lawyers, businessmen, and skilled artisans), the remaining 80 percent consisting of small farmers (51 percent), unskilled workers (22.2 percent), and semi-skilled artisans (6.5 percent). If there is any bias, it is in the opposite direction. Further gaps in the data were closed with Portuguese emigrants to Brazil.[49] We benchmarked their ABCC on the 1870s Portuguese national level of non-migrants and only used the change over time in the case of the decades between the 1820s and the 1870s for which other data were lacking.[50]

How does Portugal's numeracy rate compare with the rates of countries in our European sample? In Table 5, we present data for it between 1720 and 1910, as well as for our three European regions.[51] In the early eighteenth century Portugal was probably among the more numerate countries in Europe, but by about 1900 it was at the back of the pack. This is an indication that the underinvestment in human capital observed during the late eighteenth and early nineteenth centuries may have contributed to the decline in the numeracy rate and by extension to the subsequent decline in both height and income levels. We will therefore perform regressions to test the hypothesis of a positive correlation between the ABCC index and average height. In addition, wherever possible, we will check numeracy rates against literacy rates as a proxy for human capital.

2.6.2 Urbanization

A variety of historical studies have found a negative effect of urbanization on health trends and thereby on stature.[52] In the largest nineteenth-century cities, this was driven by rampant disease due to severe overcrowding and aggravated by the absence of proper public-health arrangements and an inadequate supply of foodstuffs. Though generally smaller, early modern cities also imposed a high urban penalty on their inhabitants, with life expectation at birth typically between 25 and 30 years.[53]

Since eighteenth-century Portugal's urbanization rate, at roughly 15 percent, was not far from the European mean,[54] one might reasonably presume that it was not a significant factor of Portugal's height trajectory compared with the rest of Europe during this century. The same conclusion is not as easy to draw however when it comes to the period from 1800 to 1910. On the one hand, Portugal's cities did not even begin to expand

[49] See Stolz, Baten, and Botelho, "Growth effects," for details.

[50] During the 1870s numeracy selectivity among emigrants and stayers was insignificant.

[51] Up to 1800 Eastern Europe is represented here by Hungary alone and thereafter also by Russia. Throughout the period under consideration Southern Europe is represented by Italy and Spain (as before), while Center Europe comprises Germany, Austria, France, the UK, Ireland, and Sweden.

[52] For a summary, see Komlos and Baten, "Looking Forward," pp. 191-210.

[53] O'Grada, "Jewish Demography", pp. 123-147.

until the 1850s, and then only modestly; in 1910 it remained one of the least urbanized countries in Europe. This should have helped narrow the height gap relative to the other countries. On the other hand, a recent study indicates that between 1840 and 1912 the biological standard of living in Lisbon -- despite the general perception that the city was profoundly unhealthy -- was somewhat superior to that in provincial, mostly rural, Portugal, even after factoring in differences in occupational structures.[55] This would contradict the beneficial impact of a slow urbanization on the relative height of the Portuguese. The net effect of 19th century urbanization on the biological standard of living is thus left unclear and can therefore only be properly ascertained by controlling for it in our regression analysis below.

2.6.3 Protein and grain supply and their relative prices

Baten has shown that protein supply and protein proximity play important roles in nutritional status.[56] In preindustrial societies, purchasing power being equal, nutritional status varied with the quantity of one's livestock, a source of high-quality protein. Cámara uses eighteenth-century Andalucía to show how a breakdown in the supply of high-quality protein could cause a decline in nutritional status.[57] Did Portugal experience a decrease in livestock production? Direct evidence that there was ever a decline in Portugal's livestock production is limited to the second half of the nineteenth century. Based on several animal censuses, it indicates that from 1852 to 1925 the total number of heads of livestock increased by a respectable 50 percent (Table 6).[58] However, when we factor in a population-growth rate of 70 percent and compositional changes in the livestock, what we have, in fact, is an overall decrease in meat consumption of ten percent for the entire period and of 16 percent if the years 1906-1925 are excluded. At the same time cereal and potato production kept pace with population growth, thanks to protectionist policies and a decline in the amount of land per capita, which encouraged a shift away from livestock production, causing (one can safely infer) a decline in nutritional quality, which would have contributed, in turn, to a decline in average height.

This scenario is corroborated by the evolution of the price relatives for grain and meat, which rose steadily during the second half of the 19th century[59], indicating that a growing scarcity of animal protein in relation to carbohydrates was indeed being felt at this

[54] Banks, *Cross-Polity*, pp. 55 ff.
[55] Reis, "Urban Premium," pp. 69-94.
[56] Baten, "Protein Supply," pp. 165-180.
[57] Cámara, "Long-Term Trends," pp. 47-63.
[58] Justino, *Formação*.
[59] Reis, "Crescimento Económico", pp. 153-169.

time. This implies a substitution in the diet of cereals for meat and a decline in the nutritional status of the population, in particular of the lower strata of society, who would have been more sensitive to these price alterations. It is confirmed by currently available evidence.[60]

For lack of data, we cannot calculate changes in diet, and by extension in nutrition, during earlier periods. We do know, however, that the market price of meat relative to that of grain steadily declined over the course of the eighteenth century and then, reversing course, increased until the late nineteenth century (Figure 8). To summarize, during roughly the first half of the period under analysis market conditions for nutritional inputs were increasingly favorable and during the second half were increasingly unfavorable to physical growth, an indicator of biological well-being.

To assess these findings in a broader context, we have compiled three series of European grain-meat price relatives.[61] Two patterns -- one involving trends over time, the other price levels -- are immediately apparent. First, the four time series move in tandem, declining throughout the eighteenth century and then rising throughout the nineteenth. This is not surprising when one considers the impact on the grain-meat price relationship of two other factors that moved together as well: demography (a proxy for pressure on fixed resources) and real per capita income, an indicator of the intensity of demand for more expensive goods. Second, as far as price levels are concerned, throughout the entire period the price of meat relative to that of grain was lower in Portugal than elsewhere in Europe. This is unexpected, since during the late nineteenth century, and probably earlier as well, its per capita consumption in Portugal was lower. Of course, relative prices are only a part of the story, one must also take into account disparities in demand driven by per capita differentials in purchasing power (Figure 6). Because in Southern Europe the primary concern was to fill bellies, the choice tended to be a high-quantity (calories) instead of a high-quality (protein) diet. As a result the Portuguese consumed a relatively large quantity of bread and a relatively low quantity of meat, despite the relatively low price (and high quality) of the latter, with potential consequences for their height evolution which we will shall test in our econometrics section.

2.6.4 Emigration

Emigration is another factor that can theoretically affect anthropometric outcomes. Emigrants sometimes are endowed with above-average height as well as human capital, and

[60] Reis, “Urban Premium”, pp. 69-94.

their departure can thus reduce the average-height level of those who remain behind. Emigration played an important role in Portuguese history, in particular during the late nineteenth and early twentieth centuries. Hatton and Williamson[62] report that decadal emigration rates increased substantially, from 1.9 percent during the 1860s to a rate of 5 percent in the first decade of the twentieth century, and then declined.[63] Portuguese emigration (Brazil was the main destination), however, does not seem to have been very selective. Baten, Pelger, and Twrdek found no significant height difference between Portuguese emigrants to Brazil and stayers.[64] The average height of 1,070 Portuguese male emigrants to Latin America (age 20-54, having been born between the 1850s and the 1880s) was at 165.1 cm, less than one centimeter greater than the overall Portuguese average. The fact that during this period the average numeracy level of Portuguese emigrants was not significantly different from that of stayers suggests that emigration was not among the factors that explain Portugal's anthropometric divergence from the rest of Europe.

2.6.5 Disease environment

Of course, height trends are also sensitive to the disease environment, since the body uses for growth only those nutrients that remain after it has used the rest for maintenance as well as for recovery from physical exertion. This means that unfavorable disease environments prevent individuals from achieving their full growth potential.

Infant mortality rates are often used as a proxy for ill health during early childhood. Unfortunately, records of the Portuguese infant mortality rate are limited to 1800, the 1860s, the 1890s, and the 1900s, a quantity insufficient for use in our regressions. However, it is safe to say that Portugal's rate was in line with that of countries in our Center Europe region -- which were better than the rates in the other two regions. This variable therefore does not help to explain the Portuguese height divergence from the European core (Figure 9). Our estimate includes, instead, urbanization rates, which can be viewed as an alternative proxy for the disease environment.

[61] Once again they represent, respectively, Southern Europe (Barcelona and Madrid); Center Europe (London); and Eastern Europe (Krakow and Danzig).
[62] Hatton and Williamson, *Age of Mass Migration*.
[63] Ferenczi and Willcox, *International Migrations*.
[64] Baten, Pelger, Twrdek, "Anthropometric History," pp. 319-333.

2.6.6 Social policies

As for the impact of European government policies on the biological standard of living since the focus was on trade policy, the primary aim being to raise government revenues, mostly through tariffs on luxury goods, such as tea and tobacco.[65]

In terms of social policy, Lindert has shown that the transfers were meager prior to the creation of the twentieth-century welfare state -- with perhaps one exception: the British Poor Law in effect from the 1790s to the 1830s. Britain was one of the first states to offer income transfers (outside of the workhouse) for adults who were not physically handicapped but otherwise needy. Since the poor were hit relatively hard by food shortages, such relief may have had a positive impact on average height. On the other hand, a Malthusian would argue that the net effect would be more children born to the poor, and blessed with a slightly better survival rate than that of their parents' generation, a combination of changes which could cause average height to decline.

As for Portugal, the only public policy that may have had a positive impact on the biological standard of living was the institution of the foundling hospital, underwritten by local governments and, to a lesser extent, by charitable institutions. In mid-nineteenth-century Portugal ten percent of newborns were abandoned, and about half of them died within a few years.[66] For Europe, this was hardly an exceptional figure and may even have been comparatively low. According to one estimate, the annual abandonment number for the Continent as a whole was 100,000;[67] the French rate was ten to twenty percent,[68] and that of Florence and Milan around thirty.

Foundlings are of interest here because they were one of the most disadvantaged groups in society in terms of the link between rearing and physical development. They were brought up, under institutional supervision, by impoverished women with a low human capital endowment, who were not their mothers and did this for purely pecuniary reasons. They were therefore prime candidates for stunting and should have lowered the average height of male cohorts as they emerged into adulthood. On the other hand, the same circumstances were responsible for an exceptionally high fatality rate – anything between

[65] Only in the event of a harvest failure did grain-exporting countries ever try to restrict food exports. Another major field of government activity was military affairs, the indirect effects of which we have previously discussed.

[66] Mapas estatisticos dos baptisados, casamentos e óbitos em cada uma das dioceses do continente do reino. Anno de 1860. Arquivo Nacional, Torre do Tombo, Lisbon.

[67] Kertzer, "Lives of Foundlings," pp. 41-56.

[68] Fuchs, *Abandoned Children,* p. 198.

300 and 700 per thousand. This would have raised the cohort's mean stature by virtue of its negative effect on the lower tail of the anthropometric distribution.

In Portugal, the guiding principle on foundlings, as elsewhere, was to minimize expenditure. Consequently, both these outcomes were present. In other words, the parsimony with which they were treated produced a disproportionate share of potentially undersized individuals in the population but also in practice a lesser number of them than if they had been brought up in ordinary households. As a result, the overall negative impact of the two influences on Portugal's anthropometric profile was likely to be small. Comparatively, the bias thus incurred, if anything, favored Portuguese heights vis-à-vis those of other countries, given that in the latter, child abandonment was on a greater scale.

Since this conclusion leaves the British Poor Law as the only government policy that could have influenced average height during the period under study, we insert a dummy variable in the regressions below to test for any such effects.[69]

2.6.7 Econometric test results

To test our hypotheses, we have aggregated the European countries into three regions, namely Southern, Center and Eastern Europe. With these three regions and Portugal we have estimated an unbalanced panel with average height as the dependent variable and real wages, human capital, relative prices for protein, urbanization and the poor law as the explanatory variables. Infant mortality could not be included but time fixed effects should pick up very serious European-wide health shocks such as the cholera epidemic during the 1830s. We restrict our analysis to the period 1720-1920, extending it no further because the twentieth-century series are characterized by a great degree of trend correlation. We thus obtain the longest time frame possible without encountering unit-root problems.

The empirical equation is as follows:

Heights= β_1 + β_2 * real wages + β_3 * human capital + β_4 * relative prices + β_5 * urbanization + β_6 * poor law + time fixed effects + regional fixed effects + ε

We took the logs of the price ratios and of the real-wage variables. For the regressions that included literacy among other explanatory variables, only nineteenth-century values were available (Appendix B). Because basic numeracy and literacy are, as a

[69] We are aware that assigning a value of one to the Center region for the existence of some form of social spending does not capture with any precision the effects of the British Poor Law, since other countries are included in this region as well. According to Lindert's study of social-welfare spending, that of the Netherlands, among others, was quite substantial, and France adopted several redistributive policies following the Revolution, an event that may have motivated the subsequent social-welfare spending in the Center regions, if not Scandinavia's, which was above average. Thus the dummy variable may be useful after all.

rule, acquired during the first decade of life, we organized the human-capital variable by birth decade. All estimations were weighted by average population size of the respective regions; the population figures refer to the countries used in constructing the other variables.[70] As the price of meat increases, so does that of protein, and heights decline. We therefore expect the relative price of meat to show a negative sign. On the other hand, both real wages and human capital should have a positive effect on heights, as opposed to the Poor Law, and perhaps also the urban-share effects.

Our first step was to regress heights on the human-capital proxies only while controlling for time effects with birth-decade dummy variables (Table 7). Both numeracy and literacy have a significant impact on our dependent variable (Table 7, Columns 1 and 2, respectively). When we include both real wages and human capital, the coefficients remain positive and significant but the coefficient size of both of these human-capital proxies decreases, indicating that the human-capital effect on heights works in part through income. When we control for regional and time-fixed effects, the results are robust: that is, both the range in magnitude of the coefficients and their significance levels remain unchanged (Table 7, Column 7). [71] The adjusted R^2 is satisfactory for all specifications and rises when real wages are included in the model. Our second step was to examine the effect of the relative price of meat on height.[72] The signs are negative in all specifications but, with one exception, insignificant, possibly because it is no more than an indirect indicator of protein availability; moreover, in economies such as those of our Center and Eastern regions, with low market integration, it was milk, not meat, which was probably the chief source of protein.

The coefficients of the urbanization variable are negative and not small, but insignificant in one case and significant in another specification. Although we doubted that this variable would help to answer the Portuguese-height question, it may in fact do so in that it may have pushed down height levels elsewhere in Europe. On the other hand, the effect of the Poor Law dummy variable is insignificant.

To mitigate the likelihood of endogeneity problems in the relationship between heights and human capital, we have instrumented numeracy with the share of Protestants and Jews in each of the European regions. This accords with the argument of Becker and

[70] Maddison, *World Economy*.

[71] Note that, on account of the small number of cases, in specifications with regional fixed effects we control for time fixed effects with half-century dummies.

[72] The variable cattle per capita proved to be insignificant, possibly because the productivity rate varied among regions.

Woessmann[73] that there was greater early human-capital formation in economies with a predominantly Protestant population than in those with a predominantly Catholic one. Given that by the eighteenth century the map of religious faiths in Europe had become quite well defined, we consider this share acceptable as an exogenous instrument. We have added the share of Jews in the population because similar religion-related educational ambitions apply to this religion (even if their numerical weight was relatively small).[74] The first-stage regression shows a significant relationship between the share of Protestants and lagged numeracy, and the second-stage one confirms our previous results. In addition, we obtain a coefficient for instrumented numeracy which is quite similar to the OLS outcome.[75]

A standard technique for determining the economic significance of such coefficients is to multiply those of the explanatory variables by their respective standard deviations. This allows us, for example, to compare the outcomes of average and high values of the explanatory variable. Thus if we multiply one standard deviation of numeracy (expressed between 1 and 0, as in Table 7), which is 0.108, by the numeracy coefficient of 14.43 in model (1), we get 1.6 cm. In other words, the model predicts that if numeracy were to rise by one standard deviation, the difference in anthropometrically measured welfare would be expressed by a change in height of 1.6 cm, an increase regarded in most of the literature as a substantial value. Hence the economic significance is given here. If we perform the same calculation with literacy (model 2, s.d.=0.262), the height increase would be 1.9 cm; with real wages, an increment of one standard deviation (0.386) would result in a height increase of 1.33 cm.

The development of these variables during the twentieth century provides another perspective on our results. Between 1910 and 1980, literacy in Portugal increased by slightly less than 40 percent, real wages by 152 percent, and heights by 8.3 cm. If we multiply the increase in literacy by its coefficient of 7.38 (Table 7, Column 2), we find that human capital alone, in our model, would have explained 7.38*0.4 = 2.95 cm of this height increase. However, the fact that the literacy coefficient decreased when real wages were included in the regression suggests that some of this change may have arrived by way of the income channel. If we multiply the wage increase during the same interval by its coefficient (Table 7, Column 5), we obtain 1.52*4.08= 6.2 cm, indicating that the

[73] Becker and Woessmann, "Was Weber Wrong?" pp. 531-596.
[74] The data are taken from http://www.hks.harvard.edu/fs/rmcclea/data.html, last accessed January 14th, 2011.
[75] The F-statistic is 15.17, p-value=0.000. Durbin and Wu-Hausman tests suggest that we cannot reject the hypothesis of the exogeneity of our instrument.

explanatory variables in our regression account for 75 percent of the Portuguese height increase during the twentieth century.

2.7 Conclusion

As recently as the year 2000, and despite decades of economic development, the Portuguese remained shorter than any other European nationality. Our aim in this paper was to determine whether they always held this record, and if not, then to determine both when and why there was a change in their average height. What we found is that in the early eighteenth century the average height of the Portuguese was on a par with that of other Europeans; that it began to diverge in the 1840s; that the divergence increased significantly during the 1870s; and that since 1890 they have been firmly positioned in last place. We scrutinized the determinants of Portugal's tardy height development by focusing on the period when the gap widened. Both OLS and IV results suggest that the modest real wage evolution, as a result of Portugal's comparatively late industrialization and slow economic growth performance, was one important determinant. Relative prices of protein and urban penalties seem to have played a negative, but less debatable role. On the other hand, we find delayed human capital formation to have been clearly of importance. If the Portuguese had achieved, for example, the average literacy level of the European center in 1860 - 0.87 instead of 0.21 - the birth cohort of the 1860s would, *ceteris paribus*, have been 7.38*0.66=4.9 cm taller and the gap which was then opening up with Europe would not have occurred.

2.8 References

A'Hearn, Brian "Anthropometric Evidence on Living Standards in Northern Italy 1730-1860." *Journal of Economic History* 63, (2003): 351-381.

A'Hearn, Brian, Baten, Joerg Crayen, Dorothee "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital" *Journal of Economic History* 68, no. 3, (2009): 783-808.

Allen, Robert C. "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War" *Explorations in Economic History* 38, (2001): 411-447.

Allen, Robert C. "Progress and Poverty in early modern Europe" *The Economic History Review* 56, no.3, (2003): 403-443.

Armstrong, W.A. "The use of information about occupations." In *Nineteenth Century Society: Essays on the Use of Quantitative Methods for the Study of Social Data*, edited by E.A.Wrigley, 191-310, Cambridge: Cambridge University Press, 1972.

Banks, Arthur S. *Cross-Polity Times Series Data*. London and Cambridge. MIT Press. 1971.

Baten, Joerg "Protein Supply and Nutritional Status in Nineteenth Century Bavaria, Prussia and France" *Economics and Human Biology* 7, (2009b): 165-180.

Baten, Joerg and Blum, Matthias "Global Height Trends and its Determinants, 1810-1984: An Overview" Working Paper Univ. Tuebingen, available under http://www.wiwi.uni-tuebing-en.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/baten_blum_ht_100331a.pdf last accessed March 31st, 2010.

Becker, Sascha O., Woessmann, Ludger "Was Weber Wrong? Human Capital Theory of Protestant Economic History" *The Quarterly Journal of Economics* 124, no. 2, (2009): 531-596.

Botticini, Maristella, Eckstein, Zvi "From Farmers to Merchants, Voluntary Conversions and Diaspora: A Human Capital Interpretation of Jewish History" *Journal of the European Economic Association* 5, no. 5, (2006): 885–926.

Cámara, Antonio D. "Long-Term Trends in Height in Rural Eastern Andalusia (1750-1950)" *Historia Agraria* 47, (2009): 47-63.

Carreras and Tafunell, "Historia Económica de la España contemporánea" Barcelona: Crítica, 2004.

Carrion and Lázaro "Was there an urban height penalty in Spain, 1840–1913?" *Economics and Human Biology* 5, no. 1, (2007): 144-164.

Cinnirella, Francesco, Komlos, John. "European Heights in the Early 18th Century" *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 94, no. 3, (2007): 271-284.

Cinnirella, Francesco "Optimists or Pessimists? A Reconsideration of Nutritional Status in Britain" *European Review of Economic History* 12, no. 3, (2008): 325–354.

Clark, Gregory. *A Farewell to Alms. A Brief Economic History of the World*. Princeton and Oxford: Princeton University Press, 2007.

Costa Leite, Joaquim da "População e Crescimento Económico" In *História Económica de Portugal 1700-2000. O Século XIX*, edited by Pedro Lains and Álvaro Ferreira da Silva, Lisbon, 2005.

Crayen, Dorothee, Baten, Joerg "Global Trends in Numeracy 1820-1949 and its Implications for Long-Run Growth" *Explorations in Economic History* 47, no.1, (2009a): pp.82-99.

Crayen, Dorothee, Baten, Joerg "New Evidence and New Methods to Measure Human Capital Inequality before and during the Industrial Revolution: France and the U.S. in the 17th to 19th Centuries", *The Economic History Review* 63, no. 2, (2009b): 452-478.

Da Silva Lopes, J.B. *Memoria Sobre a Reforma dos Pezos e Medidas em Portugal segundo o Sistema Métrico Decimal*. Lisboa, 1849.

Dores Costa, Fernando, *Insubmissão. Aversão ao serviço militar no Portugal do século XVIII*. Lisboa, 2010.

Federico, Giovanni "The first European grain invasion: a study in the integration of the European market 1750-1870" Working Paper, European University Institute, 2008.

Feliu, Gaspar (1991). *Precios y salarios en la Catalu˜na moderna*. 2 vols. Madrid: Banco de España.

Ferenczi, Imre and Willcox, Walter F., *International Migrations*, Vol I&II. New York, 1929.

Flora, Peter *State, Economy and Society in Western Europe: 1815-1975. A data handbook in two Volumes.* Frankfurt, M. Campus 1982.

Frank, Zephir "Stature in nineteenth-century Rio de Janeiro: preliminary evidence from prison records" *Journal of Iberian and Latin American History* 24, no. 3, (2006):465-489.

Fuchs, Rachel, G. *Abandoned children: foundlings and child welfare in nineteenth-century France*. State University of New York, Albany, 1984.

Garrabou, Ramon, Tello, Eenric, and Roca, Albert, "Preus del blat i salaris agrícoles a Catalunya (1720–1936)" In *La industrialització i el desenvolupament econòmic d'Espanya*, edited by Carreras, A. et. al., (1999): 422–60.

Gillis, A.R. "Literacy and the Civilization of Violence in 19th Century France" *Sociological Forum* 9, no.3, (1994): 371-401.

Graff, Harvey, J. *The Legacies of Literacy. Continuities and Contradictions in Western Culture and Society*, Bloomington, Indiana University Press, 1987.

Guardado Moreira M.J., Rodrigues Veiga, T. "A Evolução da População" In *História Económica de Portugal, 1700-2000. O século XVIII*, edited by Pedro Lains and Álvaro Ferreira da Silva (2005): pp. 35-65.

Hatton, Tim, Williamson, Jeffrey G. *The age of mass migration: causes and economic impact*. Oxford, 1998.

Heintel, Markus., Sandberg, Lars G., Steckel, Richard H. "Swedish Historical Heights revisited: New Estimation Techniques and Results" In *The Biological Standard of Living in Comparative Perspective*, edited by Joerg Baten and John Komlos, Stuttgart, 1998.

Hoffmann, W.G. *Das Wachstum der deutschen Wirtschaft seit dem 19. Jahrhundert*, Berlin, 1965.

Hofmeister, Andrea, Prass, Reiner, Winnige, Norbert „Elementary Education, Schools, and the Demands of Everyday Life: Northwest Germany in 1800" *Central European History*, (1998), 31: 329-384.

Hueso, Antonio D. C. "Fuentes antropométricas en España: problemas metodológicos para los siglos XVIII y XIX" *Historia Agraria* 38, (2006): 105-118.

Jacks, David S. "Market Integration in the North and Baltic Seas, 1500-1800" *Journal of Economic History* 33, no. 3, (2004): 285-329.

Jacks, David S. "Intra- and International Commodity Market Integration in the Atlantic Economy, 1800-1913" *Explorations in Economic History* 42, no. 3, (2005): 381-413.

Joanaz de Melo, Cristina "Contra Cheias e Tempestades: Consciência do Território, Debate Parlamentar e Políticas de Águas e de Florestas em Portugal, 1852-1886" Ph.D. Thesis, European University Institute , Florence, 2010.

Johanson, Egil “The history of literacy in Sweden” In *Literacy and social development in the West*, edited by H. J. Graft (Ed.), Cambridge: Cambridge University Press. (1981): 151-182.

Justino, David A formação do espaço económico nacional: Portugal, 1810-1913; Lisboa, 1988:

Kertzer, David I. "The Lives of Foundlings in Nineteenth Century Italy", In *Abandoned Children,* edited by Catherine Panther-Brick and Malcolm T. Smith, Cambridge: Cambridge University Press, (2000): 41-56.

Klein, Herbert “Asian and European Migration to Brazil 1880-1930” In *The Cambridge Survey of World Migration*, edited by Robin Cohen, Cambridge University Press, (2005): 205-208.

Komlos, John “Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution” *Journal of Economic History* 58, no. 3, (1998): 779-802.

Komlos, John “Anomalies in Economic History: Toward a Resolution of the “Antebellum Puzzle” *The Journal of Economic History* 56, no. 1, (1996): 202-214.

Komlos, John, Baten, Joerg “Looking Backward and Looking forward. Anthropometric Research and the Development of Social Science History” *Social Science History* 28, no 2, (2004): 191-210.

Lains, P. *Os progressos do atraso: uma nova história económica de Portugal, 1842-1992*. ICS, Lisbon, 2003.

Lains, P. “Catching-up to the European core. Portuguese Economic Growth 1910-1990” *Explorations in Economic History 40*, no. 4, (2003): 369-386.

Lains, Pedro, Ferreira da Silva, Álvaro (eds). *História Económica de Portugal*. 1700-2000, ICS, 3 Vols. Lisbon, 2005.

Maddison, Angus *The World Economy: A Millennial Perspective*, Development Centre Studies. OECD, Paris, 2001.

Martinez Vara, T. “Una estimación del coste de la vida en Santander, 1800-1860” *Revista de Historia Económica,* 15, no. 1, (1997): 87-124.

Messias, Erick “Income Inequality, Illiteracy Rate, and Life Expectancy in Brazil” *American Journal of Public Health* 93, no. 8, (2003): 1294-1296.

Mironov, Boris N. “The Development of Literacy in Russia and UDSSR from the Tenth to the Twentieth centuries” *History of Education Quarterly* 31, no. 2, (1991): 229-252.

Mokyr, Joel *Why Ireland Starved. A quantitative and analytical history of the Irish Economy 1800-1850*. London, 1983.

Nogués Marco, Pilar "Análises de la deflación española de la primera mitad del siglo XIX: una comparación international" *Revista de Historia Económica Año XXIII*, n. 2 (2005): 371-405.

Nos, Nicolau, Pujol Andreu "El consumo de proteínas animales en Barcelona entre las décadas de 1830 y 1930: evolución y factores condicionantes" *Investigaciones de Historia Económica*, no.3, (2005): 101-134.

O'Grada, Cormac "Dublin Jewish Demography a Century Ago" *The Economic and Social Review* 37, no. 2, (2006): 123-147.

Padez Cristina; Johnston F. "Secular trends in male adult height 1904-1996 in relation to place of residence and parent's educational level in Portugal" *Annals of Human Biology* 26, no. 3, (1999): 287-98.

Padez, Cristina "Secular Trend in Portugal" Journal of Human Ecology 22, (2007): 15-22.

Perez Picazo, Maria T. *Oligarquía urbana y campesinado en Murcia 1875-1902*. Murcia: Academia Alfonso X el Sabio, 1986.

Reis, Jaime "How Poor was the Periphery before 1850? The Mediterranean versus Scandinavia" In *The Mediterranean Response to Globalization before 1950*, edited by in Jeffrey Williamson and Sevket Pamuk, Routledge, London, (2000): 17-44.

Reis, Jaime "Crescimento Económico e Estatura Humana. Há um Paradoxo Antropométrico em Portugal no Século XIX?" *Memórias da Academia das Ciências de Lisboa, Classe de Letras*, XXXV, (2002/2003): 153-169.

Reis, Jaime "Human Capital and Industrialization: The Case of a Late Comer - Portugal, 1890" In *Technology and Human Capital in Historical Perspective*, edited by Jonas Ljungberg, Basingstoke, Palgrave-Macmillan, (2004): 22-48.

Reis, Jaime "Economic Growth, Human Capital Formation and Consumption in Western Europe Before 1800" In *Living Standards in the Past. New Perspectives on Well-Being in Asia and Europe*, edited by Robert C. Allen, Tommy Bengtsson and Martin Dribe, (2005): 195-226.

Reis, Jaime "«Urban Premium» or «Urban Penalty»? The case of Lisbon, 1840-1912" *Historia Agraria XIX* 47, (2009): 69-94.

Schaumburg-Lippe-Bückeburg, Friedrich Wilhelm Ernst Graf zu (1762): Regulamento para o Exercício e Disciplina dos Regimentos de Infantaria dos Exércitos de Sua

Magestade Fidelíssima feito por ordem do mesmo senhor por sua Alteza o conde Reinante de Schaumburg, marechal Lippe (Lisboa, Régia Oficina Tipográfica, 1794)

Selvagem, C. *Portugal Militar. Compêndio de História Militar e Naval de Portugal. Desde as Origens do Estado Portucalense até ao Fim da Dinastía de Bragança*. Edição Suplementar. Lisboa, 1931.

Sobral, F "Secular Changes in Stature in Southern Portugal between 1930 and 1980 According to Conscript Data" *Human Biology* 62, no. 4, (1990): 491-504.

Spijker, Jeroen J. A., Pérez, Julio, Cámara, Antonio D. "Cambios generacionales de la estatura en la España del siglo XX a partir de la Encuesta Nacional de Salud" *Revista Estadística Española* 169, no 50, (2008): 571–604.

Steckel, Richard H. "Stature and the Standard of Living" *Journal of Economic Literature* 33, no. 4, (1995): 1903-1940.

Stolz, Y. Baten, J., Botelho, T. "Growth effects of 19th century mass migrations: evidence from Brazil" University of Tuebingen Working Paper, 2010.

Tortella, Gabriel "Patterns of Economic Retardation and Recovery in South-Western Europe in the nineteenth and twentieth Centuries" *Economic History Review* 47, no. 1, (1994): 1-21.

Tóth, István, G. *Literacy and written culture in Early Modern Central Europe*, Budapest, CEU, 2000.

Valério, Nuno"Portuguese Economic Performance, 1250-2000" In *Las Claves del Desarrollo Economico. Ensayos en Homenaje al Professor Gabriel Tortella* edited by Juan Hernandez Andreu, José Luis Garcia Ruiz, José Morilla Critz and José Manuel Ortiz-Villajos, Madrid, Universidad de Alcalá Lid, 2010.

Van Zanden, Jan Luiten, de Vries, Jan, Smits, Jan Peter, van Riel, Arthur: The prices of the most important consumer goods, and indices of wages and the cost of living in the western part of the Netherlands, 1450-1800 http://www.iisg.nl/hpw/brenv.php.

van Zanden, Jan Luiten, Baten, Joerg, Foldvari, Peter, van Leeuwen, Bas "World Income Inequality 1820-2000" Working Paper (2010).

Williamson, Jeffrey G. "The Evolution of Global Labor Markets since 1830: Background Evidence and Hypothesis" *Explorations in Economic History*, no. 32, (1995): 141-196.

Williamson, Jeffrey G. "Real Wages and Relative Factor Prices in the Third World 1820-1940: The Mediterranean Basin" Discussion Paper No. 1842 Harvard Institute of Economic Research. Cambridge, 1998.

2.9 Tables

Table 1: Number of cases by recruitment decade and by birth decade

Recruitment decade	N	Birth decade	N
1760	822	1710	1
1770	184	1720	47
1790	675	1730	285
1800	226	1740	525
1820	417	1750	206
1830	528	1760	301
1840	468	1770	426
1850	3487	1780	143
1860	5574	1790	166
1870	8541	1800	331
1880	10860	1810	424
1890	2633	1820	596
1900	4034	1830	4029
1910	3058	1840	5144
1920	2819	1850	9794
1930	2314	1860	9530
		1870	2653
		1880	3872
		1890	3161
		1900	2703
		1910	2303
Total	46640	Total	46640

Notes: In this compilation, recruits younger than 18 or older than 50 were excluded, as well as extreme heights (<140 cm, > 200 cm, although the latter extreme was not reached).

Height data sources: Arquivo Distrital do Porto (310 cases), Arquivo Geral do Exército (3545 cases), Arquivo Histórico Militar (2841 cases), Arquivo Distrital de Évora (3140 cases), Reis (2002/03, 2009) (36804 cases), based on archival sources from the following archives: National archives, Torre do Tombo: http://antt.dgarq.gov.pt/ There are also links to Évora and Porto on the national archive page; Arquivo Histórico Militar: http://www.exercito.pt/SITES/AHM/Paginas/default.aspx; Arquivo Geral do Exército: http://www.exercito.pt/sites/ArqGEx/Paginas/default.aspx.

Table 2: Estimation of the height trend 1720s-1910s

	(1)	(2)	(3)	(4)
Recruitment regime	1776-1807	1763-75 & 1776-1807	1820-1856	1857-1932
Birth decades				
1720		0.19		
		(0.863)		
1730		1.01		
		(0.158)		
1740	1.53**	1.47**		
	(0.017)	(0.023)		
1750	1.44**	1.21*		
	(0.038)	(0.083)		
1760	1.58**	1.54**		
	(0.017)	(0.021)		
1770	0.88	0.75		
	(0.167)	(0.241)		
1800			1.10**	
			(0.030)	
1810			0.68	
			(0.207)	
1820			0.22	
			(0.701)	
1830			7.85***	-0.62***
			(0.000)	(0.000)
1840				-0.83***
				(0.000)
1850				-0.35**
				(0.020)
1860				-0.58***
				(0.000)
1870				-0.22
				(0.216)
1880				-0.72***
				(0.000)
1890				-0.58***
				(0.001)

1900				-1.10***
				(0.000)
age18	-2.07***	-2.08***	-1.62***	-1.12***
	(0.000)	(0.000)	(0.000)	(0.003)
age19	-0.62	-0.63	-0.98**	-0.99***
	(0.235)	(0.232)	(0.017)	(0.000)
age20	-0.67	-0.71	0.27	-0.84***
	(0.150)	(0.133)	(0.500)	(0.000)
age21	0.09	0.05	0.64	-0.72***
	(0.890)	(0.939)	(0.199)	(0.000)
age22	-0.33	-0.36	-0.37	-0.47**
	(0.526)	(0.488)	(0.418)	(0.026)
age23	-0.11	-0.17	0.47	0.02
	(0.846)	(0.765)	(0.367)	(0.944)
age45-50	-1.95	-2.13		
	(0.425)	(0.395)		
Grenadier	7.75***	7.22***		
	(0.000)	(0.000)		
Officer			7.85***	
			(0.000)	
Semi-skilled	2.24***	1.45**	2.70***	-0.23*
	(0.001)	(0.018)	(0.000)	(0.063)
Skilled	-0.57	-0.40	-0.72*	0.56***
	(0.137)	(0.256)	(0.071)	(0.000)
Semi-professional	0.97	1.28	-1.46	0.93***
	(0.591)	(0.462)	(0.428)	(0.000)
Professional	0.81	0.93*	1.25*	1.80***
	(0.119)	(0.062)	(0.080)	(0.000)
Minho	-0.26	-0.07	0.80	-0.43***
	(0.633)	(0.887)	(0.192)	(0.000)
Trás-os-montes	-3.79**	-3.74**	-4.06	-1.17***
	(0.030)	(0.019)	(0.118)	(0.006)
Beira	-0.43	-0.30	-0.41	-0.65***
	(0.348)	(0.494)	(0.312)	(0.000)
Alentejo	2.52***	2.52***	0.06	0.31**
	(0.000)	(0.000)	(0.866)	(0.049)

Algarve	3.69***	4.18***	-0.84	0.28
	(0.000)	(0.000)	(0.591)	(0.401)
Constant	163.84***	164.01***	163.73***	164.69***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,413	1,734	1,971	42,696
Wald chisq./R-sq.	666.64		710.26	438.590.01

Notes: */**/***significant at 10%/5%/1%. Constant for (1) and (2) refers to an infantry recruit born in the 1780s aged 24-50 born in the Estremadura region. Constant for (3) refers to the same, but born in the 1790s. Constant for (4) refers to the same, but born in the 1910s.

Table 3: Height by birth decades in European regions and real wages in Portugal

Region	Center	East	Portugal	South	Portugal
Variable	Height	Height	Height	Height	Real wage
1720	165.2	162.5	163.6		0.98
1730	166.0	163.3	164.3		0.92
1740	167.1	166.2	164.9		0.91
1750	167.3	165.1	165.0	166.0	0.81
1760	168.4	164.4	165.1	166.1	0.75
1770	166.8	163.0	164.4	164.7	0.71
1780	165.7	163.0	163.8	164.7	0.66
1790	165.0	161.1	163.3	166.0	0.54
1800	167.1	163.0	164.3	164.5	0.46
1810	167.1	162.3	164.4	164.1	0.49
1820	165.8	163.2	164.0	165.8	0.69
1830	165.9	163.6	163.6	162.6	0.55
1840	165.1	164.1	163.4	162.9	0.55
1850	165.7	163.6	163.7	162.7	0.46
1860	166.6	164.3	163.6	162.9	0.46
1870	167.5	166.1	164.1	163.1	0.60
1880	168.3	167.9	163.5	163.9	0.75
1890	168.8	168.3	163.7	164.4	0.62
1900	169.7	169.2	163.2	165.1	0.55
1910	170.6	169.1	164.3	165.7	0.54
1920	171.0	167.0	164.9	166.5	
1930	173.4	167.9	165.6	166.7	
1940	175.6	169.0	166.4	167.8	
1950	176.5	173.7	167.4	171.1	
1960	178.0	175.9	169.2	173.6	
1970	178.3	177.9	171.4	174.7	
1980	178.9	177.0	172.1	175.0	

Sources: Column 1-4: See Figure 4. Column 5: See Figure 6 and text.

Table 4: Regional shares by recruitment period

Recruitment period	Alentejo	Algarve	Beira	Estremadura	Minho	Trás-os-montes
1763-1774	5.8	2.1	8.8	75.5	6.3	1.6
1776-1807	5.9	1.3	12.9	69.8	9.5	0.6
1820-1856	41.0	1.2	30.2	17.8	9.4	0.4
1857-1932	4.6	0.9	7.3	69.9	16.8	0.6

Sources: see Table 1.

Table 5: Numeracy (ABCC indexes) for Portugal and European regions, 1720s-1910s

Birth decade	Center	East	Portugal	South
1720			76.4	
1750	90.0	66.0		89.0
1770	95.5			
1780	95.1			
1790	93.6			83.6
1800	95.8	82.0		85.4
1810	96.7	86.3		91.1
1820	96.7	75.4	79.3	90.9
1830	97.7	79.4	83.3	90.1
1840	96.8	84.9	85.7	90.6
1850	98.4	83.6	92.6	95.1
1860	98.1	86.2	90.8	95.0
1870	97.6	100.0	91.5	95.2
1880	98.3	100.0	94.8	96.2
1890	98.6	98.9	96.3	97.5
1900	99.8	99.8	96.3	97.4
1910	100.0	99.6	97.0	99.8

Sources: Crayen and Baten, “Global Trends”, A’Hearn, Baten and Crayen, “Quantifying”, own calculations.

Table 6: Livestock production in Portugal 1852-1925

	1852	1870	1906	1925
Total heads of livestock	5074717	5206920	6210862	7531617
Total heads standardized	1128428	1158794	1451672	1708356
Population estimated	3509134	4084446	5330751	5992116
Standardized livestock/population	0.32	0.28	0.27	0.29

Source: Justino, Formação, p. 119) – to standardize livestock, one head of cattle = 15 sheep or goats, or = 6 pigs

Table 7: Unbalanced panel Regressions of European regions 1720-1910 - dependent variable: mean height

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	LSDV	IV
Numeracy	14.43***		10.07***		16.50*		14.03***	15.33***
	(0.000)		(0.000)		(0.051)		(0.000)	(0.000)
Real Wages (log)			3.45***	2.76***	4.08***	1.86*	4.51***	1.87***
			(0.000)	(0.003)	(0.000)	(0.065)	(0.000)	(0.000)
Rel. Price of Protein (log)					-0.83	-1.85*	-0.29	-0.15
					(0.477)	(0.065)	(0.589)	(0.764)
Urbanization (log)					-0.50	-0.98**		
					(0.304)	(0.014)		
Poor Law					-1.04	-0.20		
					(0.444)	(0.824)		
Literacy		7.38***		3.41**		6.94***		
		(0.000)		(0.027)		(0.001)		
Time fixed effects	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
Region fixed effects							Incl.	
Constant	153.19***	160.20***	157.07***	162.74***	154.04***	169.69***	155.82***	153.02***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	77	45	75	44	34	32	71	67
Adjusted R-squared	0.59	0.73	0.82	0.80	0.85	0.89	0.78	0.75

Robust p-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations weighted by population share of regions. Regression (7) includes half-century dummies to control for time fixed effects, hence a Least-Square Dummy Variable model (LSDV) is estimated. Numeracy and literacy values are expressed between 0 and 1. ABCC is linearly interpolated where values were missing. Residuals of the preferred models (m5-m8) are stationary (Fisher-Test P-Value between 0.000 and 0.0278, H0 of a unit root can be rejected.). The four regions consist of: Portugal; Eastern Europe: Russia, Hungary; Center: Germany, Austria, Netherlands, UK, Ireland, France, Sweden; South: Italy, Spain.

2.10 Figures

Figure 1, Panel A: Histogram of polegadas, recruitment period 1763-1774

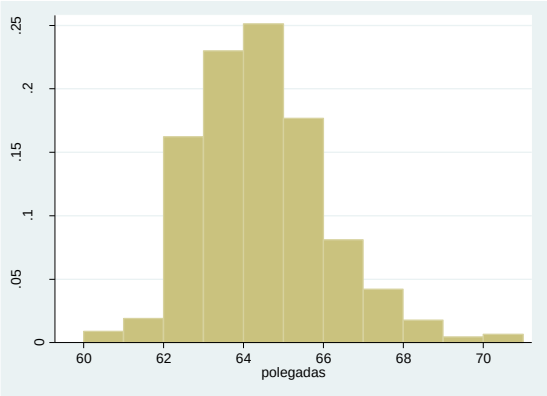


Figure 1, Panel B: Histogram of polegadas, recruitment period 1776-1807

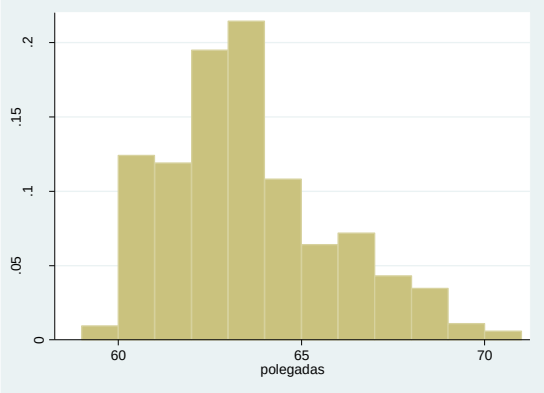
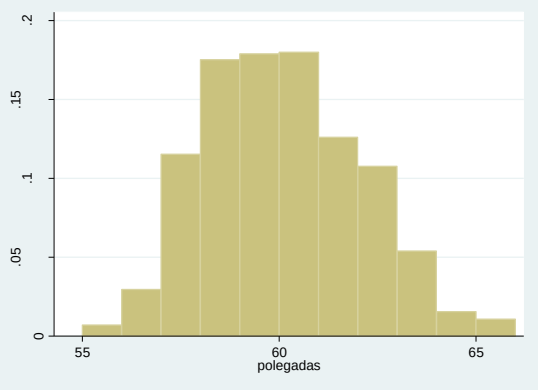
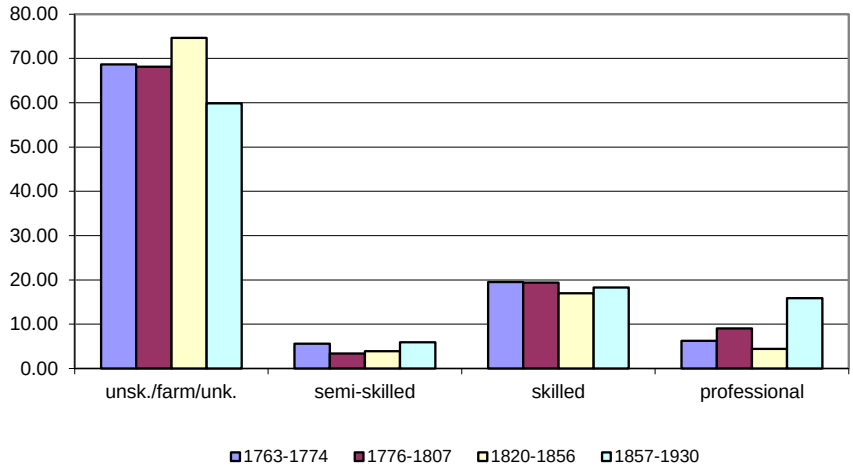


Figure 1, Panel C: histogram of polegadas, recruitment period 1820-1856



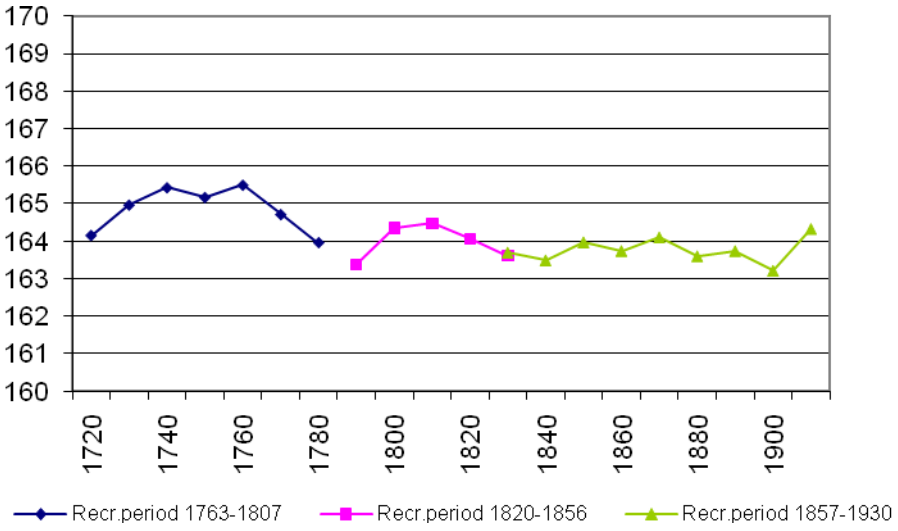
Sources: see Table 1.

Figure 2: Number of cases by occupational category and recruitment regime



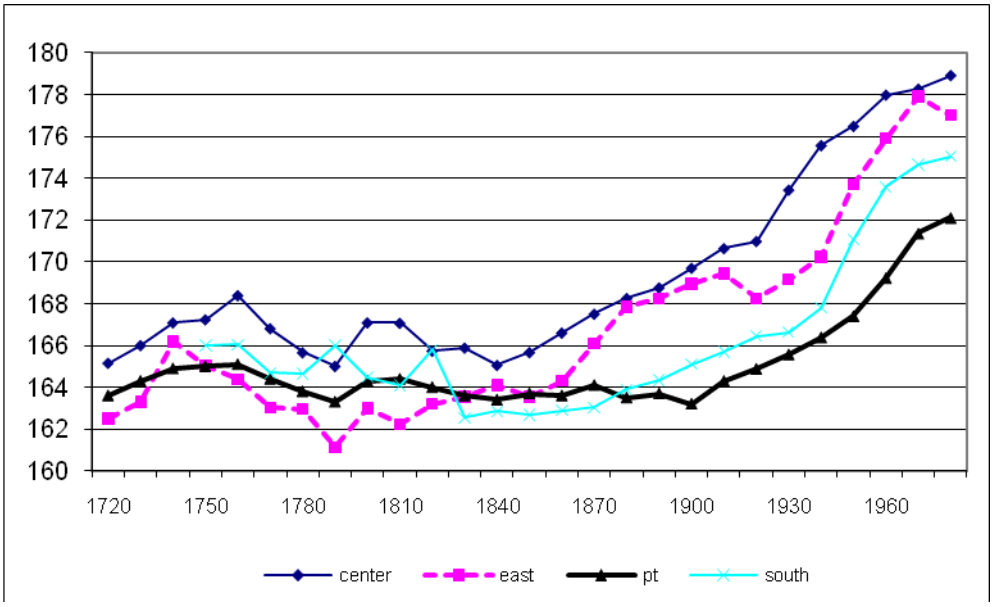
Sources: see Table 1

Figure 3: Height development of Portugal 1720-1910 by recruitment period, regionally adjusted



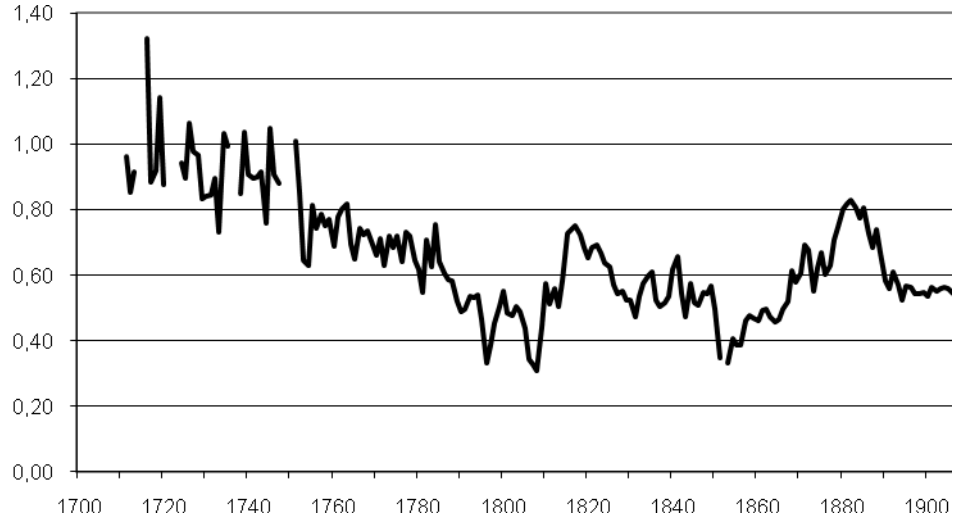
Sources: see Table 1

Figure 4: Portuguese versus European heights 1720-1980



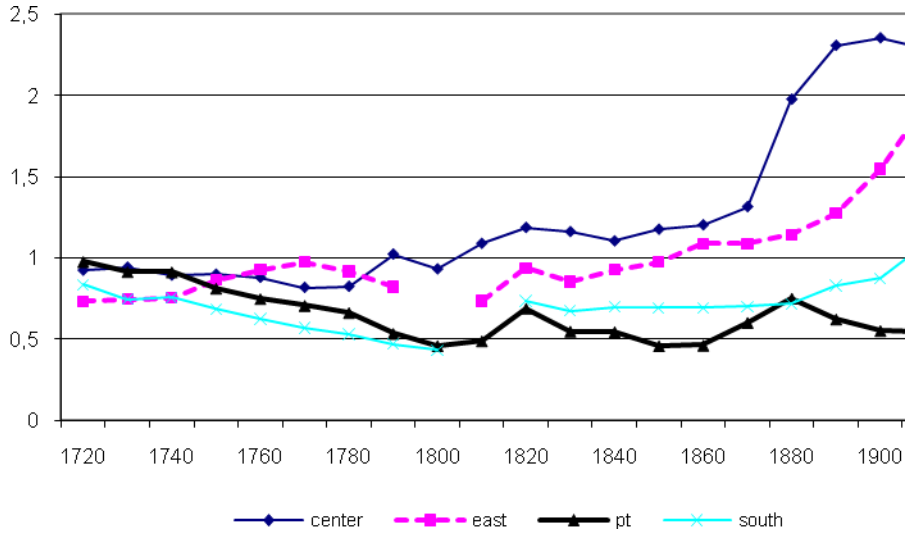
Notes: East: Russia, Hungary; Center: Germany, Austria, Netherlands, UK, Ireland, France, Sweden; South: Italy, Spain. European heights are taken from Baten and Blum, "Global Height Trends", Spijker, Perez and Cámara, "Cambios Generales", Carrion and Lázaro, "Urban height penalty", and a sample for the 18th century based on Cinnirella and Komlos, "European Heights", A'Hearn, "Anthropometric Evidence", Cámara "Long-Term Trends", Heintel, Sandberg and Steckel "Swedish Historical Heights", Cinnirella, "Optimists", Portuguese heights after 1910 are taken from Padez "Secular Trend", and Sobral, "Secular Changes". Eastern Europe in 1720/1730 refers to Russia only, Southern Europe in 1790 and 1800 is Italy only.

Figure 5: Lisbon Real Wages ("Welfare ratios") in Portugal 1720-1910.



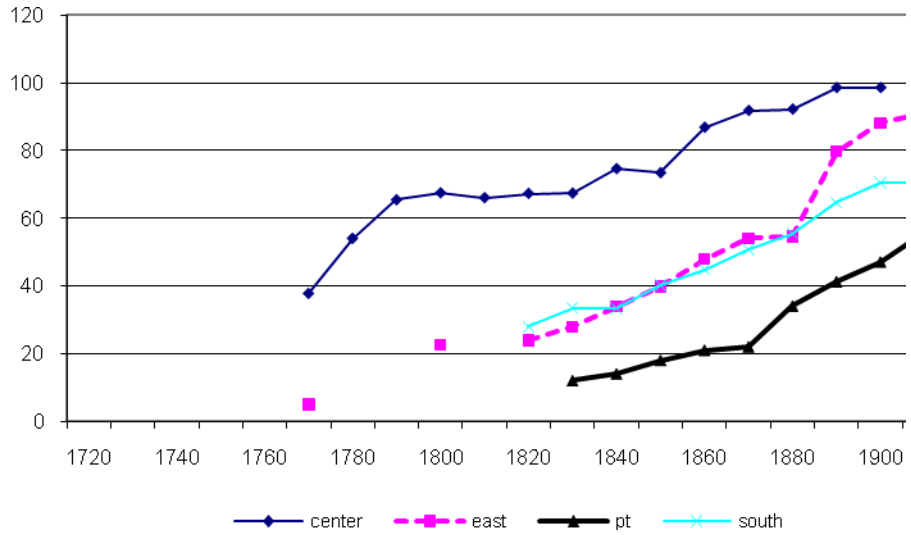
Sources: PWR Project, own estimations.

Figure 6: Welfare ratios (=real wages) in European cities



Sources: Allen, “Great Divergence”. “Center” is London, Paris, Strasbourg, Augsburg, Vienna (after 1880, only London and Paris). “South” is Madrid and Milan, “East” is Cracow and Warsaw, Lisbon represents Portugal.

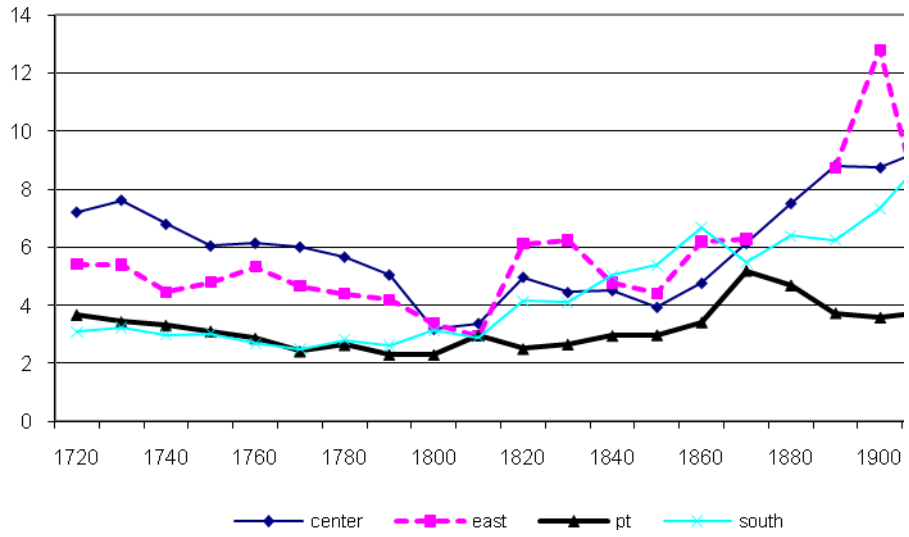
Figure 7: Literacy in Europe late 18th to early 20th century, by birth decades



Notes: East: Russia, Poland, Hungary; Center: Germany, Austria, Netherlands, UK, Ireland, France, Sweden; South: Italy, Spain

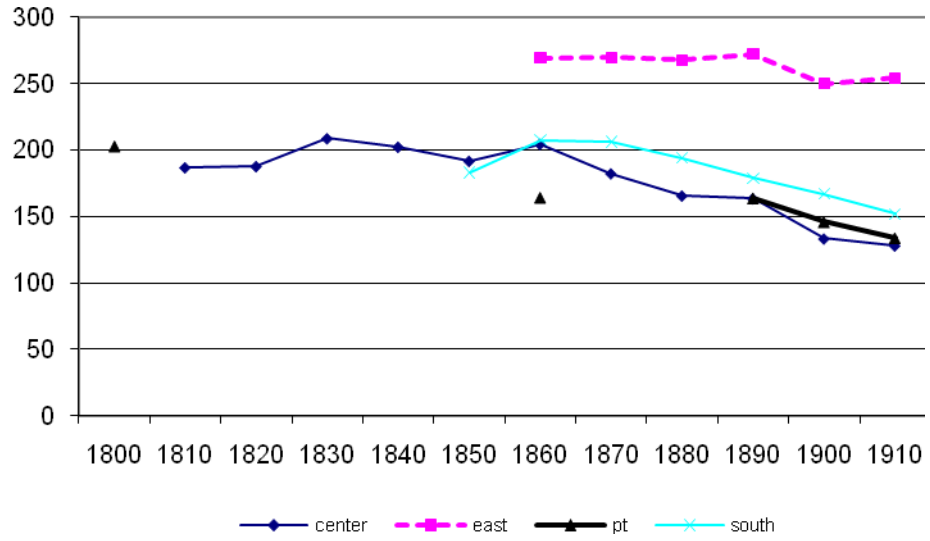
Sources: Tortella, "Patterns", for Portugal and Spain, Carreras and Tafunell, "Historia Económica" and Flora, "State", for Spain, Graff "Legacies" for it, Czech and Slovak lands: Flora, "State", Russia: Mironov "Literacy" and Russian Census 1897, Hungary: Toth, "Literacy" and Flora, "State", Poland: Flora, "State", Russian Census 1897, UN Demographic Yearbook 1963, austria: Graff, "Legacies", Germany: Hofmeister et al.,"Elementary Educaton", Flora, "State", UK: Flora, "State", Ireland: Flora , "State", France: Gillis, "Literacy", Sweden: Johanson, "History", Netherlands: Flora, "State".

Figure 8: Relative Prices of Meat and Grain for four European Cities



Notes: Kgs of wheat /rye that can be purchased for the price equivalent of one kg of meat. Southern Europe is represented by a Barcelona series until the first half of the 19th century, then Madrid (source: Perez Picazo, "Oligarquía urbana", Nicolau Nos and Pujol Andreu, "Consumo", Martinez Vara, "Coste de la Vida", Feliu, "Precios y Salarios", Nogués, "Análises", Garrabou and Tello, "Preus del blat", with special thanks to Professor Ramon Ramon, University of Barcelona), Amsterdam represents Center Europe (Van Zanden, de Vries, Smits, van Riel http://www.iisg.nl/hpw/brenv.php) and an average of Krakow and Danzig stands for the European east (Jacks, "Commodity Market", "Market Integration").

Figure 9: Infant Mortality Rates



Notes: Southern Europe (Italy, Spain), east (Russia, Poland), center (Austria, France, Germany, UK, Ireland, Sweden). Source Baten and Blum, "Global Height Trends", for Portugal: Rodrigues, "Evolução" for 1800, 1860s evidence is taken from Boletim do Minitério dos Negócios Eclesiásticos e da Justiça, no 7, pp. 124-129, Costa Leite, "Crescimento Económico", for 1900, 1910 is taken from census data.

2.11 Appendix A: In which units was height measured?

The official conversion rate of the Portuguese pre-metric inch (*polegada*) was defined in 1852 as 2.75 cm, a relatively large unit by international standards (see Da Silva Lopes 1849: 80). Having estimated height trends by recruitment period, it turns out that plausible results using this standard are obtained for the recruitment period of 1820-56, but not for the recruitment periods 1763-74 and 1776-1807 (Figure A.1). With the Portuguese official measure of 2.75 cm per inch, the mean height levels of recruits born between 1763 and 1807 would have been almost 10 cm higher than those measured between 1820 and 1856. Clearly, since the recruitment systems were different, it is possible that they might have led to a stronger selectivity in the eighteenth century. This would hardly account however for the measurement differential of some 10 cm detected above.

There are many examples from other countries in which the official measure was not used, but a different measure adopted from other countries instead.[76] After considering the historical context, we compared various possible measures with the level of the 1780s. Candidates could be, for example, the inch equivalents used in the large intercontinental Empires of the time: the British, French and Dutch Empires. The 'Rhenish inch' used in the Dutch Empire was 2.6154 cm, and the British inch was 2.54 cm. The French (Paris) inch was nearly as large as the Portuguese one (2.707 cm). Also the measure used in the Portuguese colony of Brazil (Rio de Janeiro inch: 2.5327 cm) might be a plausible candidate. Comparing these inch measures to the 1790s birth cohort of the 1820-56 recruitment period, we found that by far the most plausible was the Rhenish foot. The height levels of the 1780s according to the recruitment regime of 1776-1807 were quite similar to the heights of the 1790s birth cohort of the 1820-1856 recruitment system. Other measures in contrast would result in a strong positive or negative divergence. Is it plausible that a Rhenish (=Dutch, = North German) measure was used in Portugal? The Netherlands were an important trading partner, and often measures were exported to countries which officially had also own measures (Russia, for example). Most importantly, the restructuring of the Portuguese army in the 1760s, under Count Schaumburg-Lippe, a North-German, bureaucratic and measurement-enthusiast, could have inspired the adoption of the Rhenish measure.[77] This was kept for a few decades, until the national standardization of measures dur-

[76] Andalucia, for example, used a French measure during the early 19th century (Hueso 2006). In Brazilian prisons, Frank (2006) found a yardstick that yielded 2.73cm instead of the official 2.75cm. In Argentina, different "varas" were used in even in the same district (Silva 1997).

[77] For an account of these reforms, which were sweeping, see Bebiano (2007).

ing the early 19th century implied a switch to the official Portuguese measure. We find some evidence in the literature that the Count of Schaumburg-Lippe had, in fact, manipulated the measure. In 1763, we find an instruction on the minimum height requirement applied during that period, which refers to 62 "*Polegadas alargadas*", which means "extended", "broadened" or "widened" inches. As applying an inch that equals even more than 2.75 cm would not be plausible, in this case we conclude that the "extension" referred to the increased number of inches for a given stature. In other words, this measure yielded more inches than in case of the normal inch applied and was therefore '*alargada'*. Hence, we conclude that the Rhenish measure was used and apply this in our analysis.

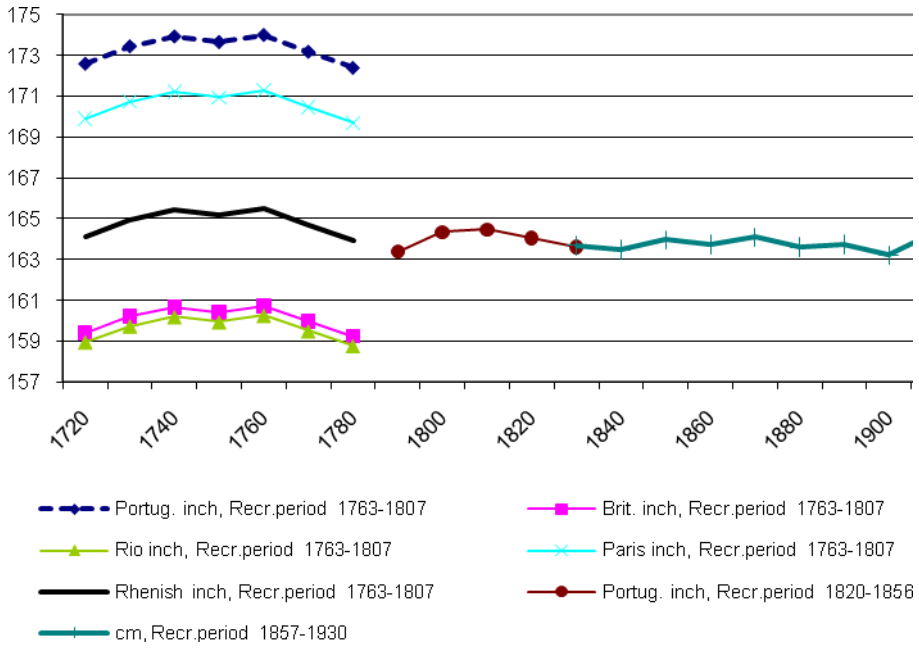
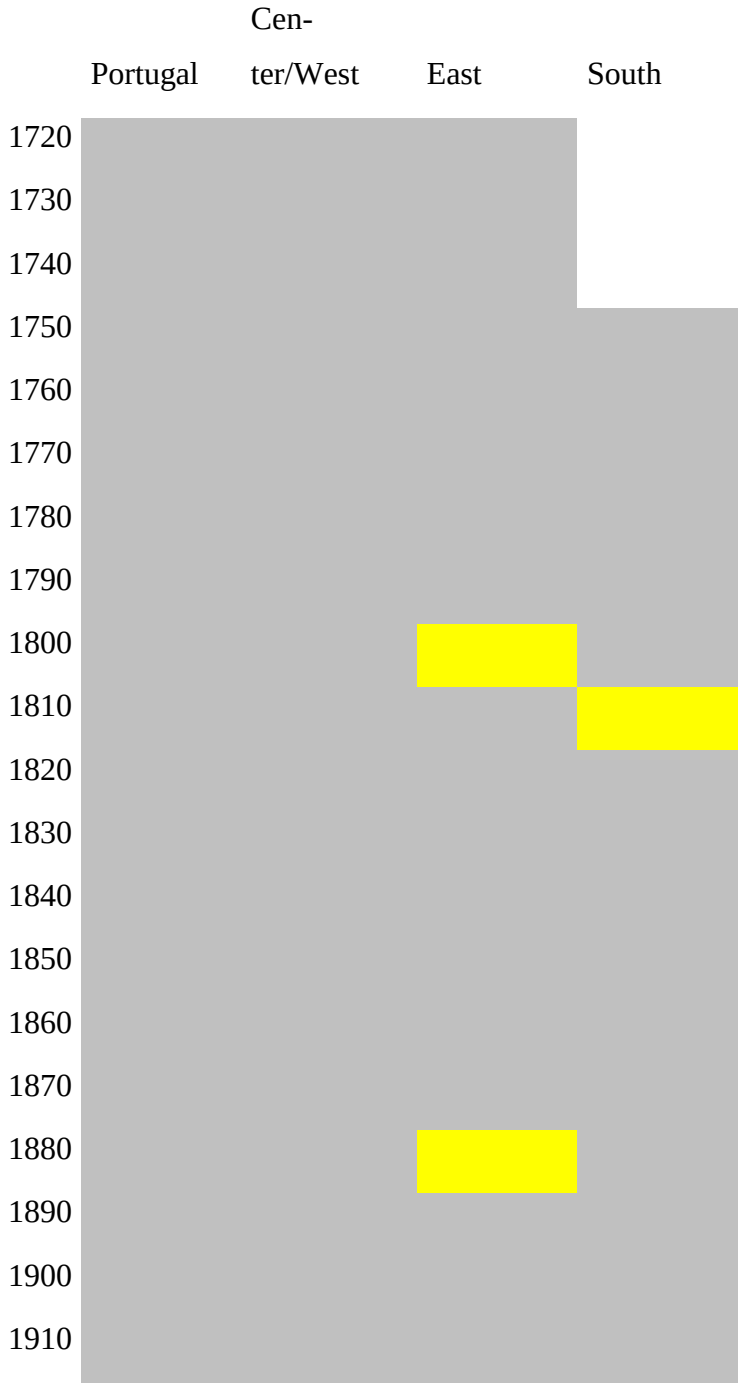


Figure A.1: Height development of Portugal 1720-1910, alternative measures for the pre-1770 period

2.12 Appendix B: Which countries and birth decade are in the various regression models?

In the table below we illustrate graphically which birth decades of which European regions were included in the regressions. For Southern Europe, three birth decades of height values are missing (1720s-1740s), hence there are 77 cases in column 1 of Table 7. Two additional missing values of welfare ratios (East 1800s, South 1810s) reduce this further in column 3, whereas the decrease of N after including literacy (only available for the 19th century) leads to a much greater reduction in N. The relative price is missing for the 1880s in the East.



2.13 Appendix C: A Cluster Analysis of European height regions

Although our distinction between three different European regions into South, East, and Center (/West) might seem a natural grouping, we were curious as to whether this might be supported by the available quantitative evidence on heights. We separated out Portugal from the beginning because this country is the main focus of our study, and we wanted therefore to keep it separate. In Figure A.3 we show a cluster analysis of the height levels of the 19th century for the other countries for which we have sufficient data. We could identify three groups which were most reasonably homogenous:

- Austria, Germany, Ireland, Netherlands, UK and Sweden
- Spain and Italy
- Hungary, Russia and – unexpectedly – France

Given that all these regions, with the exception of France, refer geographically to the South, East and Center of Europe, respectively, we moved France to the Center/West group, which can be justified by its other economic, cultural and geographic characteristics. Apart from this outlier, the quantitative evidence supports our classification into European regions.

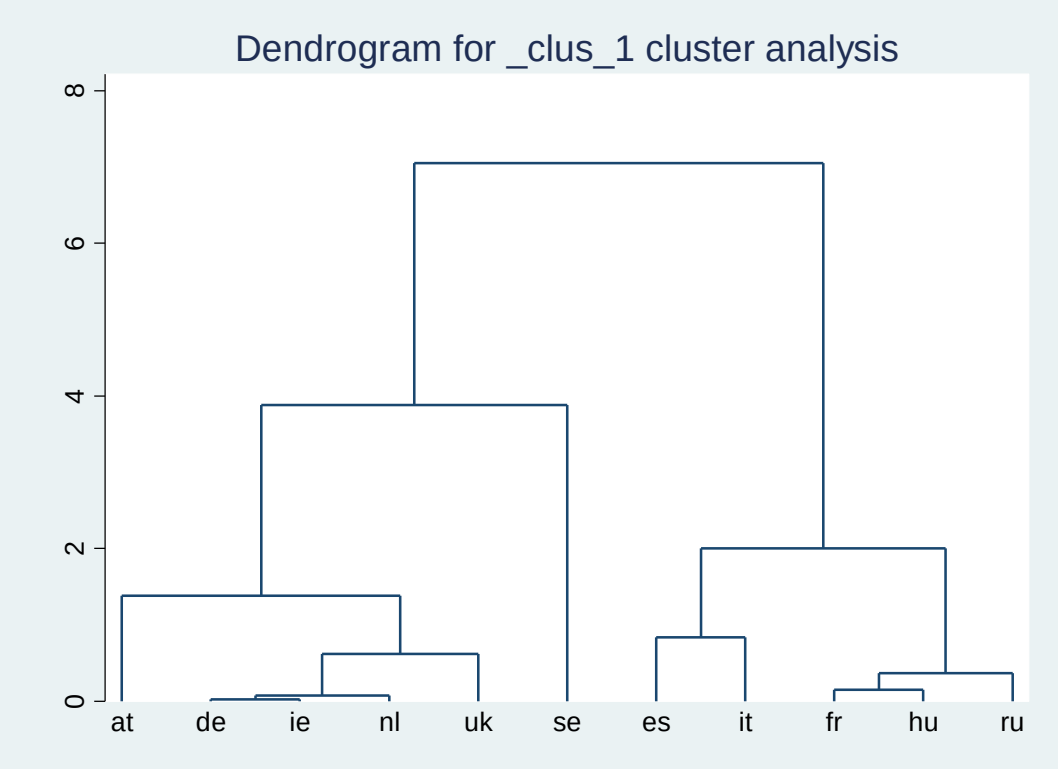


Figure A.3: A cluster analysis of 19th C European heights

2.14 Appendix D: A figure comparing the estimates of the recruitment periods of 1763-74 and 1776-1807, compared to the recruitment period of 1763-74 only

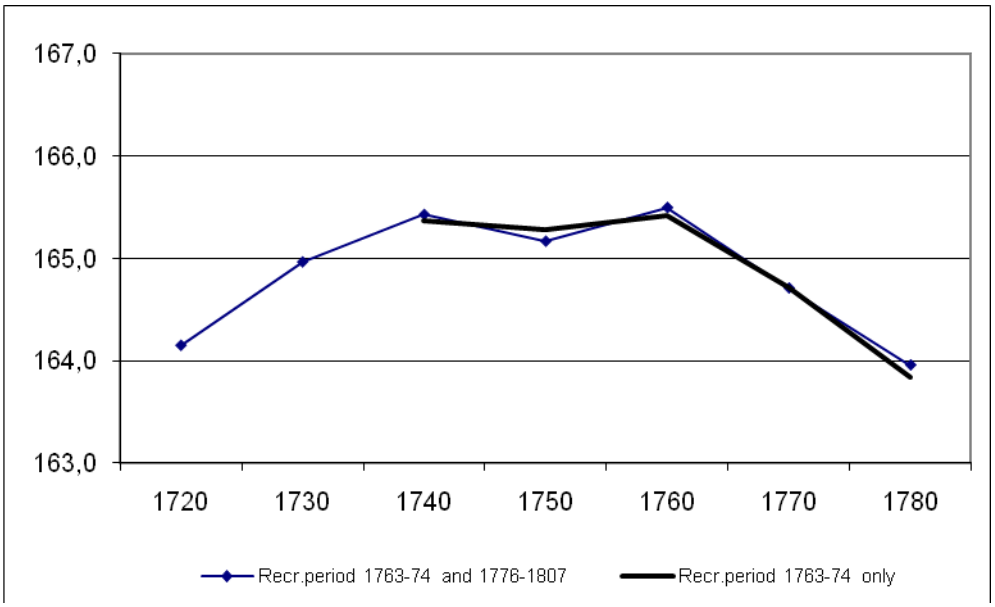


Figure A.3: Estimates of recruitment period of 1763-74 and 1776-1807, compared to recruitment period of 1763-74 only

2.15 Appendix E: Overlapping birth decades

Recruitment regimes compared	1763-74 versus 1776-1807	1776-1807 versus 1820-1856	1820-1856 versus 1857-1932
Birth decade studied	1740s	1780s	1830s
Regional dummies	Yes	Yes	Yes
Occup. dummies	Yes	Yes	Yes
Grenadier/Officer dummies	Yes	Yes	Yes
Dummy period 1763-1774	0.76		
	(0.367)		
Dummy period 1776-1807		0.26	
		(0.851)	
Dummy period 1820-1856			0.83
			(0.275)
Constant	165.56***	165.97***	164.58***
	(0.000)	(0.000)	(0.000)
N	151	60	878

Table A.4: Were the recruits born in the same birth decade, but measured under different recruitment regimes, significantly different in height?

Note: only adult recruits included (ages 22-45) because including the younger recruits would have yielded unreliable coefficients due to the small sample sizes for the individual ages. Among the birth decade of the 1740s (column 1), 121 recruits were measured under the 1763-1774 recruitment regime, and 30 under the 1776-1807 recruitment regime. The respective figures for column 2, 1776-1807 and 1820-56 (birth decade 1780s) are 38 and 22, for column 3 105 and 773. The 1780s measured under recruitment regime 1776-1807 in the general trend estimate could be included, because there were sufficient 18-21 year-olds to exceed the N=50 minimum.

3. CONVERGENCE AND DIVERGENCE OF NUMERACY: THE DEVELOPMENT OF AGE HEAPING IN LATIN AMERICA, 17TH TO 20TH CENTURIES DATA

3.1 Abstract

This study makes the first systematic attempt to trace the long-term development of Latin American numeracy: a phenomenon of great interest to economic historians in that it serves as an accurate gauge of human capital development. In order to approximate basic numeracy we use age-heaping techniques. We find that Latin America was on a path of convergence with Western Europe during the early 18th century. During the early 19th century, not only did numeracy development stagnate in some Latin American countries but differences among some of them actually increased. While numeracy rates in Argentina, Uruguay, and to a lesser extent Brazil, underwent, along with Europe, a significant increase in the late 19th century, they declined in Mexico, Ecuador, and Colombia. By performing a regression analysis, we find, even when we control for investment in education, that mass immigration contributed to human capital formation.

This chapter is based on a paper written jointly with Kerstin Manzel (Statistisches Bundesamt) and Joerg Baten (University of Tuebingen). It is accepted by the *Economic History Review*. The concept of the paper was developed jointly, analyzes and writing was equally shared.

3.2 Introduction

When Pedro de Valdivia, a Spanish conqueror and the founder of Santiago de Chile, came to the New World in 1536, 89 of the 150 Spaniards who accompanied him could not sign their names, and only one of them had received any formal education.[78] The educational level of the indigenous population of Latin America was at least as low as that of these Spaniards. The Aztecs and Mayas had some schools, but attendance was restricted to the sons of their leaders, or *caciques.* How did education develop during and after the colonial era, once Latin America had gained independence from Spain? This study will trace long-term trends of human capital from the early modern period to the 20th century and address the issue of the timing of improvement of numeracy in the New World. Was there already an international convergence in numeracy in the 18th century? Was the mass immigration of Europeans in the late 19th century a driving force in improving human capital in Latin America?

Most studies of the development of human capital restrict their view to a single country or a short time span for two reasons: information on human capital measures tends to be scarce (this is the case for most Latin American countries prior to the second half of the 20th century) and those measures that are far apart in time or space do not permit comparison. We begin our comparative analysis of human capital development in the early 20th century and then move back in time to the 18th, and in some cases the 17th, century. One important component of human capital is numeracy, i.e., the "ability to count, keep records of these counts, and make rational calculations".[79] We employ the age-heaping technique, which permits one to calculate the proportion of a given population able to provide to census takers their exact age, rather then providing a rounded age**.** This indicator of basic numeracy is a precondition for developing more advanced skills, including literacy. The age-heaping approach thus captures a basic form of human capital that is useful in this context, primarily because it is not linked to an individual's mother tongue. Moreover, this proxy is constructed coherently over time and space[80], thereby permitting the comparison of numeracy values from various sources and countries. Having been introduced into modern economic historiography by Mokyr[81], the concept of age heaping has recently led researchers

[78] Austin, *Education*, p. 1.
[79] Emigh, 'Numeracy', p. 653.
[80] Crayen and Baten, 'Numeracy', p. 89.
[81] Mokyr, *History*, p. 245.

to cultivate a new field of research.[82] The literature shows that numeracy is closely correlated with other human capital indicators, such as literacy and schooling. (A summary of the methodology can be found in Appendix C)

Latin America offers interesting sources that permit historians to trace numeracy over the long term. The colonial powers carried out population counts and regularly collected data on their colonies' populations. A great number of these sources have survived and can be used to estimate numeracy trends, despite biases, which need to be factored into any conclusions. We pose three fundamental questions: How did numeracy develop in Latin America from the 17th century on? Which Latin American countries led the field, at any given time, when it came to education? At what point in time may we speak, in reference to numeracy, of a convergence between Latin America and Western Europe?

Our findings reveal that during the 18th century many Latin American countries made such rapid numeracy progress that their levels converged with those of Western European countries, but that by the early 19th century this progress had begun to level off. Considerable differences emerged in the early 20th century, with the Southern Cone countries (chiefly Argentina, Chile, and Uruguay) at the upper end of the scale and Colombia, Ecuador and Mexico at the other. By performing a regression analysis, we find, even when we control for investment in education, that mass immigration contributed to human capital formation.

The remainder of the paper is structured as follows. Section 2 reviews the literature on education in Latin America. Section 3 presents the data sources we used to construct a new database of numeracy development in Latin America and discusses the representativeness of our samples. Section 4 shows the estimates for Argentina, Brazil, Colombia, Ecuador, Mexico, Peru, and Uruguay. Section 5 compares the estimated numeracy trends of these Latin American countries with those of the U.S. and Europe. Finally, Section 6 draws conclusions.

3.3 Literature review: Latin American human capital development in the very long run

Long before the arrival of the Spanish conquistadores three major cultures had devised systems by which to perform complex numerical calculations. The Mayas used a vigesimal

[82] A'Hearn, Baten, and Crayen, 'Quantifying quantitative literacy; Baten, Crayen, and Manzel, 'Zahlendisziplin'; de Moor and van Zanden, 'Leeftijdstapelen'; Clark, *Alms*; Crayen and Baten 'Numeracy'; Manzel

-- that is, based on 20 -- number system, represented by bars and dots. Their time intervals were the *tun* (360 days), the *winal* (20 days) and the *k'in* (one day).[83] The Aztecs combined simple numbers to signify larger ones; for instance, 399 was represented by (15+4) x 20+15+4.[84] The Incas devised *quipo*, a calculation technique, based on knotted strings that permitted them to develop a sophisticated administrative system including population counts.[85]

During the colonial period, schooling was seen as a method to "civilize" the native elites, undermine indigenous customs, and spread the Catholic religion. Education was in the hands of various religious orders that had arrived in Latin America during the 16th century: primarily the Jesuits, but also some Franciscans and Dominicans.[86] The Jesuits established missions, primarily among the indigenous populations, and often learned their languages in order to facilitate communication and thus conversion. When, in 1767, the Spanish King, Charles III, expelled the Jesuits from the entire continent, they left tens of thousands of indigenous people unprotected in their missions, many of which decayed, but some of which were taken over by Franciscans or the Dominicans.[87] Schools were scarce in colonial Latin America, especially before the 19th century. Moreover, attendance was restricted to the sons of the European elite and of the *caciques*, and classes were conducted almost exclusively in Spanish.[88] Apart from what scattered Jesuit missions provided before they were expulsed, there was virtually no formal education in Brazil in the 18th century.[89] What is more, the few schools that survived thereafter did not create much in the way of human capital, as the teachers were underpaid and the schools understaffed.[90]

Regional disparities tended to be considerable. For example, in post-independence Mexico the funding and administration of schools was often handed over to local municipalities, a policy that worsened regional disparities, with prosperous cities benefiting at the expense of rural areas.[91] A similar situation prevailed in Brazil, where the educational system was decentralized shortly after independence, resulting in a decline in the quality of

and Baten, 'Gender equality'; Cinnirella, 'British nutritional status'; Mironov, 'Novaya istoricheskaya demografia'; O'Grada, 'Jewish demography'.

[83] Closs, 'History of mathematics', p. 143.

[84] Conant, *Numeracy*.

[85] Julien, *Inka*.

[86] Leininger Pycior, *Education*; The Jesuits, for example, arrived in Peru and Mexico in 1568 and 1572 (Merino and Newsom, 'Jesuit missions') and the Franciscans in Mexico in 1524 (Vaughan, 'Mexican education').

[87] Merino and Newsom, 'Jesuit missions'.

[88] Bakewell, *History*, p. 90.

[89] Baer, *Brazilian economy*.

[90] Ibid.

[91] Vaughan, 'Mexican education'.

elementary education in some regions.[92] Compulsory primary education was theoretically introduced over the late 19th and early 20th centuries throughout Latin America, although not reinforced consistently. Generally, schools were poorly endowed.[93] One reason often cited to explain this decline in the quality of the educational system is the reluctance of the ruling elite to finance public schools; keeping the laborers and peasants uneducated reinforced their hold on power.[94] Lindert has called primary public education "the kind of education that involves the greatest shift of resources from upper income groups to the poor."[95] He discusses a number of positive and negative factors that influenced the decision to introduce large-scale, tax-financed, universal primary schooling. Rural elites in countries like Argentina, Chile and Venezuela who benefited from independence objected modernization endeavours including the provision of rural schools for fear of losing their bases of power.[96] The point of view of a member of the landed elite was, 'Why should I pay taxes to provide for public schooling when it will only spread discontent among the poor, including the day laborers who work on my own estate and incite them to rebel?'.

Statistics on the development of education in the New World are scanty. Oxford University's Latin American economic-history database is limited to the second half of the 20th century. Earlier literacy estimates exist for selected countries[97], even though Javier Núñez criticizes the estimates for the period around 1900, arguing that there was a lack of comparability between different definitions of literacy due to the fact that censuses were taken in different years and were confined to particular sectors of the population.[98] Moreover, in some countries "literacy" was defined as "the ability to read," in others as "the ability to read and write." Núñez's solution is to combine population censuses with marriage registrations (since they require signatures) and with crime records (since they indicate whether or not the criminal was literate). This methodology allows him to re-estimate literacy rates for the turn of the 20th century for a number of Latin American countries. He finds that literacy rates in Brazil in the first half of the 20th century were low, and concludes that this was due to the fact that Brazil abolished compulsory primary education in 1891 and reintroduced it only in 1934.

[92] Martìnez-Fritscher, Musacchio and Viarengo, 'Brazilian education'.
[93] Meyer Loy, 'Colombian education'.
[94] Mariscal and Sokoloff, 'Comparative education'.
[95] Lindert, *Social spending*.
[96] Morse, 'Urbanization'.
[97] Newland, 'Elementary education'; Engerman, Haber, and Sokoloff, *New world economies*; Astorga, Bérges, and FitzGerald, 'Living standard'.
[98] Núñez, 'Literacy evolution'.

Benavot and Riddle[99] performed a wide-ranging assessment of late-19th- and early-20th-century school-enrollment rates. Their data show a moderate overall increase in Latin American primary-school enrollment rates, more pronounced in the urbanized countries such as Argentina, Uruguay, and Chile, where there was a relatively large proportion of European immigrants, and in the British colonies of the Caribbean as well. However, one has to bear in mind that school-enrolment rates measure input, not output, and that even today the quality of education in some Latin American countries is poor, as measured by high repetition rates and low share of pupils finishing school, in spite of rising enrolment rates.[100]

During the 20th century, major differences among Latin American countries persisted. Astorga, Bérges, and FitzGerald[101] estimate literacy rates among those above the age of 15 in the LA6 countries (i.e., Argentina, Brazil, Chile, Colombia, Mexico, and Venezuela) at 33% in 1900, 60% in 1950, and 89% in 2000. For the remaining Latin American countries literacy rates were 32% in 1920, 46% in 1950, and 82% in 2000.[102] Moreover, most of the improvement in literacy took place from 1900 to 1939 for the LA6, but among the others did not occur until 1940-1980.

Núñez estimated that educational development in 19th-century Latin America trailed that of Italy and Spain by three or four decades, not on account of a lack of financial resources and resistance on the part of the elite but rather on account of the Independence Wars.[103] During the early 19th century, the newly independent countries suffered from political instability along with a drop off in international trade**,** and capital flight.[104]

European immigration was a decisive factor in the development of education in Latin America. Brazil subsidized immigration extensively in the late 19th century, a policy that Baer deems an effective alternative to investment in education.[105] Luis Bertola and José Antonio Ocampo[106] suggest a typology of Latin American countries: those in which indigenous and mestizo people compose a large proportion of the population, those with a relatively high proportion of African descendents and those with a relatively large proportion of European immigrants, such as Argentina, Chile, Uruguay, and Cuba. Such immi-

[99] Benavot and Riddle, 'Education expansion'.
[100] For Mexico, see Palafox, Prawda, and Velez 1994; for Latin America and the Caribbean, see: Wolff, Schiefelbein, and Valenzuela, 'Primary education'.
[101] Astorga, Bérges, and FitzGerald, 'Living standard'.
[102] Cuba is excluded.
[103] Núñez and Tortella, *La maldición divina*.
[104] Haber and Klein, *Brazilian independence*.
[105] Baer, *Brazilian economy*.
[106] Bertola and Ocampo, *Desarollo*, p.21.

grants not only had a positive impact on the promotion of education[107] but also constituted a source of teachers.[108] In Brazil, which had a far lower immigration rate than did Argentina, human capital development in those states where the rate was relatively high was significantly greater than where it was not.[109]

In sum, several (mostly qualitative) studies on Latin American educational development exist, but estimates of human capital measures taken prior to the late 19th century, which would permit long-range comparisons, are lacking. For instance, the study by Manzel and Baten of the differential between male and female age heaping among Latin American and Caribbean countries begins only with the post-1880s period.[110] This study fills the gap.

3.4 Data Sources

Beginning with that of the ambitious *Visita General* of the Viceroy Francisco de Toledo, in 1572, population counts in the New World were carried out regularly. The first were intended to provide a detailed overview of the territory and its inhabitants and did not contain systematic age statements. Most *visitas* focused on small regional units and were repeated every five to ten years. These were followed, in the next century, by the *empadronamientos de tributarios*, population counts to determine how much individuals owed in taxes. *Padrones de población* (population counts with a limited geographical or social scope) and partial censuses carried out during the 18th and 19th centuries covered larger regions and a larger portion of the population.[111] In 1776, the Council of the Indies started a series of systematic census records.[112] At the same time, Portugal, eager to learn more about the inhabitants of her colonies, carried out systematic enumerations.[113] By means of this information the colonial powers could better calculate their tax rates and learned about the total potential of their territorial possessions. For the post-colonial period, information on the first few decades is scarce; censuses of the republics are available, but most of them date back only to the late 19th and 20th centuries.

---

[107] This despite the fact they were not as well educated as immigrants to the United States, and in many cases had emigrated because they had not encountered success in their native countries.

[108] Thorp, *Economic history*, p.37.

[109] Stolz, Baten, and Botelho, 'Mass migration'.

[110] Manzel and Baten 'Gender and numeracy'.

[111] Mellafe, *Social history*, pp. 148-170.

[112] Platt, *Census records*, p. 8.

[113] The only groups excluded from the Brazilian counts were regular troops, ecclesiastics, and those Indians who, having successfully resisted attempts at cultural assimilation, were excluded from these censuses (Alden, 'Brazilian population').

In fact, Latin America is the only continent in the developing world for which such early population counts are widely available.[114] These records are particularly valuable in that they provide a considerable quantity of detailed information; records for the birth cohorts of the late 17th and 18th centuries include a given individual's age, gender, birthplace (in most cases), ethnicity or caste, marital status, occupation, and in some cases an itemized list of the family property. In many cases only aggregated data were available for the 19th and early 20th century.[115]

Ethnic background played a crucial role in the allocation of rights and duties in colonial society. In New Spain, for example, Spaniards, Mestizos, and Indios paid tributes, each at a different rate, to their *encomenderos* (from the verb "to entrust"), landowners who functioned as trustees, and whose tax obligations were, in turn, partly calculated according to the tributes paid by the indigenous people who lived and worked on their estates.[116] The data currently available permit an in-depth study of age heaping from the colonial era to the middle of the 20th century. Our evidence, spanning Argentina, Brazil, Colombia, Ecuador, Mexico, Peru, and Uruguay,[117] represents a large part of the Latin American continent, around 85% of today's population.[118]

An important question is whether our various sources are representative of a given society during the period under study. Population censuses were in theory universal, representing all social strata in the area under consideration. However, the native population viewed officials of the Spanish Crown as invaders, and the purposes of such a census (chiefly tax collection and military recruitment) led to fear and distrust on the part of the populace. Many demographers are therefore convinced that most of the census data are flawed by undercounting. Gootenberg[119] goes so far as to assume that all of Peru's population estimates are inaccurate. Some men aged 16 to 36 attempted to avoid the military

[114] Platt, *Census records*, p. 7.

[115] For the 19th century, the information is even more abundant; however, the period of the wars of independence features, not surprisingly, several gaps.

[116] McAlister, 'Social structure'; Mellafe, *Social history*; throughout the 18th century, however, the New Spain's socio-ethnic stratification was modified by the relaxation of certain traditional ethnic barriers (McAlister, 'Social structure'). For example, the newly created militia regiments provided opportunities for professional, and therefore material and social, advancement not only to whites but also, for the first time, to Creoles and free coloured people. In 18th-century Mexico, however, Indians, continued to be banned from military service, and mestizos were discouraged from enlisting (Vinson, 'Mexican militia').

[117] While national borders changed during the colonial and post-colonial periods, we refer to national borders as they exist today.

[118] Because Central America was sparsely populated and had no mineral resources, the Spanish Crown deemed it of little interest (Mabry, *Colonial Latin America*, 58), and it was therefore rarely the object of a census.

[119] Gootenberg, 'Population'.

draft.[120] Women declared themselves to be widows or spinsters in order to protect their husbands and sons.[121] In the *Censo de Revillagigedo* (1790-1794) in Mexico City, for example, widows are over-represented. While this and other draft-dodging techniques led to an undercount of the overall population, they do not bias our numeracy estimates as long as the data concerned are not strongly correlated with the men's educational status. The available evidence suggests that military-draft avoidance was widespread.[122] Did the opportunity costs of more educated persons lead to a downward bias in numeracy in the colonial period? Given the higher opportunity costs of richer and more educated persons, we would expect that our samples would be negatively biased where army service was perceived as an activity which should be avoided particularly by the upper social strata. On the other hand, human history is full of examples in which belonging to warrior castes or other types of military activity was not avoided in favor of economic and trading activity, especially in societies in which returns to land were partly appropriated in a quasi-feudal way by those who were powerful. Rich Latin American landowner's sons would enter the militia or the army normally as officers, because they would otherwise lose contact with the center of power which granted their landownership.

Given the fact that its financial resources were limited and that it had reason to doubt the fidelity of Latin American military forces, the Spanish colonial government's attitude toward them was cautious.[123] Until the 1760s, it often opted for small, part-time militias; but toward the end of the century it reversed its policy, and armies began to expand. One would expect this trend would trigger an increase in avoidance behavior on the part of the educated classes, which in turn would translate into a downward bias in our numeracy estimates. Can we observe this in the evidence? When we compare the 1744 military and the 1771 non-military Buenos Aires samples (Figure 2), we find no significant difference in the overlapping birth decade, that of 1710-1719. If anything, the early, military motivated sample had a slightly higher numeracy in this birth decade, but the difference is very small. Similarly, the numeracy trend in the 1818 military sample of Buenos Aires is situated at a level that would have been reached had the trend of the earlier samples continued in a linear way. To conclude, comparing the different samples of the Argentinean case does not support a strong downward bias for militarily motivated samples, not even during the late 18th century, when army service might have been extended over previ-

[120] Orozco y Berra, *Historia Mexico*; Alden, 'Brazilian population'; generally, census takers in English, Spanish, and Portuguese colonies were confronted with such problems as the dispersion of the population over a huge area and passive resistance" (Alden 'Brazilian population', p. 181).

[121] Arrom, *Mexican women*.

[122] With the possible exception of the very rich; Orozco y Berra, *Historia Mexico*, p. 72.

ous militia service. Evidently the sons of landowners and wealthy farmers, who constituted the majority of Argentina's socioeconomic elite, did not avoid the military draft in significant numbers.

Similarly, censuses for the purpose of tax collection prompted only moderate avoidance behavior: that is, the wealthy tended to conceal a portion of their wealth, instead of trying to avoid being registered at all. For example, it is hard to imagine that the authorities allowed Buenos Aires's upper class to avoid being listed among the city's inhabitants, whereas it is easy to imagine that by means of a bribe a rich man could get his tax base modified in his favor. Such avoidance behavior would therefore not lower the numeracy rate, since it did not cause a distortion in the population base of the tax-census data, and therefore does not falsify our numeracy estimates. Alden describes similar avoidance behavior in Mexico, Brazil and elsewhere in Latin America.[124]

A further strategy to assure the comparability of our sources is to run a panel regression of our numeracy estimates on a dummy variable that controls for the specific purpose of certain censuses: specifically, prison and military ones, with others, taken for other purposes, used as a reference category (Table 3). The 1700 Peru census, the Argentina ones of 1744 and 1818, and the 1846 and 1868 Uruguay ones were undertaken for military purposes. The 1871 Peru sample is taken from prison records. We control for country- and time-fixed effects. We find that neither the institutional context of prison nor military motivation was significantly different from the general-purpose census that served as reference. Consequently, even data flawed by an undercount can provide reliable information on the development of numeracy, as an element of education, in the colonial societies of Latin America. Some bias is unavoidable; we minimize its impact by comparing samples whenever possible.

An important issue is whether each member of a given household in the age range under study (23-62) reported his own age or whether it was the household head who reported the ages of all the household members. In the latter case, since the household head may not have been certain of their exact ages, age heaping, except in the case of the heads themselves, may have occurred. When we compared male household heads and non-heads registered in the 1744 Buenos Aires census, we found, in fact, no substantial differences, even though one might expect that the educational status of the heads would be higher.[125]

[123] We thank an anonymous referee for this hint.

[124] Alden, 'Brazilian population'.

[125] The difference in the ABCC index, not counting slaves, is only 48 vs. 46. We compared age heaping for the entire age range, 23-62, only, because the data set was too small to permit us to compare them by decade.

We conclude that the possibility that a household head was the source of data on other household members does not pose a serious problem. In addition, census takers sometimes noted that a certain person claimed to be 30 years of age but "but looked considerably older".[126] Such notes are strong evidence that census takers collected their data directly from each individual concerned, instead of accepting second-hand information provided by the household head, and that they resisted the temptation to adjust obviously erroneous statements of age.

Table 1 and Appendix A contain additional details on the sources that we used to construct a numeracy series for colonial and post-colonial Latin America. Some of our sources were restricted to a capital city, others to a region. The most challenging data sets derive from prison records; the concentration of prisoners on 23 and 24 years of age results in a skewed age distribution, and therefore in some cases in an upward bias for the youngest prisoners in the ABCC index. (For a discussion of the ABCC index, see Internet Appendix C.)

In order to assess the representativeness of our samples in detail, we compare the ethnic composition of our census samples with that of the entire population for all cases in which the literature provided the necessary information to do so (Table 2). It is noteworthy that classification by skin colour was abolished in most countries after independence; our comparisons according to skin colour are therefore limited to the colonial-era samples. The 1744 and 1771 Argentine samples provide an accurate representation of the population of Buenos Aires with the exception of its blacks and mulattos, slightly underrepresented in the first of the two.[127] As for the Mexico census samples, those of both Mexico City and Oaxaca in 1777 underrepresent the indigenous population, whereas that of Durango errs on the other side. The 1700 Lima sample overrepresents Spaniards, while the proportion of Spaniards in the prison sample for 19th-century Lima is smaller than that of the Lima population as a whole.[128] While we have no information on Brazil's ethnic composition prior to the 1830s, according to the 1830 census slaves constituted 36% of the population in the districts in Minas Gerais, São Paulo, and Paraná, a figure that seems reasonable in light of

Fortunately, household heads were not as a rule older than other household members, so age plays no role here.

[126] Cook, *Numeración*, p. 34.

[127] During the 19th century, the percentage of blacks and mulattos in the Buenos Aires censuses plummeted, from approximately 25% in those of 1810, 1822, and 1838 to 1.8% in that of 1887 (Andrews, *Afro-Argentines*).

[128] Fisher's data (Fisher, *Peru*) refer to the late 18th century. Afterwards, Lima's population evolved, the proportion of mestizos to whites and Indians gradually increasing.

the fact that slaves constituted 38% in the 1838 Rio de Janeiro census.[129] Thus it is safe to conclude that, while not all census samples reflect the exact share of each ethnic group in colonial society, our data are free of any serious bias.

Another important issue is whether migrants should be included in the individual samples. Since the focus of this study is the long-term formation of human capital in Latin America during the colonial and postcolonial periods, when immigrants were an important part of the population, they are indeed included. This decision enables us to compare individual and aggregated samples, since immigrants were always included in the latter. However, some shifts in the numeracy graphs during periods of mass immigration are to be expected, since the numeracy rate was generally higher among immigrants to Latin America than it was among the native population.[130]

Can we also compare age heaping and literacy in Latin America to double-check the methodology for the sources just described? To show explicitly that age heaping is a good indicator of educational status, we compare age-heaping and literacy data from samples representative of the 1869 and 1895 Argentinean censuses. These samples contain information on 38,776 and 43,897 inhabitants, respectively, from 23 to 62 years of age. In 1869, roughly 78% of the population were unable to read and write, while 21% declared themselves to be literate (1% gave no answer). By 1895 the literate share had had soared to 49%.[131] Figure 1 shows that age-heaping was far more extreme among illiterates than among literates; the spikes of their rounded ages are much higher, indicating that they were less willing or able to report their exact age. Literates in the 1840s birth decade had a ABCC rate of exact-age reporting of 81.5%, illiterates an index of only 60.5%.[132] The correlation on a provincial level between the share of illiterates and the Whipple index yields a highly significant coefficient, of 0.87. We can therefore safely conclude that in this context age heaping is a valuable, informative human capital indicator.

[129] Graden, 'Slave trade'; the proportion in the northeastern states, however, tended to be higher than elsewhere in Brazil; estimates for Salvador are around 42% (ibid.). The 1838 census of Rio de Janeiro only distinguishes between slaves and free people. The figure refers to slaves, one has to bear in mind that not blacks were slaves and that there was an important share of mulatoes.

[130] Stolz, Baten, and Botelho, 'Mass migration'.

[131] Only native-born Argentines are considered.

[132] These figures are comparable to those of Newland (Newland, 'Elementary education'), who calculates that in 1900 52% of Argentines over the age of 10 were literate.

3.5 Age-heaping trends from colonial times to the 20th century

In order to assess the long-term development of numeracy in Latin America, we begin with a review of Argentina, for which we possess the most comprehensive data. Afterwards, we will proceed to discussing the other country cases, for which the data set is still substantial, but has some gaps and selectivities.

3.5.1 Argentina

For Argentina, we provide the first long-term estimate of numeracy values, beginning in the 17th century. The values for Buenos Aires in the late 17th and early 18th centuries indicate a very low numeracy rate (Figure 2). Initially, less than 40% of the population of Buenos Aires reported an exact age. The 1744 and 1771 censuses of the capital are mutually coherent, the value of the youngest cohort of the first similar to that of the oldest cohort of the second. Overall, there is a remarkable, 30%, increase in numeracy between the earliest cohort and the mid-18th century. A sample from the 1818 Buenos Aires census covers the birth decades of the late 18th century. Although its main aim was to assess military capacity, it included females and those who were not household heads. The fact that the line is flat suggests stagnation during the late 18th century, but at a high level, of around 75-80%. For the Buenos Aires birth cohort of the 1810s the numeracy levels are about 70%, far below the one for the late 18th century and almost as low as the one for the 1750s. The birth cohorts of the period of the wars of independence (roughly 1808-1829) indicate an increase, but from a very low level, below that of the mid 19th century. We would characterise the increase between the lows of the 1810s through the 1840s as recovery because only the late 18th century levels were reached again.

For the 1810s, we also have the first nationwide estimate for Argentina: lower than in Buenos Aires (where in 1810 it was approximately 9%), suggesting that those who lived in the capital had greater access to the educational system and benefited from its economy, which offered more opportunities for skilled workers than were available elsewhere. After the 1840s, a continuous increase in numeracy took place, both in the capital and the rest of the country. The upward trend continued steadily until 1880, when, according to 1947 census data, age heaping ended. However, some age heaping occurs in the overlapping birth decade of the 1880s, as one can observe by examining the data for the youngest age group in the 1914 census. This gap of numeracy for the same birth cohort between two census years is typical for periods of mass immigration of more skilled migrants, relative to the native population. Although immigration rates declined after 1914, the influx remained

considerable; the proportion of numerates -- that is, of those who declared their exact age -- in the census of 1947 would have been smaller than it was if it had not been for this selective immigration.

In order to estimate Argentina's national numeracy trend, we carried out an OLS regression, controlling for the share of males in each census (between 0.47 and 0.77) and controlling for a census of the capital with a dummy variable (Table 4). The coefficient of the capital and the male share is statistically significant and positive. The resulting time dummies, based on an assumption that males constitute 50% of the population, indicate that there was a positive development that persisted over time, with a steep upward trend from 1680 to 1760 (Figure 3). However, the regression results also indicate that there may have been a temporary stagnation, or even a temporary deterioration, in the numeracy rate from the late 18th to the mid 19th centuries, even after controlling for gender and for the greater access to schools in the capital. During the second half of the 19th century, however, numeracy development in Argentina was clearly positive, thanks in large part not only to the influx of substantial numbers of numerate Western European immigrants but also to the relatively early introduction, in 1884, of compulsory primary schooling.[133] Although the educational level of these immigrants was generally lower than that of those who headed to the United States, it was superior to that of the native Argentines, and thus fuelled the rapid growth in human capital there.[134]

3.5.2 Brazil

Our earliest evidence regarding numeracy in Brazil dates from the birth decade of 1700 (Figure 4), our most recent was a census taken in 1950. The sample for the 1700-1740 birth cohorts, from a census of the São Paulo district of Sorocaba, does not contain data on ethnicity; that for the 1770-1800 birth cohorts, derived from 1830 census data collected in the states of São Paulo, Paraná, and Minas Gerais, contains data on men and women from both the slave and the free populations. In 1830 slaves constituted over 36% of Brazil's south-eastern population: a far larger percentage than that found in Argentina and Uruguay at the time. A'Hearn, Baten and Crayen showed that during early educational development, ABCC values of numeracy begin to increase first, before literacy values move upward.[135] In other words, ABCC values of around 50 percent often correspond with literacy values

[133] Cortés Conde, *El progreso*.
[134] Núñez and Tortella, *La maldición divina*, p. 371.
[135] A'Hearn, Baten, and Crayen, 'Quantifying', p. 789.

below 10 percent. Knowing one's exact age is a much more basic skill than reading and writing. This is also true for the slave population, which was largely illiterate. A'Hearn et al. studied U.S. African Americans who were born in slavery.[136] With data taken from the 1870 US census the authors show that numeracy values of native blacks is around 70%, while literacy only reaches a value of 35. If we apply this difference to the 1830 Brazil evidence, where some groups of slaves exhibit an ABCC of only 45, their literacy must have been below 10 percent, hence almost illiterate.[137] In all three states, slaves show significantly less numerical abilities than the free population.

How do our two early samples compare? Together they describe an overall upward development in numeracy throughout the 18th century. What exactly happened during the 30-year period between our two samples remains unclear, but their levels suggest that numeracy first increased and then stagnated or even decreased during the Napoleonic wars, which prompted the Portuguese court to flee from Lisbon to Rio de Janeiro, where it remained in power for over a decade. Brazil achieved independence quite peacefully, at least by Latin America's standards.[138] The 1890 census sample, which can be regarded as representative of the Brazilian population, indicates that numeracy levels rose modestly through the birth decade of the 1830s to that of the 1860s. The 1920 census sample is slightly positively biased as detailed age information is only reported for the regional capitals of the Brazilian districts. The census data of 1950 is again nationally representative. The trend of the 1830s to 1860s continues almost linearly to the trend of the 1890s to 1920s. The slightly upward biased urban sample in between confirms this development. Although there was a solid improvement, Brazil had not yet overcome age heaping by the 1920s, in contrast to Argentina and Uruguay. However the improvement in numeracy during the era of mass migration is considerable and as already in the Argentinean case, one can speculate about human capital enhancing effects from international migrants who came to Brazil.

[136] Ibid.

[137] This value refers to the 1770 birth cohort in Paraná. Slaves in the other birth cohorts were for the most part in the range of 35 to 60 ABCC points.

[138] To which degree are the early samples biased? Before the coffee boom and the large migration waves set in after the 1870s, the Southeast was actually not a region with higher human capital than the Brazilian average (except the urban Federal District of Rio de Janeiro), whereas the South was. We can consider the regional numeracy of the birth decade 1830s based on the 1890 census. In this decade, the ABCC indexes in Sao Paolo and Minas Gerais were some 3 percent below the national average, whereas the one on Paraná was some 3 percent above the national average. Assuming similar regional differences for the inhabitants in the 1830 population lists, the slight negative regional bias cannot be very large (around minus 1 percent). For the 1772 census, in which only Sao Paolo is represented, it might be in order of the 3 percent negative bias (i.e., the Brazilian national figure might be even higher).

3.5.3 Colombia

Evidence regarding Colombian numeracy is scarce. Calculations from our primary sources suggest that numeracy levels were relatively high during the early 18th century throughout the country, and especially so in the merchant city of *Cartagena de las Indias* (see Appendix B). The fact that ABCC indexes for a broad sample of provinces in the first half of the 19th century were all of the same level suggests that numeracy stagnated or improved only slightly between the early 18th and early 19th centuries. The 1928 census reveals a low numeracy rate for the birth decades of the 1880s and 1890s.

To what degree is each of the three early Colombian samples representative of the country? In Table 5 we report by region the number of observations for the 1777, 1870, and 1928 samples. These are sorted by the ABCC numeracy value in the 1930s birth decade, for which we possess data from all regions. Bogotá has the highest numeracy levels, with 98% of the population reporting an exact age, the centre and the Amazonas regions the lowest. Neither Bogotá nor the second-most numerate region, Eje Cafetero, is represented in these three early census samples, whereas the two least numerate regions are, suggesting that there may be some downward bias among them. The 1870 census represents a broad mix of regions. In 1777 and 1928, a higher share of regions fell into the lower half of the 1930s numeracy spectrum. Of course, the relative ranking probably changed between 1777 and 1930, but the regional composition may at least partly account for differences among the three samples. However, that there was stagnation between the 18th and late 19th centuries is quite likely, since the first and third samples were similar to one another in their regional composition.

Can this poor performance during the second half of the 19th century be explained by negative events that hampered efforts to improve Colombia's educational system? In fact, the many changes of government during this period and, more important, the civil war of 1876-1877 had a negative influence on educational levels. Ramirez and Salazar[139] find that funding of public education fluctuated with each change of regime. Moreover, during the civil war educational reforms were abandoned. Schools were transformed into hospitals, teachers were turned into soldiers, and governmental spending on education was suspended.

To summarize: Colombia's numeracy levels compared well with those of other Latin American countries at the beginning of the 18th century but stagnated during the 19th,

[139] Ramirez and Salazar, *Educaiòn en Colombia*.

resuming their upward course, toward convergence with those of its more advanced neighbors, only after 1900.

3.5.4 Ecuador

Our Ecuador data, derived from an 1871 census, provide complete coverage of the western provinces of Manabi, on the coast, and Azuay and Pichincha, inland: age, gender, occupation, and birthplace. There is no information on race or ethnicity in the tradition of all republican censuses. The sample comprises 71,545 observations. The birth decades, ranging as they do from 1800 to 1840, cover the critical period of independence. The 1800-1809 birth decade attains a numeracy level of only 54 ABCC points, much lower than the levels recorded elsewhere in Latin America (Appendix B). Between 1810 and 1830 there is a very slight improvement, and during the 1830s birth decade Ecuador's numeracy level rises sufficiently to converge with that of another Andean state, Colombia, but their levels remain the lowest in our panel. The numeracy level of the 1880s birth decade, represented by the 1950 census, is slightly below that of the 1840s birth decade, suggesting that levels stagnated throughout the 19th century. It is only after 1890 that Ecuador's level begins to rise, overtaking that of Mexico during the 1910-1919 birth decade.

3.5.5 Mexico

Our Mexican numeracy estimates begin in 1680 for the provinces of Hidalgo, Guanajuato, and Oaxaca, in central and southern Mexico (Figure 5).[140] The ABCC index suggests that only 40% of those born in these provinces in the 1680s were able to state their exact age. A number of sources indicate that by the mid 18th century basic numeracy had improved, with values varying between about 60 and 65%, and our sources for Coahuila (a northern province) during the late 18th century suggest that this trend continued. The census of Guadalajara (1821/22) in central Mexico reports considerably lower numeracy for the late 18th century but this was a regional phenomenon. Since northern Mexico's provinces were sparsely populated, but decently well-off in per capita terms, and therefore spent more on education than did those in the centre and the south,[141] it is possible that Guadalajara's lower numeracy level reflects this regional disparity. Because data for the birth decades from 1800 to 1860 are missing, we cannot prove that numeracy levels stagnated around the

[140] In order to have an adequate number of observations per birth decade, we were obliged to pool regions thus. For further details, see the Internet appendix B.

[141] With the exception of the federal district of Mexico City (Vaughan, 'Mexican education').

time of independence, but the fact that the 1870s birth decade's numeracy levels were low suggests that this was the case. For the birth decades from 1880 on (as represented by the 1930 and 1950 census samples), numeracy indexes remained quite low in Latin American comparison. To estimate the numeracy in Mexico we use a regression analysis similar to the one we used for Argentina, controlling for the share of males in each census and including a dummy variable for a census of the capital (Table 4, Columns 2 and 3). We also controlled for the indigenous percentage of the Mexican population, although it turned out to be insignificant. The trend in Mexico City, likewise, was not significantly different from that of the country as a whole, whereas the male share was significant for Mexico, indicating that in the 19th century gender differences in numeracy were greater in Mexico than in Argentina; results in line with those for the early 20th century.[142] Many of the time-dummy coefficients were insignificant (Table 4, for a graphical representation of time-dummy coefficients see Figure 3). Until 1750, Mexico's numeracy levels were higher than those of Argentina, but then they stagnated, whereas Argentina's soared.

3.5.6 Peru

The Spanish conquest of South America began with Peru in 1532. For at least the next 200 years thanks mostly to Bolivia's silver mines and Lima's proximity to the sea, the capital remained the continent's most important city. Lima's residents were mostly European, only a small portion of the indigenous population living within the city walls.[143] We have data for very early birth cohorts, those of the 1640s and 1650s, which are characterized by very low numeracy levels; those of the next two decades show some improvement. Age statements are of male household heads in Lima only and might therefore actually overestimate the general numeracy of the population. Because the data were limited to age statements made by male household heads, the estimates may indicate a higher level of numeracy than in fact existed among the general population.

Because we do not have age data for the birth decades of the 18th century, there is a 160-year gap, terminating in the 1820s with a sample of Lima prison inmates, many of whom were natives of the capital or the coastal region.[144] The numeracy estimates, especially those derived from data on the youngest prisoners, those born in the 1850s and 1860s, appear to be upwardly biased, when we compare them with the first birth cohort of

[142] Manzel and Baten, 'Gender and numeracy'.
[143] Mabry, *Colonial Latin America*, p. 59.
[144] Twrdek and Manzel, 'Peruvian living standards'.

the 1940 representative census of Lima. This 1940 census also allows us to estimate the numeracy gap between Lima and the rest of Peru: an unusually high difference of 17% for the 1880s birth decade. There are not sufficient census data from earlier decades to permit us to estimate Peruvian numeracy in a regression analysis; instead, we provide the available data in a table in the Internet Appendix B. For the same reason, we restrict our interpretation to the city of Lima.

3.5.7 Uruguay

Most of our Uruguay data derive from Montevideo. Numeracy in the capital in the early 18th century is at the same high, 60%, level as that for Buenos Aires: not surprising when one considers that Montevideo was founded in 1726 mainly by *Porteños* (inhabitants of Buenos Aires).[145] Two early-19th-century samples from the city's prison population display some random fluctuation, but their rates are not as low as those of rural Soriano and Maldonado (Appendix B). As we explained in the "Data Sources" section 3, such random fluctuation is associated with prison data. The fact that Europeans constituted as much as 60% of this population helps to explain the relatively high level of numeracy in those samples.

However, the evidence suggests that throughout the late 18th century numeracy trends were nearly flat. If we assume that prison samples from the mid 19th century are, despite some variability, representative of the general population, then the numeracy level of the latter was about 80 to 85%. For 20th century birth decades there is the nationwide census of 1963. Age heaping has disappeared; this means that between the mid 19th century and 1900 Uruguay's overall numeracy rate must have risen sharply. To summarize: numeracy in Montevideo is at a high level in the early 18th century, stagnates in the late 18th century, and then, along with the rate in the rest of the country, soars during the late 19th century, perhaps thanks to mass immigration. If we assume that the mid-19th-century numeracy levels in rural regions were lower than those in the capital, then it follows that this increase would have been from a lower starting point and therefore even steeper if we had been able to calculate earlier rates for the entire country as well as for Montevideo.

Because we have little data for rural Uruguay, and the other explanatory variables are not reported systematically we cannot apply a regression analysis in order to adjust our results to account for the capital effect and other variables. We therefore decided to study the available data on Montevideo and Soriano/Maldonado up until the mid 19th century and

display the trend for the whole country around 1900 graphically and in a table provided in Internet Appendix B.

3.6 Comparison of Latin American countries

We will now use a LOWESS regression, carrying out a weighted linear least-square fit of the data points, to estimate the general trends of average numeracy, thereby reducing short-term fluctuations in order to compare the the seven countries under study (Figure 6).[146] Data points near the point for which the response is being calculated are given greater weight than are the others.[147]

Argentina's numeracy levels are relatively low at the start of the period under study, but during the early 18th century they rise faster than do those of the other six countries, and by the 1750s reach those of Mexico, Uruguay (data for this period are mostly for Montevideo), and Colombia. After 1810 Mexico and Colombia began to lag behind the other five (but the Colombian lag may have been partly due to a regional bias). The numeracy levels of these two countries seem to have suffered the most from the wars of independence, the political instability of the early republican governments, capital flight, and the disruption of internal trade. The early-19th-century numeracy levels of Ecuador, for which we provide fairly reliable evidence, are extremely low until the end of the century.

In Mexico the conflicts between centralists and federalists (who will evolve into the conservative and the liberal parties) lasted until the 1860s, when President Benito Juárez, introduced a law that made primary education mandatory as well as free. During the *Porfiriato* (the presidency, spanning the years 1876-1911, of the war hero José de la Cruz Porfirio Díaz Mori), two decades later, primary schooling began to spread.[148] But especially the *Porfiriato* was also characterized by high inequality. Nevertheless, our trend for Mexico shows a small improvement in numeracy during these later decades of the 19th century, although we might want to interpret this as a convergence from an actually very low starting level.

The Brazilian region around São Paulo had favorable levels during the early 18th century. Around 1850, Brazil was on the fourth place, behind Lima, Uruguay, and Argentina, even if Brazil achieved independence relatively peacefully, and suffered less econom-

[145] After initially six families had arrived from Buenos Aires, they were followed by a second and third group of 13 and 25 families, respectively, who came from the Canary Islands. We thank an anonymous referee for his hint.
[146] LOWESS stands for "Locally Weighted Sum of Squares."
[147] Cleveland, 'Smoothing scatterplots'.
[148] Andrade de Herrera, 'Education in Mexico'.

ic dislocation than did most other Latin American countries during their wars of independence. From the mid 19th century on, Brazil remains in the middle, trailing these three but trailed by Mexico, Colombia, and Ecuador until they start to catch up, after 1900. The Mexican Revolution of 1910, however, causes the Central American country to fall back even behind Ecuador and Colombia and to drop into last place.

How did European immigration to Latin America affect numeracy trends there? A study by Klein describes immigrants that raised the levels in Argentina and Brazil, often constructing their own schools and charitable institutions.[149] In addition, their entrepreneurial activities had a significant socioeconomic impact on Argentina, Brazil, and Uruguay. In order to assess whether the impact of immigration on numeracy development in Latin America was systematic or not, we constructed a small panel using the numeracy time series presented here (Figure 8). When we regressed these numeracy values on schooling estimates[150] and a dummy variable which controlled for substantial immigration that surpassed a certain threshold, we obtained a significantly positive relationship between schooling and numeracy (Table 6).[151] Even after we have controlled for schooling investments, the immigration variable has a p-value of 0.005. The R^2 suggests that about 40% of the numeracy variation is explained with these two variables (Adj. R^2=0.38). We assume that many other variables are reflected by the schooling investments, such as the quality of institutions and ethnic variation. Thus it seems safe to say that mass immigration from Europe to a given Latin American country had a positive impact on human capital formation there.

To summarize: numeracy levels improved in all seven countries in our sample, but less so in Mexico, Ecuador, and Colombia than in the other four, and almost all of them suffered a stagnation or even a decline during the late 18th or early 19th century. Mass immigration from Europe provided human capital enhancing effects, even when schooling investments are controlled for.

3.7 Comparison of Latin American countries with European countries and the US

How do the trends compare with those in Western Europe, Eastern Europe, and the US (Figure 7)? For Western Europe we use the ABCC index for the UK reported in A'Hearn, Baten, and Crayen (2009), and for East-Central Europe we use comprehensive Hungarian

[149] Klein, *Migration to Brazil*.
[150] Benavot and Riddle, 'Education expansion'.
[151] We define it as 'more than 5,000 per decade and per country'.

data from the same source.[152] As for the US, we draw on the research of Fischer[153] into the 17th-century colonies that would later become the US. Working with a sample of some 4,000 individuals in Essex County, Massachusetts, in the years 1636-72 (that is, the 1620s and 1630s birth cohorts, for the most part), Fischer calculates the ratio between those colonists reporting ages with multiples of ten and a 10-year moving average. This ratio can be transformed into an ABCC index of 69% reporting exact ages. Another sample features adult males in Westchester County, New York (N = 259), most of them members of the 1670s birth cohort; their age-numeracy level was roughly 73% (Wells 1975).

Beginning with the birth decade of the 1800s we are on firmer ground, thanks to the censuses of 1850, 1870, and 1900. Because age heaping was greater in the South than elsewhere in the US during the 1800s, the increase in numeracy between the 1670s and the 1800s is slightly underestimated.

As for Western Europe, 16th-century English emigration records suggest an ABCC index of about 76% for the birth cohorts around 1600.[154] By about 1700 it would be close to 93%, on its way to 100% by the end of the next century.

During the 18th century Hungary's numeracy levels were on a par with Mexico's: slightly superior, that is, to Argentina's. However, by about 1780 trends in all seven Latin American countries under study, along with Hungary, had improved sufficiently to reduce the gap separating them from Western Europe and the US, which had been at 50% in 1650, to about 30%. However, Mexico's and Argentina's levels stagnated in the early 19th century, whereas Hungary's continued to rise, not only converging with but in fact surpassing, for a short while, those of the US, which stagnated, particularly in the South, until the middle of the century. During the late 19th century Argentina's trend resumed its upward course, reaching the basic-numeracy levels of Western Europe around 1880, whereas Mexico fell farther behind.

3.8 Conclusion

This paper makes the first systematic attempt to estimate the long-term development of human capital, measured in terms of basic numeracy, for seven Latin American countries from the 17th to the 20th century. Despite the existence of a large number of sizable 19th and

[152] Moreover, we use some additional early-19th-century data as well (Baten and Benyus, 'Hungarian numeracy').
[153] Fischer, *Ageing*.
[154] It should be noted, however, that the numeracy levels of emigrants tend to have an upward bias.

20th century censuses and of more limited ones from all four centuries, the data set suffers from some gaps and flaws, which we discussed intensively in this study. The large time span covers part of the colonial rule, the early post-independence period, the era of the first globalization as well as the first half of the 20th century. It thus permits new insights into the socioeconomic development of the New World. We discuss in detail the issue of source selectivity, and where possible adjust for regional composition before estimating long-term trends.

There was an overall increase in numeracy throughout the entire region and the entire period under study. Argentina, Mexico, and Peru (Lima) started, in the 17th century, with relatively low numeracy levels but by the 1780s they had reduced the gap separating them from Western Europe and the US from 50 to 30%. It is thus safe to say that until the late 18th century numeracy levels in Latin American countries were quite good. However, during the late 18th and early 19th centuries -- particularly the first two decades, marked by the wars of independence -- numeracy levels stagnated in many of these countries even as Western Europe's soared, and so the gap widened once again. In the late 19th century, numeracy gaps within Latin America increased, as well, Argentina, Uruguay, and Peru (Lima) at the upper end of the scale and Ecuador, Mexico, and Colombia at the lower. Brazil's numeracy levels stagnated until the 1860s but then began to improve. Immigration from Western Europe was positively associated with numeracy (even after we controlled for investment in public education). The fact that Ecuador, Mexico, and Colombia benefited relatively little from immigration and invested less in schooling helps to explain the fact that their numeracy levels stagnated in the late 19th century, and contributes to our understanding of their long-term histories.

What are the wider implications of the results? First, Latin America, which is partly still a developing region of the 21st century, was on a stable human capital growth path during the 18th century. We can quantify the long-term human capital retardation due to political conflicts of the early 19th century, which helps us to understand the more general relationship between conflict and human capital formation. Second, the historical evidence of immigration effects can be cautiously interpreted as a policy option for less-developed regions to invite skill-selective immigration if they are able to integrate (and attract) those immigrants.

3.9 References

A'Hearn, B., Baten, J., and Crayen, D., 'Quantifying quantitative literacy: age heaping and the history of human capital', *Journal of Economic History*, 69, (2009), pp. 783-808.

Andrade de Herrera, V., 'Education in Mexico: historical and contemporary educational systems', In Le Blanc Flores, J., ed., *Children of La Frontera: Binational Efforts to Serve Mexican Migrant and Immigrant Students*, ERIC Clearinghouse on Rural Education and Small Schools, (Charleston, SC, 1996), pp. 25-29.

Andrews, G. R., *The Afro-Argentines of Buenos Aires, 1800-1900*, (Madison, 1980).

Alden, D., 'The population of Brazil in the late eighteenth century: a preliminary study', *The Hispanic American Historical Review*, 43 (1963), pp. 173-205.

Apolant, J. A., *Génesis de la familia uruguaya*, 4 vols, (Montevideo 1975).

Arrom, S. M., *The women of Mexico City 1790-1857*. (Stanford, CA, 1985).

Astorga, P., Bérges, A. R., and FitzGerald, E. V. K., 'The standard of living in Latin America during the twentieth century', *The Economic History Review*, 58 (2005), pp. 765-796.

Austin, R., *The state, literacy, and popular education in Chile, 1964-1990*. (Lanham, MD, 2003).

Bachi, R., 'The tendency to round off age returns: measurement and correction', *Bulletin of the International Statistical Institute*, 33 (1951), pp. 195-221.

Baer, W., *The Brazilian economy: growth and development* (Boulder, 6th edn. 2001).

Bakewell, P. J., *A history of Latin America: c. 1450 to the present* (Malden, MA, 2004).

Barman, R. J., *Brazil: the forging of a nation, 1798-1852* (Stanford, CA, 1994).

Baten, J., and Benyus, R., 'Numeracy and protestantism in Hungary', Univ. of Tuebingen working paper (2009).

Baten, J., Crayen, D., and Manzel, K.,'Zahlenfaehigkeit und Zahlendisziplin in Nord- und Westdeutschland, 16.-18. Jahrhundert', *Jahrbuch fuer Wirtschaftsgeschichte*, 2 (2008), pp. 217-229.

Baten, J., Pelger, I, and Twrdek, L., 'The anthropometric history of Argentina, Brazil and Peru during the 19th and early 20th centuries', *Economics and Human Biology*, 7 (2009), pp. 319-333.

Benavot, A., and Riddle, P., 'The expansion of primary education, 1870-1940: trends and issues', *Sociology of Education*, 66 (1988), pp. 191-120.

Bertola, L. (2004) An overview of the economic history of Uruguay since the 1870s [WWW document]. URL http://eh.net/encyclopedia/article/Bertola.Uruguay.final, [accessed on 29 September 2010].

Bertola, L., Ocampo, J.A. *Desarrollo, vaivenes y desigualdad. Una historia económica de América Latina desde la Independencia.* Madrid, Segib, 2010.

Caillet-Bois, R. R., *Documentos para la historia argentina,* vols X and XI, (Buenos Aires, 1920-55)

Calvo, H. and Meisel, A., *Cartagena de Indias en el siglo XVIII* (Catagena, 2005), p. 326.

Cinnirella, F., 'Optimists or pessimists? A reconsideration of nutritional status in Britain, 1740–1865', *European Review of Economic History*, 12 (2008), pp. 325–354.

Clark, G., *A farewell to alms. A brief economic history of the world.* (Princeton, NJ, Oxford, 2007).

Cleveland, W. S., 'LOWESS: a program for smoothing scatterplots by robust locally weighted regression', *The American Statistician*, 35 (1981), p. 54.

Closs, M. P., 'Maya Mathematics', in Grattan-Guiness, I., ed. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 2 vols, (2003).

Coatsworth, J. H., 'Economic and Institutional Trajectories in Nineteenth-Century Latin America', in J .H. Coatsworth, and A. M. Taylor, eds., *Latin America and the World Economy Since 1800* (Cambridge, MA, 1998), pp. 149-175.

Conant, L. L., *The number concept: its origin and development* (New York, 1923).

Cook, S. F., and Borah, W., *Ensayos sobre la historia de la población: México y el Caribe* (Madrid, 1999)

Cook, N. D., *Padrón de los indios de Lima en 1613*, N.D. Cook ed. (Lima 1968).

Cook, N.D., *Numeración general de todas las personas de ambos sexos, edades y calidades q se hà echo en esta ciudad de Lima, año de 1700*. N.D. Cook, ed., (Lima 1985).

Corona Baratech, C. E., 'Notas para un estudio de la sociedad en el Río de la Plata durante el Virreinato', *Anuario de Estudios Americanos*, 8 (1951), pp. 59-167.

Cortés Conde, R., *El progreso Argentino, 1880-1940* (Buenos Aires, 1985).

Crayen, D., and Baten, J.,'Global trends in numeracy 1820-1949 and its implications for long-run growth', *Explorations in Economic History*, 47 (2010), pp. 82-99.

De Moor, T., and Van Zanden, J.-L., 'Uit fouten kun je leren. Een kritische benadering van de mogelijkheden van 'leeftijdstapelen' voor sociaal-economisch historisch onderzoek naar gecijferdheid in het pre-industriële Vlaanderen en Nederland', *Tijdschrift voor Economische en Sociale Geschiedenis*, 5 (2008), pp. 55-86.

Emigh, R. J., 'Numeracy or Enumeration? The uses of numbers by states and societies', *Social Science History*, 26 (2002), pp. 653-698.

Engerman, S. L., Haber, S., and Sokoloff, K. L., 'Inequality, institutions, and differential paths of growth among New World economies', in *Institutions, Contracts, and Organizations*, C. Menard, ed., (Cheltenham, 2000).

Fischer, D. H., *Growing Old in America*. (New York 1977).

Fisher, J. R., *Bourbon Peru, 1750-1824*, (Liverpool 2003).

Graden, D.T., 'An act "even of public security": slave resistance, social tensions and the end of international slave trade to Brazil, 1835-1856', *The Hispanic American Historical Review*, 76 (1996), pp. 249-282.

Gootenberg, P., 'Population and ethnicity in Peru: some revisions', *Latin American Research Review*, 26 (1991), pp. 109-157.

Haber, S., and Klein, H. S., 'The economic consequences of Brazilian independence', in S. Haber, ed., *How Latin America fell behind: Essays on the economic histories of Brazil and Mexico, 1800-1914*, (Stanford, CA, 1997).

Hernández Sánchez-Barba, M., 'La población hispanoamericana y su distribución social en el siglo XVIII', *Revista de estudios políticos*, 78 (1954), pp. 111-142.

Instituto Nacional de Estadística, Geografía e Informática. *Censo de población de la Ciudad de México, 1790* [CD ROM]: *Censo de Revillagigedo*. (Aguascalientes, 2003).

Julien, C. J., *Die Inka : Geschichte, Kultur, Religion*. (Munich 2001).

Klein, H., 'Asian and European migration to Brazil', in R. Cohen, ed., *The Cambridge Survey of World Migration*, (Cambridge 1995), pp. 208-214.

Leff, N. H., 'Long-term viability of slavery in a backward closed economy', *Journal of Interdisciplinary History*, 5 (1974), pp. 103-108.

Leininger Pycior, J., 'The history of education in Latin America: an overview', in P. J. Rodolf, ed., *History of Education*, (New York 1984), pp. 33-49.

Lewis, W. A., *Tropical development, 1880-1913* (London and New York 2006).

Lindert, P., *Growing public. Social spending and economic growth since the eighteenth century*, 2 vols. (New York, 2004).

Mabry, D. J., *Colonial Latin America*, (Coral Springs, FL, 2002).

Maddison, Angus (2007) World population, GDP and per capita GDP, 1-2001 AD, [WWW document]. URL http://www.ggdc.net/maddison/, [accessed on 29 September 2010].

Manzel, K., and Baten, J., 'Gender equality and inequality in numeracy: the case of Latin America and the Caribbean, 1880-1949', *Revista de Historia Económica*, 27 (2009), pp. 37-73.

Martinéz-Fritscher, A.C., and Musacchio, V., 'The great leap forward in education in Brazil: the political economy of education in Brazil, 1890-1940' Working paper, Harvard Business School 10-075, (2009).

Mariscal, E., and Sokoloff, K. L., 'Schooling, suffrage, and inequality in the Americas, 1800-1945', in Haber, S., ed., *Political Institutions and Economic Growth in Latin America*, (Stanford, 2000).

McAlister L. N., 'Social structure and social change in New Spain', *The Hispanic American Historical Review*, 43 (1963), pp. 349-370.

Mellafe, R., *Historia social de Chile y América* (Santiago de Chile, 2004).

Merino, O., and Newsom, L. A., 'Jesuit missions in Spanish America: The aftermath after the expulsion', *Revista de História de America*, 118 (1994), pp. 7-32.

Meyer Loy, C., 'Primary education during the Colombian Federation: the school reform of 1870' *Hispanic American Historical Review*, 51 (1971), pp. 275-294.

Mironov, B. N., 'Novaya Istoricheskaya Demografia Imperskoy Rossii: Analycheski obsor Corremennoy Istoriografii', *Vestnik Sankt-Peterbursgskovo Universiteta*, 4 (2006), pp. 62-90.

Mokyr, J., *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850* (London and Boston 1983).

Morse, R. M., 'Trends and patterns in Latin American urbanization, 1750-1920', *Comparative Studies in Society and History*, 16 (1974), pp. 416-447.

Myers, R., 'Accuracy of age reporting in the 1950 United States census', *Journal of the American Statistical Association*, XLIX (1954), pp. 826-831.

Newland, C., 'The estado docente and its expansion: Spanish American elementary education, 1900-1950', *Journal of Latin American Studies*, 26 (1994), pp. 449-467.

Núñez, C. E., and Tortella, G. (eds.). *La maldición divina: ignorancia y atraso en perspectiva histórica* (Madrid 1993).

Núñez, J., 'Signed with an X: methodology and data sources for analyzing the evolution of literacy in Latin America and the Caribbean, 1900-1950', *Latin American Research Review*, 40 (2005), pp. 117-135.

O'Grada, C., 'Dublin Jewish demography a century ago' *UCD Centre for Economic Research Working Paper Series*, (WP 06/01 2006).

Orozco y Berra, M., *Historia de la ciudad de México, desde su fundación hasta 1854*, Volume 112 of the Series "Sep-Setentas" published by the Secretaría de Educación Pública (SEP) (Mexico City. 1980).

Palafox, J. C., Prawda, J., and Velez, E., 'Primary schooling in Mexico', *Comparative Education Review*, 38 (1994), pp. 167-180.

Parro, A. A., 'Census of Peru: 1940', *Geographical Review*, 32 (1942), pp. 1-20.

Pérez Cantó, M. P., *Lima en el siglo XVIII* (Madrid, 1985).

Platt, L. D., *Census records for Latin America and the Hispanic United States* (Baltimore, MD, 1998).

Ramirez, M.T., and Salazar, I., 'Surgimiento de la educación en la República de Colombia, ¿En que fallamos?' *Borradores de Economia*, 454 (2007).

Reis, J., 'How helpful is age heaping? An exploration using Portuguese data', Univ. of Lisbon working paper (2008).

Ruggles, S., Trent Alexander, J., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M., Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis, MI: University of Minnesota, (2010).

Sianesi, B., and Van Reenen, J., 'The returns to education: macroeconomics', *Journal of Economic Surveys*, 17 (2003), pp. 157-200.

Somoza, J., and Lattes, A., 'Muestras de los dos primeros censos nacionales de población, 1869 y 1895', Documento de Trabajo Instituto T. Di Tella, 46, (Buenos Aires, 1967).

Stolz, Y., Baten, J., and Botelho, T., 'Growth effects of 19th century mass migrations: the case of Brazil', Univ. of Tuebingen working paper (2010).

Thorp, R., *Progress, poverty and exclusion: an economic history of Latin America in the 20th Century* (Baltimore, 1998).

Twrdek, L., and Manzel, K., 'The seed of abundance and misery: Peruvian living standards from the early republican period to the end of the Guano era (1820-1880)', *Economics and Human Biology,* 8 (2010), pp. 145-152.

UN: *Demographic Yearbook, Special Topic Population Censuses*. (New York 1955).

Vaughan, M.K., 'Primary education and literacy in nineteenth-century Mexico: research trends, 1968-1988', *Latin American Research Review,* 25 (1990), pp. 31-66.

Vinson, B., 'Race and Badge: Free-Colored Soldiers in the Colonial Mexican Militia', *The Americas* 56 (2000), pp. 471-496

Wells, R. V., *The population of the British colonies before 1776: a survey of census data* (Princeton, NJ, 1975)

Werner, M. S. *Concise Encyclopedia of Mexico* (Chicago, 2001)

Wolff, L., Schiefelbein, E., and Valenzuela, J., 'Improving the Quality of Primary Education in Latin America and the Caribbean', *World Bank Discussion Papers,* 257 (1994), Washington.

3.10 Tables

Table 1: Data Sources

Country/Region	Year	Observations (age range 23-62)	Indiv. data?	Bias relative to total population?
Argentina				
1744 Buenos Aires	1744	3,179	yes	urban, military census, incl. slaves
1771 Buenos Aires	1771	11,140	yes	urban, incl. slaves
1778 Buenos Aires	1778	895	yes	urban
1818 Buenos Aires	1818	890	yes	capital, perhaps no slaves
1869 Argentina	1869	38,776	yes	no
1869 Buenos Aires	1869	5,005	yes	capital
1887 Santa Fé	1887	1,102	yes	regional
1895 Argentina	1895	43,897	yes	no
1895 Buenos Aires	1895	7,818	yes	capital
1914 Argentina	1914	3,286,844	no	no
1947 Argentina	1947	7,729,939	no	no
Brazil				
1772 São Paulo	1772	2,347	yes	regional, excl. slaves
1830 Brazil	1830	267,595	no	regional, incl. slaves
1870 São Christovão	1870	691	yes	urban upper-class, incl. slaves
1890 Brazil	1890	5,536,449	no	no
1920 Brazil	1920	1,010,056	no	no, representative sample
1950 Brazil	1950	13,798,696	no	no
Colombia				
1777 Cartagena	1777	2,431	yes	merchant city
1777 Colombia	1777	1,554	yes	regional
1870 Colombia	1870	2,387	yes	various regions
1928 Colombia	1928	567	yes	various regions
1963 Colombia	1963	6,058,045	no	no
Ecuador				
1871 Ecuador	1871	29,151	yes	regional
1950 Ecuador	1950	2,211,838	no	no
Mexico				
1740-44 Hidalgo/Guanajuato/Oaxaca	1740-44	1,228	yes	regional
1777 Mexico - Central	1777	4,379	yes	regional
1777 Mexico - City	1777	608	yes	capital

1777 Mexico - North	1777	705	yes	regional
1790 Mexico - City	1790	4,212	yes	capital, all households headed by Spaniards, and mestizos
1821 Guadalajara	1821	16,625	no	regional
1823 Coahuila	1823	1,598	yes	regional
1930 Mexico	1930	4,967	yes	regional
1950 Mexico	1950	9,934,234	no	no
Peru				
1700 Lima	1700	2,797	yes	capital, household heads
1866-1909 Lima prison	1866-1909	4,392	yes	capital, prisoners
1940 Lima	1940	352,755	no	capital
1940 Peru	1940	2,370,166	no	no
Uruguay				
1772 Montevideo	1772	1,362	yes	capital
1791 Montevideo	1791	2,96	yes	capital
1834-36 Soriano/Maldonado	1834-36	1,166	yes	regional
1846 Montevideo prison	1846	1,565	yes	capital, prisoners
1868 Montevideo prison	1868	1,268	yes	capital, prisoners
1963 Uruguay	1963	1,290,319	no	no

Table 2: Ethnic Composition of the Early Census Samples

Ethnic group	Composition estimates in previous literature	Composition estimates in our census samples
Argentina: Buenos Aires 1744	Corona Baratech (1951), Bs As 1744	
White/Spanish	80.2	87.9
Indigenous and Mestizo	2.9	3.73
Black and Mulatto	16.9	6.56
Argentina: Buenos Aires 1771	Corona Baratech (1951), Bs As 1770	
White/Spanish	66.8	66.8
Indigenous and free Black	4.8	5.8
Slave	28.4	21.9 (Black slaves 13.58)
Mexico 1777: Oaxaca		
White/Spanish	12.33	5.45
Pardo	3.75	5.29
Indigenous	83.92	12.68
N/A	0	64.44
Mexico 1777: Durango		
	Cook and Borah (1999), Table 20b	
Mestizo	81.61	33.3
Indigenous	18.39	66.3
Mexico 1777: Mexico City	Hernández Sánchez-Barba (1954), 127, urban population in Mexico, ca. 1794	
White/Spanish	49.27	48.91
Casta	26.64	34.58
Indigenous	24.0	5.30
Other	0	11.21
Mexico 1790: Mexico City	Hernández Sánchez-Barba (1954),127, urban population in Mexico, ca. 1794	
White/Spanish	49.27	48.40
Casta	26.64	21.12
Indigenous	24.0	22.37
Black	0	0.56

Other	0	7.55
Peru 1700: Lima	Pérez Cantó (1984), Lima 1700	
White/Spanish	56.5	97.11
Indigenous	11.7	1.5
Mulatto	9.7	0.75
Black	22.1	0.3
Peru 1871: Lima prison	Fisher (2003), 56, Peru in late 18th century	
White/Spanish	38.46	17.69
Indigenous	7.69	21.31
Casta	9.6	25.22
Black	44	13.90

Table 3: Regression of ABCC Outcomes on Reasons for Enumeration, All Countries

Dependent Variable	ABCC
Military	4.98
	(0.403)
Prison	4.88
	(0.198)
Country-Fixed Effects	included
Time-Fixed Effects	included
Constant	27.70***
	(0.004)
Observations	171
R-squared	0.74

in parentheses

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

Note: Reference Category is an enumeration reason other than military or prison, the country is Argentina, and the birth decade is the 1640s.

Table 4: Regressions of numeracy (ABCC Index) on Male Share, Capital Effect, and Birth Decade for Argentina and Mexico

	(1)	(2)	(3)
Dependent Variable	ABCC	ABCC	ABCC
Country	Argentina	Mexico	Mexico
Capital	6.41***	-0.43	-1.37
	(0.000)	(0.917)	(0.812)
Male share	0.31**	1.59**	1.52*
	(0.037)	(0.024)	(0.059)
Indigenous share			-0.06
			(0.789)
b1680	-80.30***	-30.25***	-31.06**
	(0.000)	(0.005)	(0.022)
b1690	-82.35***	-19.18*	-19.99
	(0.000)	(0.054)	(0.118)
b1700	-74.16***	-14.12	-14.94
	(0.000)	(0.145)	(0.234)
b1710	-64.16***	-15.72**	-16.38
	(0.000)	(0.047)	(0.127)
b1720	-56.28***	-19.07**	-19.69*
	(0.000)	(0.033)	(0.090)
b1730	-52.03***	-14.40*	-14.89
	(0.000)	(0.071)	(0.164)
b1740	-41.13***	-14.28*	-14.77
	(0.000)	(0.073)	(0.167)
b1750	-43.30***	0.51	0.38
	(0.000)	(0.960)	(0.976)
b1760	-29.70***	-8.99	-8.88
	(0.000)	(0.218)	(0.384)
b1770	-33.38***	-6.87	-6.65
	(0.000)	(0.377)	(0.534)
b1780	-26.68***	-4.21	-3.98
	(0.000)	(0.585)	(0.708)
b1790	-27.80***	-10.26	-10.03

	(0.000)	(0.194)	(0.353)
b1800			
b1810	-40.45***		
	(0.000)		
b1820	-34.65***		
	(0.000)		
b1830	-29.51***		
	(0.000)		
b1840	-27.12***		
	(0.000)		
b1850	-19.67***		
	(0.000)		
b1860	-17.81***		
	(0.000)		
b1870	-9.34*	-13.25	
	(0.060)	(0.170)	
b1880	-3.42	-7.71	
	(0.409)	(0.416)	
b1890	-0.00	-4.60	-4.94
	(1.000)	(0.550)	(0.682)
b1900	-0.00	6.68	6.41
	(1.000)	(0.480)	(0.595)
b1920		13.46	13.19
		(0.164)	(0.283)
Constant	84.71***	-9.51	-4.26
	(0.000)	(0.770)	(0.918)
Observations	44	36	32
R-squared	0.98	0.70	0.69

Note: p-values in parentheses. */**/*** denote statistical significance at the 10, 5, and 1% levels, respectively. The constant represents the numeracy of female, non-capital inhabitants born in 1910-19 (in Col. 3 non-indigenous).

Table 5: Regional Distribution of Observations in Colombia, Sorted by Regional Numeracy in 1930-39

Region	ABCC, birth decade 1930s	Number of observations on numeracy in		
		1777	1870	1928
Central	88	236		
Amazonia	91		1016	419
Andina Sur	91		287	
Andina Norte	91	447		
Orinoquia	91		176	
Pacifico Norte	93		184	148
Caribe	93	99	100	
Pacifico Sur	93		597	
Eje Cafetero	95			
Bogotá	98			

Table 6: Regression: Impact of Immigration and Schooling on Numeracy Development

	(1)
Dependent Variable	ABCC
Immigration Dummy	12.68***
	(0.005)
Schooling	0.46**
	(0.042)
Constant	67.17***
	(0.000)
Observations	32
R-squared	0.42

in parentheses

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

Note: Immigration is measured by an indicator variable that is 1 when a country experiences substantial immigration, defined as over 5,000 arrivals per decade. Data on immigration are from Mitchell (1975) and schooling data are from Benavot and Riddle (1983); for the numeracy estimates, see text. The schooling data allow us to include the following countries and decades (decades during which immigration was substantial are provided in parentheses): Argentina 1870 and 1890-1910 (1860-1910), Brazil 1870-1920 (1880-1920), Colombia 1880-1930, Ecuador 1890-1920, Mexico 1870 and 1890-1920 (1900-1920), Peru 1890-1910, and Uruguay 1900-1930 (1840 and 1900-1930). Results also hold with controls for time-fixed effects, but in this case the schooling estimates are no longer significant.

3.11 Figures

Figure 1: Age statements by literacy status (1869 census of Argentina in 1869)

N

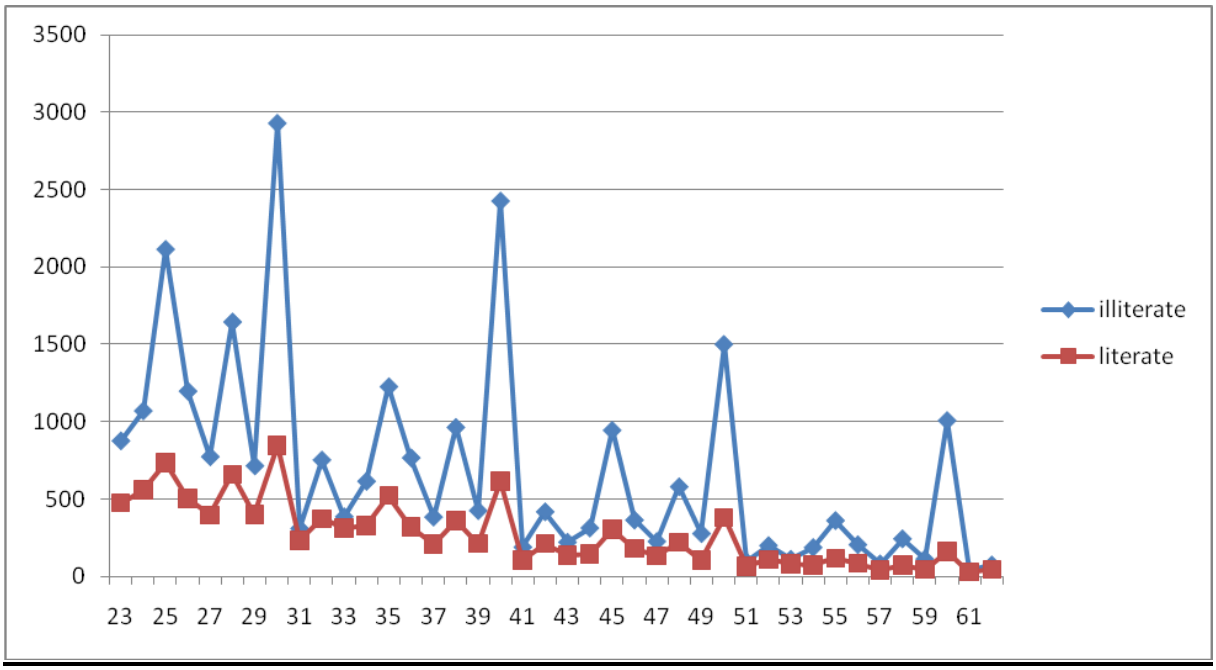
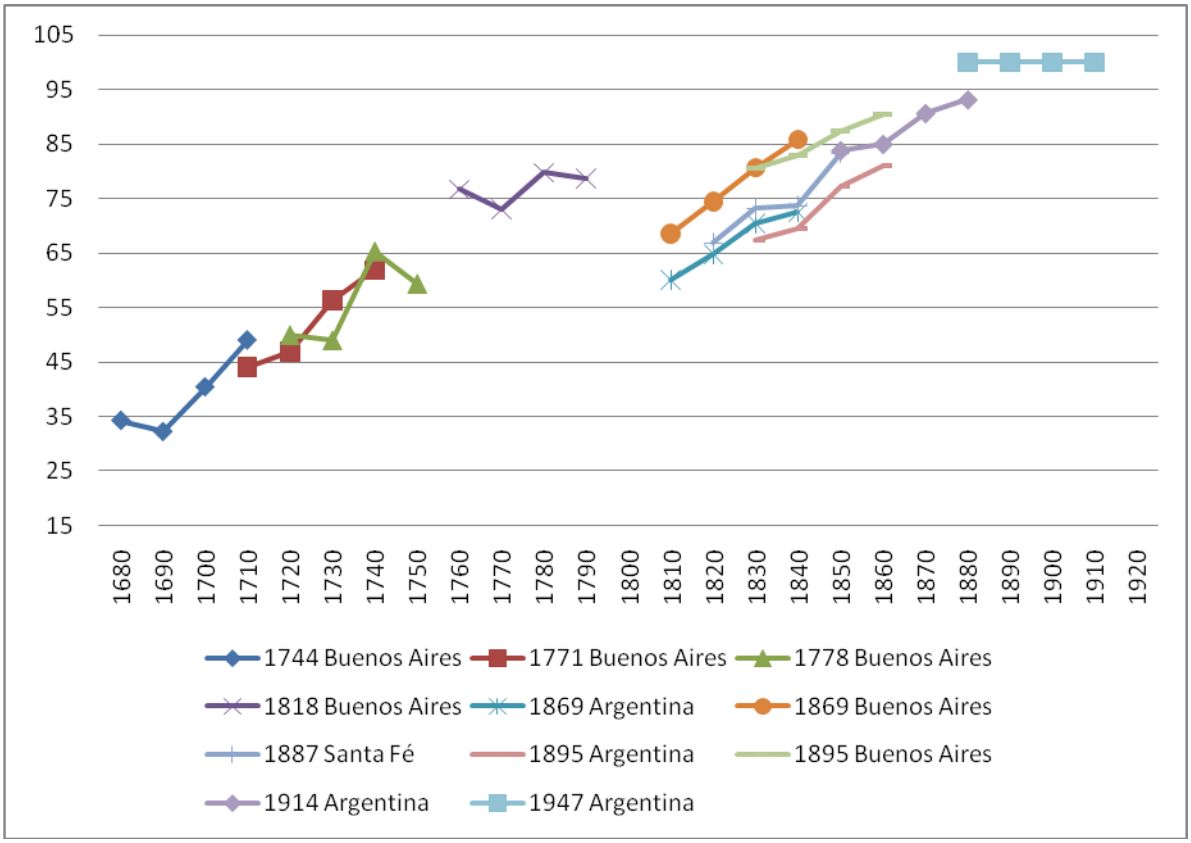
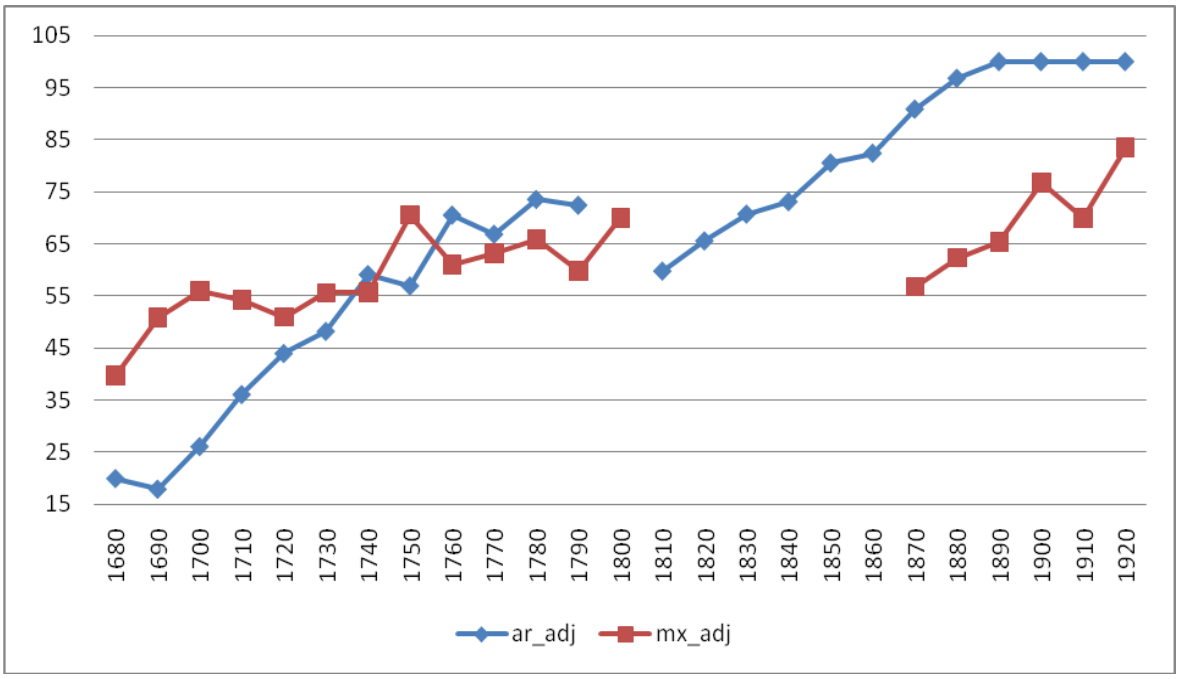


Figure 2: Argentina - ABCC index of basic numeracy (vertical axis) by birth decades



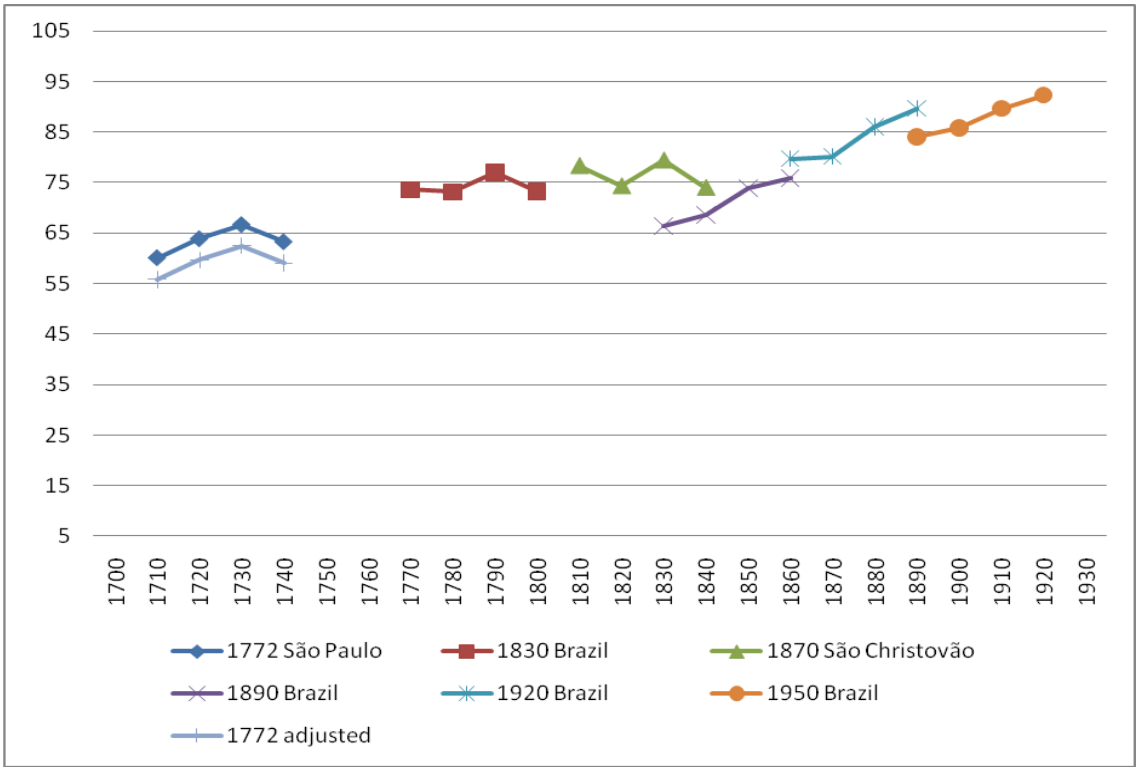
Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Figure 3: Trends of the ABCC index for Argentina and Mexico, (vertical axis) by birth decades, controlling for capital effect and gender composition



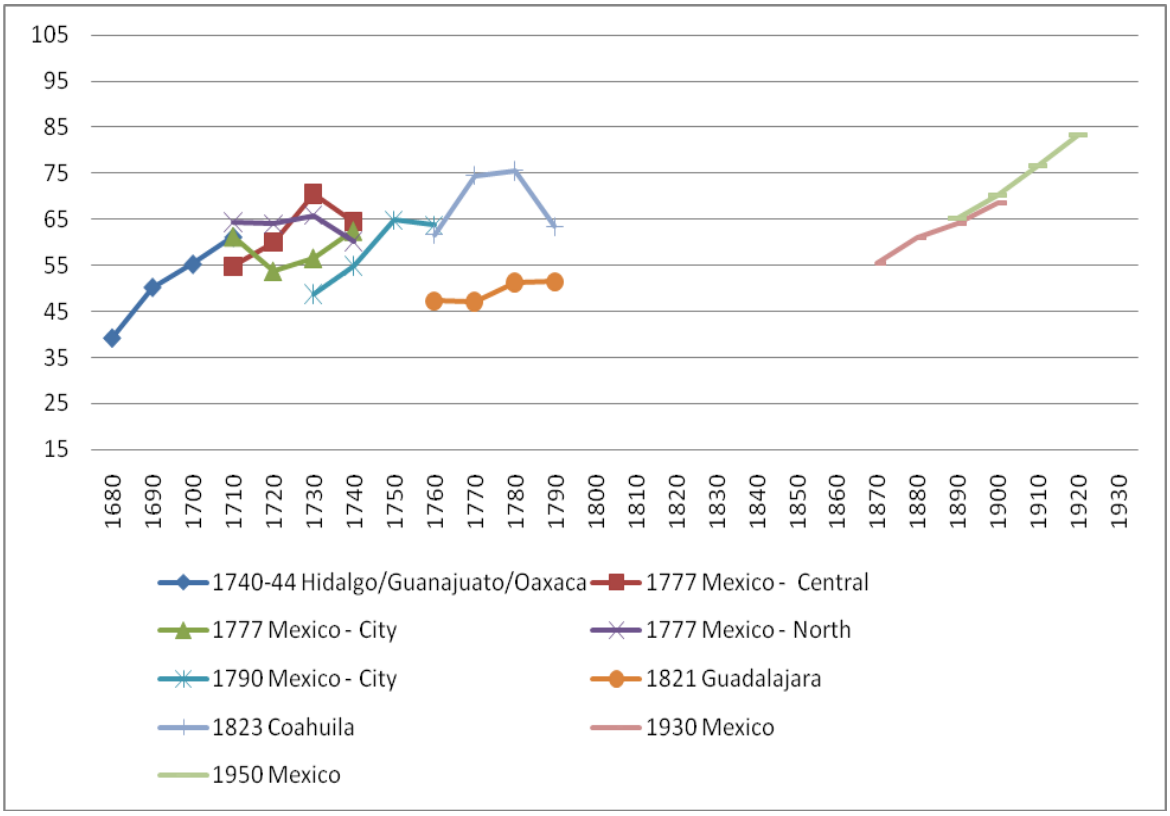
Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Figure 4: Brazil - ABCC index of basic numeracy (vertical axis) by birth decades



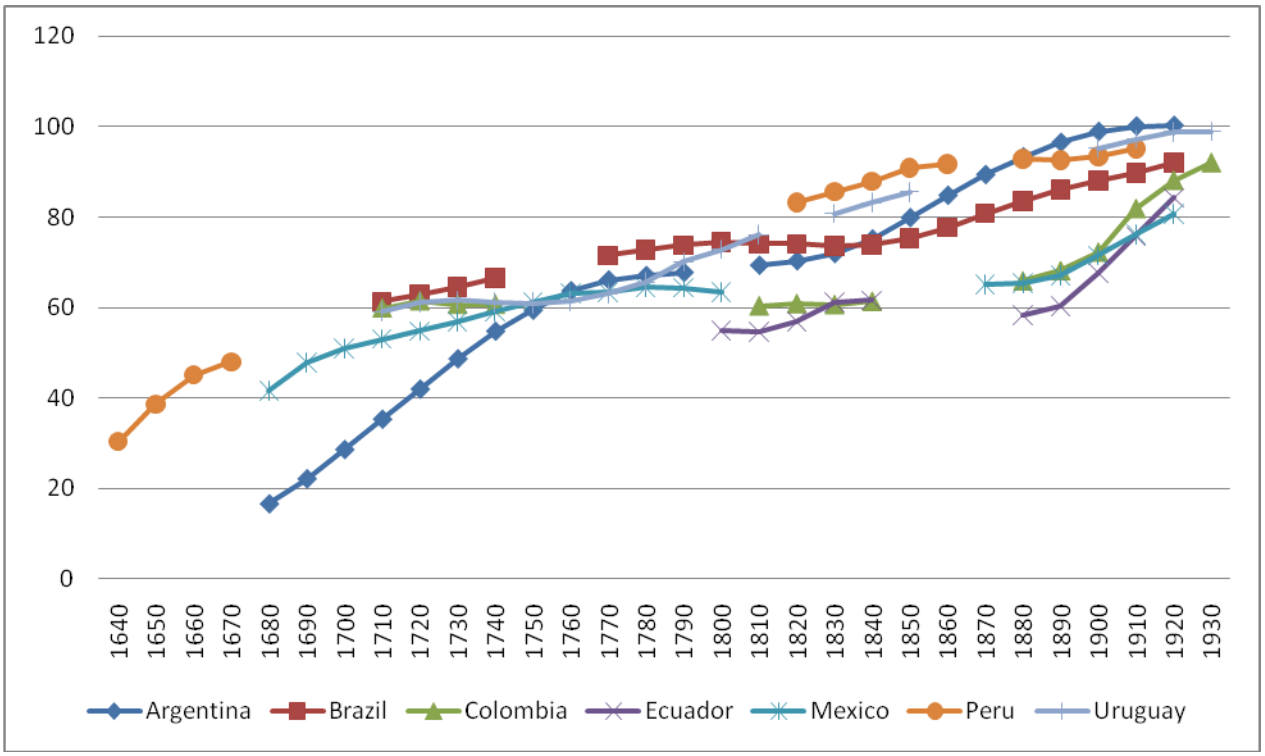
Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Figure 5: Mexico - ABCC index of basic numeracy (vertical axis) by birth decades



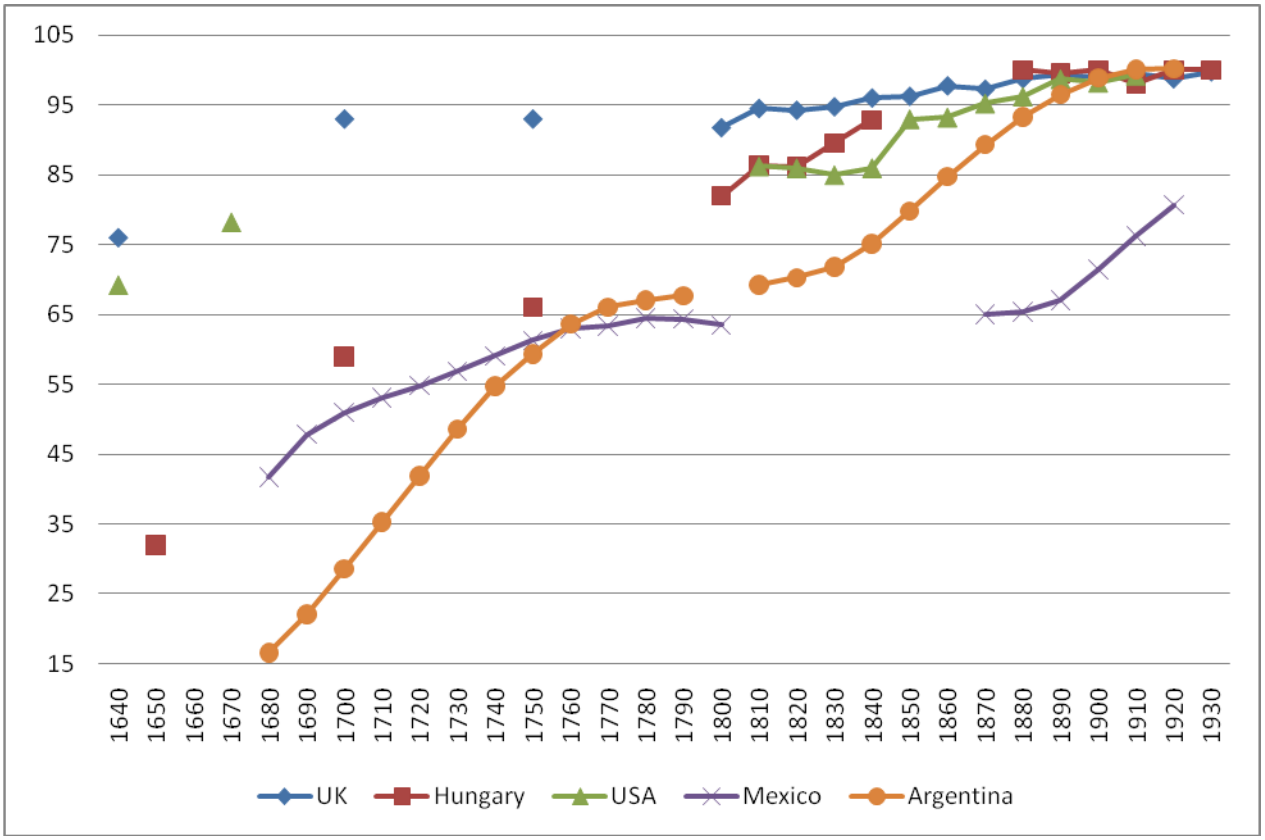
Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Figure 6: Argentina, Brazil, Colombia, Ecuador, Mexico, Peru (Lima), and Uruguay -- ABCC index of basic numeracy (vertical axis) by birth decades, LOWESS-transformed



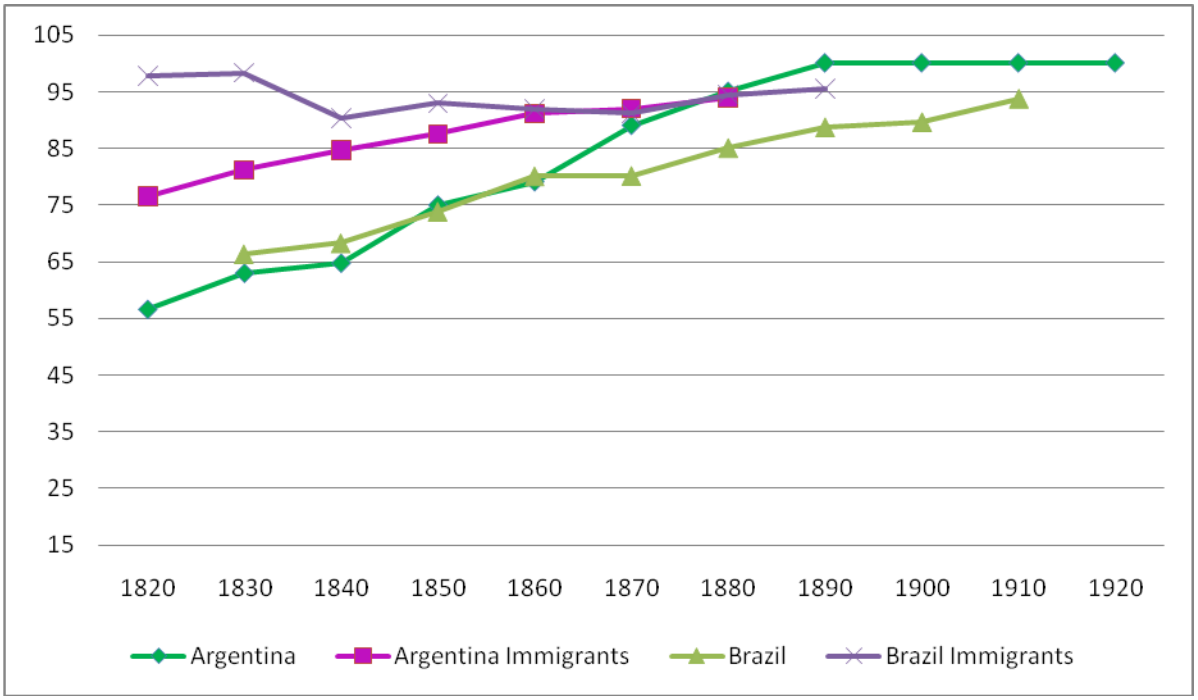
Note : Only the values for Lima are included in the Peruvian data. For Argentina and Mexico, estimates are based on regressions (controlling for capital effect and male share). LOWESS bandwidth is 0.5. The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1. Non-representative samples, such as that of São Christovão, for Brazil, are excluded.

Figure 7: ABCC numeracy indexes for Argentina, Mexico, the UK, Hungary, and the US (vertical axis) by birth decades



Notes: For Argentina and Mexico, estimates are based on regressions (controlling for capital effect and male share). The value "UK 1640" actually refers to the ABCC of around 1600; the earliest US value is for Massachusetts 1620s and 1630s. Hungary during the early 19th century refers to the Hungarian part of the Habsburg monarchy. Sources for Hungary and the UK are from A'Hearn, Baten, and Crayen (2009). For Hungary, the US, and the UK during the 19th century, see Crayen and Baten (2009). Some of the Hungarian values come from Baten and Benyus (2009). The US values are based on the Integrated Public Use Micro Samples (IPUMS); see Ruggles et al. (2004). The decades refer to birth decades (1800-1809, etc.). We thank Rita Benyus, Brian A'Hearn, Dorothee Crayen, and the IPUMS team for providing data.

Figure 8: Numeracy of immigrants and natives in Brazil and Argentina



Note: Numeracy of Argentine Data taken from the Argentine censuses of 1869, 1895, and 1914 and also from Brazilian data on native numeracy; see text. Brazilian immigrants are documented in Stolz, Baten, and Botelho (2010).

3.12 Appendix A: Data Sources (to be provided on an Internet page)

Sources for Argentina

There are a large number of primary sources that provide information on the development of numeracy in Argentina, the earliest being the 1744 Buenos Aires military census, reproduced in the *Documentos para la Historia Argentina*[155]. For Buenos Aires we also have 1771 census data. We use information on 11,140 individuals in this data set (children and the elderly are excluded from our calculations). Data from the military census of 1818 fill the gap between 1771 and 1869. In addition, we incorporate into our analysis the samples of the first two national population censuses, those of 1869 and 1895, collected by Somoza and Lattes[156], which contain extensive information on a representative nationwide sample of the population. These data provide not only insights into regional differences but information on urban and rural areas as well. The time series is complemented with aggregated data from the 1914 and 1947 censuses.

Sources for Brazil

Our earliest evidence for Brazil, dating from 1772, consists of a single sample of a very early census of the São Paulo district of Sorocaba, representing the birth decades 1700-1750 (N=6,279). While this census does not include slaves, it does include *agregados*, akin to live-in servants in pre-industrial rural northern Europe. *Agregados* were free but mostly unpaid, performing household tasks in exchange for food and shelter[157]. We also include a complete aggregated 1830 census evidence for Paraná, Minas Gerais, and São Paulo, which includes information also on slaves and thus can be deemed representative of late-18th -century Brazil, or at least the southern region thereof. Further evidence comes from the 1890 and 1920 censuses. The 1920 data are limited in that only the provincial capitals are reported in sufficient age-specific detail. Last but not least, we have the nationwide aggregated 1950 census.

Sources for Colombia

[155] Caillet-Bois, *Documentos para la historia argentina,* vols X and XI.

[156] Somoza and Lattes 'Muestras'.

[157] Stolz, Baten, and Botelho, 'Mass Immigrations', p. 9.

The Colombian census data derive from the National Archive in Bogotá, the earliest being population counts from the year 1777 from several central and Northern provinces.[158] The census records of Cartagena de las Indias have been studied extensively by Calvo and Meisel.[159] Information on four districts of this important merchant city has survived. The 1870 census' information on Cauca, Magdalena, Chocó, Quindio, and other departments and thus hints offers hints about the development of basic numerical abilities in a wide variety of provinces. Finally, the 1928 census of Putumayo, Vaupés, and Chocó provides individual age data.

Sources for Ecuador

Our Ecuador data, mostly for the year 1871, derive from the *Archivo Nacional de Quito*. The complete enumeration, extending from the 1860s to 1893 and comprising 71,545 observations, covers the provinces of Manabi, Azuay, and Pichincha. The census offers complete breakdown of the enumerated population. To account for the birth decades of the late 19th and early 20th centuries, we include the published ABCC values taken from the 1950 nationwide census.[160]

Sources for Mexico

The *Archivo de las Indias* in Seville houses a considerable wealth of primary sources for Mexico. Our sample includes population censuses from 1740 to 1743 for Hidalgo, Guadalajara, and Oaxaca. We also had access to age data from 1777 for Mexico City, Durango, Chihuahua, Baja California, Oaxaca, Puebla, and Veracruz. The *Censo de Revillagigedo,* our most important source for all of northern Latin America, and carried out in Mexico between 1790 and 1794, was the first census equipped with a standard format for listing individuals by name, age, sex, and family status.[161] Those census forms that have survived provide information on 15 Mexico City districts, from which we have drawn a sample of 4,212 individuals.

The Guadalajara Census, a joint project of researchers from Mexico and the US, aims to preserve and to provide public access to census data for Guadalajara, a province in the Western Pacific area of Mexico. Another source provides data from the 1823 census of

158 We thank Adolfo Meisel for providing additional data 1777 Cartagena de las Indias sample.

159 Calvo and Meisel, *Cartagena*, p. 326.

160 Manzel and Baten, 'Gender equality', pp. 37-73.

161 Werner, *Concise Encyclopedia*.

Coahuila, a province bordering Texas. For the 20th century, we have data from 1930 on several other provinces and aggregated data from the 1950 nationwide census.

Sources for Peru

Our earliest source of Peruvian age data is a 1770 Lima population census considered one of the most important of the colonial period.[162] Because its main purpose was to determine the number of men available for military service in case of foreign aggression, the data are limited to the age, profession, and race of male household heads. Thus we have information on no more than about 3,000 of Lima's 37,000 inhabitants[163], and none at all on the female portion of Lima's population: a situation that virtually guarantees a numeracy overestimation -- all the more nearly certain because Peru's indigenous population was underrepresented in the sample since it was underrepresented in Lima, relative to the proportion that it constituted of other cities' populations.[164] Our analysis of numeracy in Peru is therefore primarily an analysis of the educational development of Lima's elite. However, we also have access to the 1940 nationwide census data, which shed light on the birth cohorts from the 1880s onwards, permitting us to compare numeracy rates in Lima with that in Peru generally. In addition, we have access to a Lima prison sample that extends from 1866 to 1909. Although one assumes, understandably, that this prison sample overrepresents Lima's underclass, in fact the age heaping to be found in the portion of this sample spanning the 1880s is quite similar to that of Lima's general population.[165]

Sources for Uruguay

The National Archive in Montevideo contains a variety of data sets valuable for the study of the evolution of numeracy in Uruguay: prison records (1847-1868), the 1832 census of Soriano, and the 1836 census of Maldonado. Finally, there is also the published nationwide census of 1963.

[162] Cook *Numeración.*

[163] Pérez Cantó, *Lima,* p. 185.

[164] Mabry., *Colonial, p. 58.*.

[165] Baten, Pelger and Twrdek, 'Anthropometric History', p. 19.

Data Sources

Argentina:

Military Census 1744: Reproduced in Caillet-Bois, R. R. (1919); **Census 1771**: Archivo Nacional Buenos Aires, Sala IX, Documentos de Gobierno, Censos y padrones; **Census Alcalde Matheo 1778; Census of Santa Fé 1887:** http://www.digitalmicrofilm.com.ar /censos/geografico.php; **Census of Buenos Aires 1818**: Archivo Nacional Buenos Aires, Sala IX, Documentos de Gobierno, Censos y padrones, **Census of Argentina 1869**: Somoza and Lattes (1967); **Census of Argentina 1895:** Somoza and Lattes (1967), **Census of Argentina 1914**: Archivo Nacional Buenos Aires, Sala IX, Documentos de Gobierno, Censos y padrones, **Census of Argentina 1947**: UN (1955): Demographic Yearbook, p. 311.

Brazil:

Census São Paulo, Sorocaba 1772: Arquivo Hist«orico Ultramarino, Lisboa, Portugal, AHU_ACL_CU_,Cód.1270, **Census of Paraná, Minas Gerais, São Paulo 1830: # Census of Brazil 1890**: Biblioteca do IBGE, [http://biblioteca.ibge.gov.br/]; **Census of Brazil 1920**: http://biblioteca.ibge.gov.br/ **Census of Brazil 1950**: UN (1955), Demographic Yearbook, p. 313.

Colombia:

Census of Cartagena 1777: Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 9 and 58, Meisel (2005); **Census of Media Granada 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 11; **Census of Magdalena 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 15-16; **Census of Mogotes 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 38; **Census of San Juan Girón 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 53; **Census of Sativa 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 61; **Census of Bolivar 1777:** Archivo Nacional de Bogotá, Microfilm 23, Volume 8 (1), No. 62; **Census of Cauca 1870:** Archivo Nacional de Bogotá, Microfilm 2, No. 4; **Census of Magdalena 1870:** Archivo Nacional de Bogotá, No. 6; **Census of Quibdo 1870:** Archivo Nacional de Bogotá, Microfilm 2, No. 15; **Census of Quindio 1870:** Archivo Nacional de Bogotá, Microfilm 2, No. 19; **Census of Putumayo 1928:** Archivo Nacional de Bogotá; **Census of Vaupés 1928:** Archivo Nacional de Bogotá; **Census of Chocó 1928:** Archivo Nacional de

Bogotá; Census of Colombia 1963: UN (1972): Demographic Yearbook. For the regional breakdown in Table 5 we used the IPUMS International Sample on the Colombian Census of 1963; see Ruggles et al. (2004).

Ecuador

Census for all Ecuador excluding the Amazonas region and the Galapagos Archipelago. It was taken between **1861 and 1893**. Archivo Nacional de Quito. **1950** Census of Ecuador. See Manzel and Baten 2009.

Mexico

[Mexico 1740-44] Census of Ixmiquilpan 1740: Archivo General de Indias, Ind, 107; **Census of Pozos 1743:** Archivo General de Indias, Ind, 107; **Census of southern central Mexico 1743:** [Place unreadable, Platt (1998): Tlazazalca, Michoacán, Tetela del Rio or Guerrero], Archivo General de las Indias, Ind, 108; **Census of Chichihualtepec 1743:** Archivo General de las Indias, Ind, 108;

[Centro 1777] Census of Ciudad de los Angeles 1777: Archivo General de las Indias, Mex, 2578; **Census of Piaxtla 1777:** Archivo General de las Indias, Mex, 2578; **Census of Totoltepec 1777:** Archivo General de las Indias, Mex, 2579; **Census of Nopalucan 1777:** Archivo General de las Indias, Mex, 2579; **Census of Quanquecholan 1777:** Archivo General de las Indias, Mex, 2579;

[Norte 1777] Census of San Gregorio 1777: Archivo General de las Indias; **Census of Real de Minas de Nuestra Senora del Rosario 1777:** Archivo General de Indias, Gua, 103 and Gua, 250;**Census of los Remedios 1777:** Archivo General de las Indias, Ind, 1526; **Census of San José de Animas 1777:** Archivo General de las Indias; **Census of San José de Pimas 1777:** Archivo General de las Indias; **Census of Suchil 1777:** Archivo General de las Indias; **Census of Penol 1777:** Archivo General de las Indias; **Census of Tonanchi 1778:** Archivo General de las Indias;

[Mexico City 1777] Census of Mexico City 1777: Archivo General de las Indias;

Census of Revillagigedo 1790: Instituto Nacional de Estadística, Geografía e Informática (2003); **Census of Guadalajara 1821:** Guadaljara Census project [http://www.fsu.edu/]; **Census of the Municipality of Monclova 1822-23**; **Censo del Pueblo de San Francisco de Tlaxcala 1823**; **Censo de la Hacienda de Castaños y Bajan 1822-23**; **Censo de la Hacienda de Alamo 1823**; **Censo de la Hacienda de Encinas 1823**; **Censo de la Hacienda de San Vicente el Alto 1823**; **Censo de la Hacienda de Santa Ana 1823**;

Censo de la Hacienda de San Juan Bautista 1823; **Censo de la Hacienda de San José 1823**; **Censo de la Hacienda de San Ignacio del Paso Tapado 1823:** Grupo Exploradores Coahuiltecos [http://mx.geocities.com/camino_real_ mva/]; **Census of Guanajuato 1930:** FSI, Microfilm 4107114; **Census of Minas de Luz 1930:** FSI, Microfilm 4107114; **Census of Mineral de los Llamitos 1930:** FSI, Microfilm 4107114; **Census of Ahualuco 1930:** FSI, Microfilm 4107751; **Census of Coyuca de Benitez 1930:** FSI, Microfilm 4107141; **Census of Tepoztlán 1930:** FSI, Microfilm 4107265; **Census of Mezquital 1930:** FSI, Microfilm 4107065; **Census of Tetecala 1930:** FSI, Microfilm 4107265; **Census of Tlaltizapan 1930:** FSI, Microfilm 4107265; **Census of Tetecala 1930:** FSI, Microfilm 4107265; **Census of Tlaltizapan 1930:** FSI, Microfilm 4107265; **Census of Mexico 1950:** UN (1955), Demographic Yearbook, p. 304.

Peru

Census of Lima 1700: Reproduced in Cook (1985); **Prison sample Lima 1871:** Manzel and Twrdek (2009); **Census of Peru 1940:** Parro (1942).

Uruguay

Padrón of Aldecoa 1772: Reproduced in Apolant (1975), Volume III; **Census of Soriano 1834:** Archivo Nacional de Montevideo; **Census of Maldonado 1836:** Archivo Nacional de Montevideo; **Prison sample 1846:** Archivo Nacional de Montevideo, **Prison sample 1868:** Archivo Nacional de Montevideo; **Census of Uruguay 1963:** UN (1972): Demographic Yearbook, p. 214.

United Kingdom

Before 1800: A'Hearn, Baten, and Crayen (2009). After 1800: Crayen and Baten (2009), Baten and Benyus (2009).

United States

Census of Westchester County: Wells (1975); **Census of 1850, 1870 and 1900:** A'Hearn, Baten, and Crayen (2009), based on Integrated Public Use Micro Samples (IPUMS). See Ruggles et al., Integrated Public Use.

3.13 Appendix B: Whipple and ABCC indexes (to be provided on an Internet page)

Data Source	Birth decade	Observations	Whipple	ABCC
1744 Buenos Aires	1680	303	363	34
1744 Buenos Aires	1690	501	371	32
1744 Buenos Aires	1700	941	338	40
1744 Buenos Aires	1710	1,434	304	49
1771 Buenos Aires	1710	949	324	44
1771 Buenos Aires	1720	1,832	313	47
1771 Buenos Aires	1730	3,358	275	56
1771 Buenos Aires	1740	5,001	253	62
1778 Buenos Aires	1720	105	300	50
1778 Buenos Aires	1730	153	304	49
1778 Buenos Aires	1740	243	239	65
1778 Buenos Aires	1750	394	262	59
1818 Buenos Aires	1760	88	193	77
1818 Buenos Aires	1770	190	208	73
1818 Buenos Aires	1780	254	181	80
1818 Buenos Aires	1790	358	186	79
1869 Argentina	1810	3,221	260	60
1869 Argentina	1820	6,598	241	65
1869 Argentina	1830	11,142	218	70
1869 Argentina	1840	17,815	210	72
1869 Buenos Aires	1810	352	226	69
1869 Buenos Aires	1820	826	202	74
1869 Buenos Aires	1830	1,493	177	81
1869 Buenos Aires	1840	2,334	157	86
1887 Santa Fé	1820	114	232	67
1887 Santa Fé	1830	234	207	73
1887 Santa Fé	1840	280	205	74
1887 Santa Fé	1850	474	167	83
1895 Argentina	1830	4,001	230	67
1895 Argentina	1840	7,989	222	69

1895 Argentina	1850	13,875	191	77
1895 Argentina	1860	18,032	176	81
1895 Buenos Aires	1830	682	177	81
1895 Buenos Aires	1840	1,415	169	83
1895 Buenos Aires	1850	2,548	150	87
1895 Buenos Aires	1860	3,173	139	90
1914 Argentina	1850	340,213	165	84
1914 Argentina	1860	574,992	160	85
1914 Argentina	1870	922,034	137	91
1914 Argentina	1880	1,449,605	127	93
1947 Argentina	1880	1,140,200	96	100
1947 Argentina	1890	1,697,562	100	100
1947 Argentina	1900	2,286,936	99	100
1947 Argentina	1910	2,605,241	98	100
1772 São Paulo	1740	968	247	63
1772 São Paulo	1730	639	233	67
1772 São Paulo	1720	409	244	64
1772 São Paulo	1710	331	260	60
1830 Brazil	1800	120,000	208	73
1830 Brazil	1790	73,325	193	77
1830 Brazil	1780	46,458	211	73
1830 Brazil	1770	27,812	212	74
1870 São Christovão	1810	59	186	78
1870 São Christovão	1820	143	203	74
1870 São Christovão	1830	239	182	79
1870 São Christovão	1840	250	204	74
1890 Brazil	1830	586,793	235	66
1890 Brazil	1840	1,021,027	226	69
1890 Brazil	1850	1,605,498	205	74
1890 Brazil	1860	2,323,131	197	76
1920 Brazil	1860	102,312	181	80
1920 Brazil	1870	180,316	179	80
1920 Brazil	1880	279,862	156	86

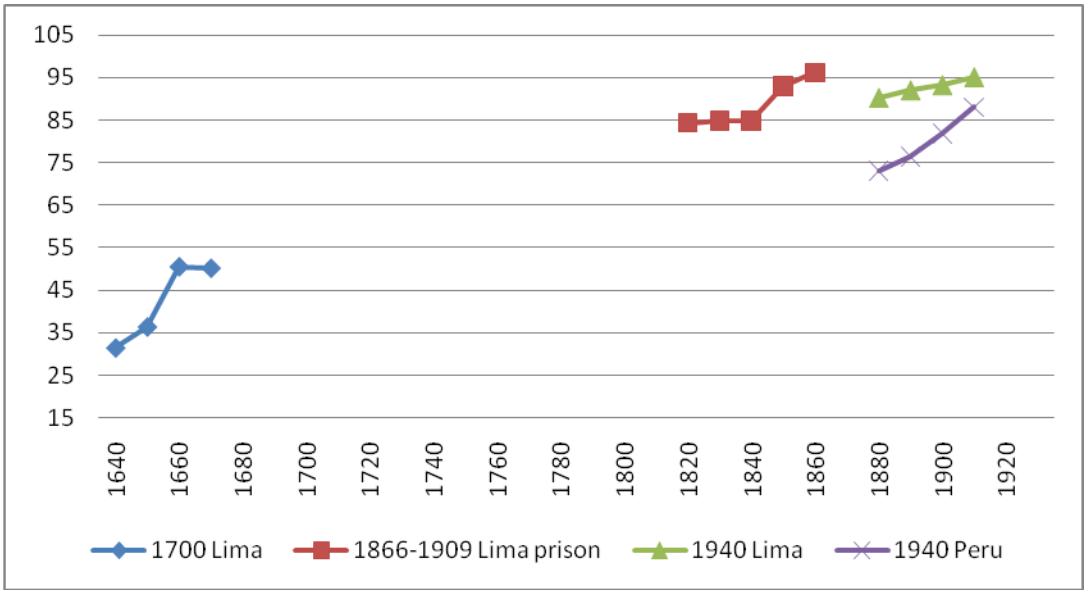
1920 Brazil	1890	447,566	141	90
1950 Brazil	1890	2,221,106	164	84
1950 Brazil	1900	3,844,441	157	86
1950 Brazil	1910	5,774,083	141	90
1950 Brazil	1920	8,143,411	131	92
1777 Cartagena	1710	269	257	61
1777 Cartagena	1720	476	251	62
1777 Cartagena	1730	667	207	73
1777 Cartagena	1740	1,019	275	56
1777 Colombia	1710	168	265	59
1777 Colombia	1720	254	289	53
1777 Colombia	1730	398	241	65
1777 Colombia	1740	734	275	56
1870 Colombia	1810	254	260	60
1870 Colombia	1820	399	259	60
1870 Colombia	1830	557	251	62
1870 Colombia	1840	1,177	250	62
1928 Colombia	1880	85	288	53
1928 Colombia	1890	203	268	58
1928 Colombia	1900	279	223	69
1963 Colombia	1900	727,666	175	81
1963 Colombia	1910	1,178,083	160	85
1963 Colombia	1920	1,796,228	142	89
1963 Colombia	1930	2,356,068	132	92
1871 Ecuador	1800	1,531	280	55
1871 Ecuador	1810	3,053	287	53
1871 Ecuador	1820	4,800	275	56
1871 Ecuador	1830	8,049	254	61
1871 Ecuador	1840	11,718	238	66
1950 Ecuador	1880		290	53
1950 Ecuador	1890		256	61
1950 Ecuador	1900		234	67
1950 Ecuador	1910		195	76

1950 Ecuador	1920		163	84
1740-44 Hidalgo/Guanajuato/Oaxaca	1680	118	343	39
1740-44 Hidalgo/Guanajuato/Oaxaca	1690	189	299	50
1740-44 Hidalgo/Guanajuato/Oaxaca	1700	348	279	55
1740-44 Hidalgo/Guanajuato/Oaxaca	1710	573	255	61
1777 Mexico - Central	1710	550	281	55
1777 Mexico - Central	1720	777	260	60
1777 Mexico - Central	1730	1,437	218	70
1777 Mexico - Central	1740	1,615	242	64
1777 Mexico City	1710	53	255	61
1777 Mexico City	1720	86	285	54
1777 Mexico City	1730	170	274	57
1777 Mexico City	1740	299	250	63
1777 Mexico - North	1710	70	243	64
1777 Mexico - North	1720	125	244	64
1777 Mexico - North	1730	184	236	66
1777 Mexico - North	1740	326	259	60
1790 Mexico City	1730	341	305	49
1790 Mexico City	1740	755	280	55
1790 Mexico City	1750	1,271	241	65
1790 Mexico City	1760	1,845	245	64
1821 Guadalajara	1760	1,438	311	47
1821 Guadalajara	1770	3,017	312	47
1821 Guadalajara	1780	4,975	295	51
1821 Guadalajara	1790	7,195	294	51
1823 Coahuila	1760	150	253	62
1823 Coahuila	1770	307	202	75
1823 Coahuila	1780	483	198	76
1823 Coahuila	1790	658	246	63
1930 Mexico	1870	514	278	55
1930 Mexico	1880	908	256	61
1930 Mexico	1890	1,432	243	64
1930 Mexico	1900	2,113	226	68

1950 Mexico	1890	1,147,619	239	65
1950 Mexico	1900	2,028,193	220	70
1950 Mexico	1910	2,855,705	194	77
1950 Mexico	1920	3,902,717	167	83
1700 Lima	1640	274	374	31
1700 Lima	1650	515	354	36
1700 Lima	1660	896	298	51
1700 Lima	1670	1,112	299	50
1866-1909 Lima prison	1820		163	84
1866-1909 Lima prison	1830		161	85
1866-1909 Lima prison	1840		161	85
1866-1909 Lima prison	1850		128	93
1866-1909 Lima prison	1860		115	96
1940 Peru	1880	287,100	208	73
1940 Peru	1890	441,660	194	76
1940 Peru	1900	698,569	172	82
1940 Peru	1910	942,837	148	88
1940 Lima	1880	36,162	139	90
1940 Lima	1890	62,904	132	92
1940 Lima	1900	104,355	127	93
1940 Lima	1910	149,334	119	95
1772 Montevideo	1710	119	261	60
1772 Montevideo	1720	151	265	59
1772 Montevideo	1730	371	236	66
1772 Montevideo	1740	721	232	67
1791 Montevideo	1740	52	260	60
1791 Montevideo	1750	112	281	55
1791 Montevideo	1760	132	248	63
1834-36 Soriano/Maldonado	1770	110	286	53
1834-36 Soriano/Maldonado	1780	245	218	70
1834-36 Soriano/Maldonado	1790	365	222	70
1834-36 Soriano/Maldonado	1800	446	257	61
1846 Montevideo prison	1790	130	231	67

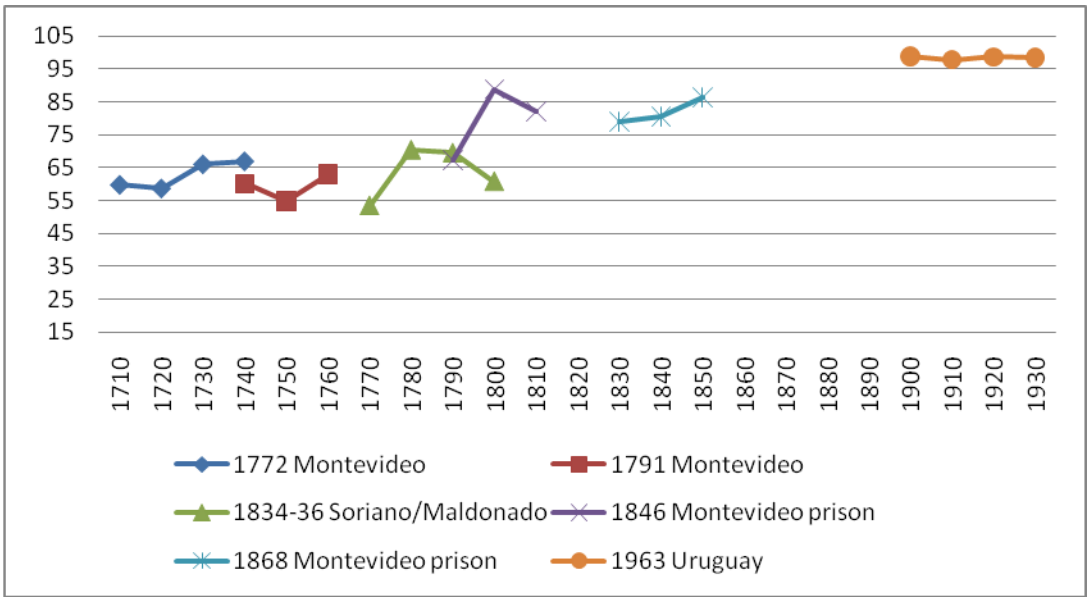
1846 Montevideo prison	1800	470	145	89
1846 Montevideo prison	1810	965	172	82
1868 Montevideo prison	1830	130	185	79
1868 Montevideo prison	1840	345	178	80
1868 Montevideo prison	1850	793	154	86
1963 Uruguay	1900	234,244	104	99
1963 Uruguay	1910	297,306	109	98
1963 Uruguay	1920	373,428	105	99
1963 Uruguay	1930	385,341	106	99

Peru - ABCC index of basic numeracy (vertical axis) by birth decades



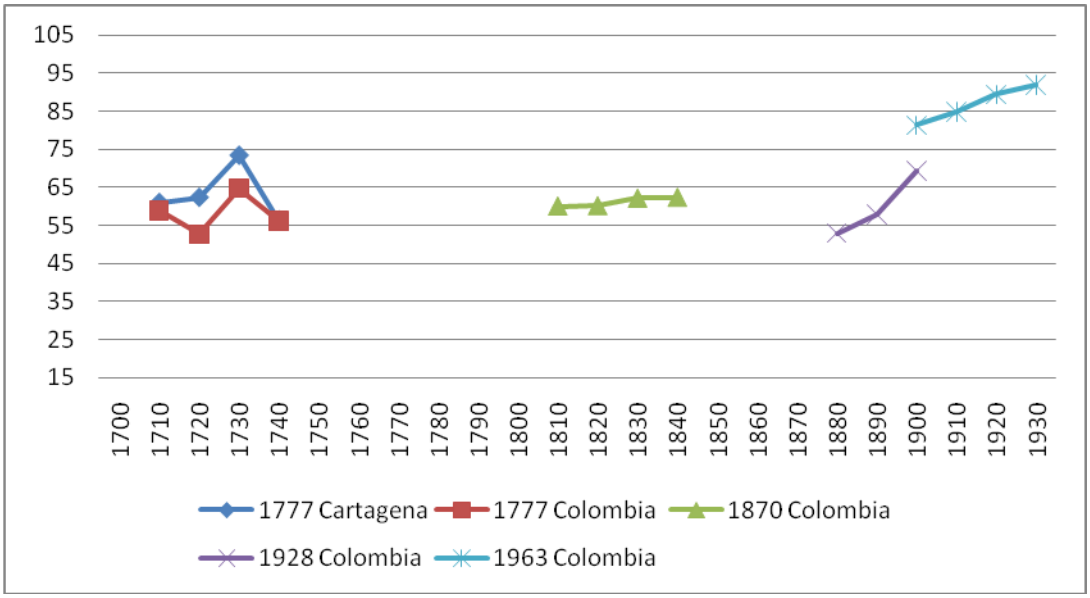
Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Uruguay - ABCC index of basic numeracy (vertical axis) by birth decades



Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

Colombia - ABCC index of basic numeracy (vertical axis) by birth decades



Note: The decades refer to birth decades (1800-1809, etc.). Sources: See text and Table 1.

4. GROWTH EFFECTS OF 19TH CENTURY MASS MIGRATIONS: "FOME ZERO" FOR BRAZIL?

We estimate a long-run trend of Brazilian human capital that extends back to the very beginning of the 18th century. With new data on selective immigration during the era of mass migrations at the end of the 19th century, we show that human capital endowment of international migrants can induce effects on economic development that persist until today. According to our estimations, the effect of selective immigration on real GDP per capita in the year 2000 is significant and equals around 75 US $ overall. As a reference, this value equals the amount poor Brazilians get to supplement their subsistence in the "*Fome Zero*" (Zero Hunger) program. We argue that human capital formation is a highly path-dependent and persistent process.

This chapter is based on a paper written jointly with Joerg Baten (University of Tuebingen) and Tarcísio Botelho (Universidade Federal de Minas Gerais). It is currently under review at the *European Review of Economic History*. The concept of the paper was developed jointly; analyzes and writing equally shared.

4.1 Introduction

19th century mass migrations drove some 55 million Europeans to the New World in search of a better life (Hatton and Williamson 1998). The magnitude of migratory flow has been unprecedented in history. This has been possible due to a conjuncture of events such as the transport cost revolution, a sharp rise in population growth in the old world and labor scarcity in the New World, as documented well in the literature (Hatton and Williamson 1998, Cohn 2009, Balderas and Greenwood 2010). Only recently, however, the skill selectivity of these migrants has shifted to the centre of attention to scholars (Stolz and Baten 2010, Cohn 2009, Viarengo and Murtin 2010). Evidence that allows measuring the impact of selective migration on destination country's economies is therefore a clear research desideratum. The human capital which migrants incorporate, their individual skill-set, is an important channel through which migration can have persistent effects on destination country's economic prosperity. Human capital includes formal education, but also entrepreneurial spirit and motivation. These capabilities and skills will not be applied in the home country's economy, but in the economy of the receiving nation. Selective migration can have persistent effects on the process of human capital accumulation of receiving nations and thus for economic growth in general. The direction of this effect depends very much on how immigrants compare to the destination country's population. In this context, research providing long-run data on human capital endowment of both migrants and native population is particularly important, as it sheds light on the long-term relationships, persistence and path-dependency of these processes.

This article provides new evidence on Brazilian human capital formation in the very long-run, from the early 18th to the 20th centuries. Furthermore, we collected new data on human capital endowment of 19th century migrants to Brazil, whose educational level was higher than that of native Brazilians. We use the age-heaping method (Mokyr 1983, A'Hearn, Baten and Crayen 2009), which provides a numeracy indicator particularly suitable for early stages of human capital formation. It exploits the tendency of less educated people to round their age instead of stating an exact age. Numeracy is highly correlated with other human capital output indicators such as literacy. Its main determinant is schooling investments (Crayen and Baten 2009). We argue that - apart from the migratory transfer of skills - the relatively high human capital levels of immigrants induced spillover effects to overall Brazilian human capital endowment which persistently enhanced economic growth in Brazil. Endogenous growth theory states a theoretical relationship between human capital development and economic growth (Lucas 1988, Romer 1994, amongst many others). Testing this relationship empirically is complex. Capolupo (2008) and others (for

example, Cohen and Soto 2007) show that the impact of human capital might have even been underestimated in previous research due to a lack of consistent and long-run data on human capital. In this study we are interested in the long term impact of human capital formation on economic performance. Does 19th century human capital development still impact growth performance today? In a cross-section of 91 countries, we show that numeracy levels in 1900 have a strong significant impact on real GDP per capita in the year 2000, even after controlling for a set of control variables. That means countries which already performed well in terms of human capital around the year 1900, still have a high real GDP per capita in the year 2000, whereas countries which performed badly in terms of numeracy around 1900, today are situated more at the bottom of the global income distribution. This exercise provides some evidence, that human capital endowment a hundred years ago still impacts growth performance today.

In a counterfactual, we show that the effect of migratory human capital transfer is also important economically. For illustration, consider that in the year 2002, 44 million Brazilians received between 50 and 95 Brazilian *Reais* per month to amend their subsistence with the "*Fome Zero*" program initiated by President *Lula da Silva*.[166] For illustration: the effect is larger than one of these payments, namely around 75 US $ or somewhat over 130 *Reais*. This would be the estimated magnitude of the persistent human capital effect on today's income that selective mass migrations around the turn of the last century still has.

When it comes to 19th century mass migrations, mainly the destinations of North America and Argentina have been at the centre of scholarly attention (Hatton and Williamson 1998). The Brazilian experience has somewhat been neglected, although it received over 4.1 million immigrants between its independence in 1822 and 1926, ranking after the US and Argentina among the top three most favorite immigration countries during the era of mass migrations (Ferenczi and Willcox, 1929, important exceptions are, for example, Klein 1995, Balderas and Greenwood 2010). Brazil was the last country in the Western World to abolish slavery, consequently the demand for immigrant labor on the plantations of the thriving coffee sector created an immigrant flow that was attracted to a large extent by favorable immigration policies of the Brazilian government which paid passages or granted access to land. Also owners of large coffee plantations were willing to pay the travel costs for immigrants and their families, who then worked off their debts on the plantations. For Brazil it was necessary to be financially attractive, because the country had to compete with Argentina and the US, whose economies provided higher income per capita

and hence higher wage gaps in relation to European source countries. This created stronger pull forces to these economies than to Brazil (Holloway 1980, Stolcke 1986).

The article is structured as follows: in the next chapter, we clarify the relation between human capital formation, selective migration and economic growth. We also propose possible channels via which human capital spillovers could have taken place. In the next section we review the literature on human capital formation in Brazil since the 18th century. Then, we introduce and discuss the data. The following section displays our regression results and the counterfactual. In a next step, we look at regional human capital distribution and immigration, and verify whether high absolute immigration in a state correlate with numeracy increases. We end with a conclusion.

4.2 Economic Growth, Human Capital Formation, and International Migration

Thirty years ago, endogenous growth theory has provided new ways of thinking about economic growth, by incorporating the source of economic growth into both sides of equations. Human capital plays a central role in endogenous growth theories. Generally, there exist three broad strands of literature. One strand of literature focuses on growth driven by capital stock, where capital is broadly defined to include human capital (Romer 1994), whereas Lucas (1988) and others have provided models that incorporate human capital as a separate factor that becomes the driving force for economic growth. A more recent strand of literature puts the process of knowledge creation via innovation and R&D into the centre of attention (Grossman & Helpman, 1991; for a comprehensive survey of the topic see Drinkwater et al 2003). All of these approaches have been applied to the field of international migration to answer the question of how migration affects economic growth in the long run. Central to all of these approaches is the assumption that migrants become a part of the destination country's human capital stock, which then affects economic performance. Empirical studies using one of these models commonly conclude that international migration can affect economic growth through the skill composition of migrants which they apply in the destinations' labor market but not in their source countries (Reichlin and Rustichini 1993, Walz 1995, Lundborg and Segerstrom 1999, 2000 as in Drinkwater et al 2003).[167] In destination countries, the question is whether migrants are high or low skilled

[166] The „*Fome Zero*" program was introduced 2002, by *President Luis Inácio Lula da Silva*. It assists ca. 44 million Brazilians with monthly payments between 50 and 95 Brazilian *Reais* to amend their nutrition expenditures. In Brazil, nowadays the share of people living from less than 1$ per day is well over 20 percent.

[167] For source countries, emigration of highly skilled individuals, the phenomenon of a brain drain, can be problematic, however the possibility of migration might also induce further investment in human capital and

and how they compare to natives.[168] Although migration is typically not included in growth regressions, there exists evidence from the field of economic history that international migration induces growth effects in receiving nations for the case of Argentina (Taylor 1997). Taylor uses a Computable General Equilibrium Model to capture the effects of migration on the Argentinean economy during the age of mass migration. He finds that immigration causes a significant increase in GDP. CGE models have the advantage that possible internal migration due to international immigration can be controlled for and multiplier effects can be captured. However, the outcome of the analysis is very sensitive to initial assumptions of a variety of model parameters that have to be chosen or estimated beforehand.

Recently, the Brazilian case of the state of São Paulo has been investigated by Carvalho Filho and Colistete (2010). Based on historical data from agricultural censuses and education statistics, these authors find a positive and enduring effect of the presence of foreign born immigrants on the supply of public instruction, as well as negative effects of land concentration.

In this study, we quantify the long-term impact of selective migration on human capital formation and economic growth in Brazil. Human capital formation is a highly persistent and path dependent process. Looking at numeracy levels in the year 1900 versus higher educational attainment of the total population provided by Barro and Lee (2001) (Figure 1), we find a strong positive relationship. Both are indicators of human capital, however, while numeracy captures a basic form of human capital, the latter is an indicator suitable for economies today, capturing a very advanced form of human capital, namely average years of higher schooling in the total population. It is evident that countries performing poorly in terms of human capital in the year 1900, like Egypt or Bangladesh, are still underperforming in the year 2000. On the other hand, countries like the US or Japan, which already had solved their numeracy problem around 1900, are among the top-performers in the year 2000 (Figure 2). The graph shows the path dependency of human capital formation in the very long run. Due to this path dependency, and the theoretical relation of economic growth and human capital, we would also expect the long-run relationship between early human capital formation and economic output to hold.

We first estimate the effect of human capital endowment in the year 1900 measured with age heaping techniques on log GDP per capita in the year 2000, controlling for a set

therefore affect factor endowment positively (Beine et al 2001, Stark et al. 1998) Brücker (2007) therefore concludes that effects on the source country's human capital endowment are ambiguous.

168 The question is whether the immigrants are substitutes or complements to the native population, because this determines if there will be a crowding out of natives on the labor market, or migrants' skill composition is complement natives so that growth effects can be realized (Borjas 1994).

of geographical and institutional variables.[169] In a next step, we calculate a counterfactual, assuming that the migrants' human capital had not arrived in Brazil, to assess whether there are any economically important long-term effects. Further, we look at regional dispersion of human capital and correlate this with immigration statistics, to find out, whether human capital enhancing effects took place mainly in the Brazilian states which were also predominantly destinations for the immigrants. This would give further support to the hypothesis that the human capital of migrants produced spillover effects to the overall Brazilian human capital formation.

Through which channels does the human capital of migrants affect economic growth? Migrants take their human capital endowment with them, which comprises their crude manpower, but also entrepreneurial spirit, innovative ideas and motivation, as well as formal and informal education and training. Migrants are not a random sample of the source country population (Todaro 1969, Williamson and Hatton 1998). Positive or negative selection in terms of formal education, relative to the source country's populations highly depends on skill premia or relative inequality of source and destination countries (Borjas 1994, Stolz and Baten 2010). It has been argued in the literature that the great majority of migrants however, are positively selected in terms of their willingness to take risks, entrepreneurial spirit and motivation. This leads to three channels, via which the migrants' human capital induces positive spillover effects to destination countries' human capital stock:

The first channel is a sheer level effect: 19th century immigrants exhibited a formal and informal education and training that was better than that of native Brazilians. The immigrants stemmed from societies that already had implemented school systems and provided compulsory formal education for most children (with the exception of Middle Eastern immigrants). They were also better trained for tasks in factories due to earlier industrialization in their home countries. Even though it is true that immigrants often were less educated than the people staying in Europe (Klein 1995), they often were more educated than native Brazilians (Merrick & Graham 1979), a fact that is supported by our data. Merrick and Graham (1979) state, in terms of literacy, that the average education level of immigrants was higher than that of native born Brazilians. They estimate that migrants had twice the literacy rate and three times the level of secondary and higher education than the

[169] Alternatively, we tried a set of world regional fixed effects, the results were similar.

native population.[170] The first channel might therefore be simply the higher education of migrants compared to the destination country's population.

The second channel is related to entrepreneurship. Campos Araújo and his co-authors (2006) stress analogies between the mindset of an entrepreneur and a migrant. The motivation to start a new life in a geographically and culturally remote target country is very similar to the willingness to risk and the ambition to start a new business. Also, immigrants are often driven to engage in entrepreneurial activities because of restrictive labor markets. Particularly in 19th and early 20th century Brazil, where there was no social security system whatsoever, immigrants had to start their own businesses to survive if they were not working on plantations or in manufacturing, or after they had paid off their debt of assisted passages to the plantation owners. Most of them successfully did so. For example, by 1934, over 40 percent of the coffee production in São Paulo was produced by the 14.5 percent foreign population of the state, which shows that, as Baer (2001) states, the immigrants were "economically ambitious people". Additionally, if immigrants wanted to maintain their cultural identity and lifestyle and consume the services and products they were accustomed to, they had to produce them, provide them or initiate trade with the home country to obtain them. This channel would induce trade, craftsmanship and employment in both immigrant communities and among the native population.

A third possible channel could be human capital spillovers from migrant human capital to the local population. The higher education of immigrants could spill over to the native population as they initiated successful behavior. Immigrants built their own schools, hospitals and charitable institutions, in Rio de Janeiro and São Paulo (Klein 1995). These institutions were channels through which the native population might have enjoyed greater access to education and social care. Schwartzman and Brock (2003) also state that German, Italian, and Japanese immigrants created their own schools, sometimes with the support of the governments of their countries of origin, or foreign priests.

In the Brazilian case, one has to bear in mind that the majority of the migrants arrived in the great cities of *Rio de Janeiro* and *São Paulo* via the ports where they disembarked the steamships after their journey across the Atlantic. They mainly went to the center and the southern part of Brazil (Klein 1995). Hence, human capital spillovers would have predominantly taken place in and around these areas with high immigrant concentra-

[170] However, Fritscher, Musacchio and Viarengo (2009) argue that the increase in human capital in the period 1890-1940 cannot be explained by European immigration, because these migrants were among the least educated in Europe. Still, they do not provide data on immigrant selectivity. In our data, selectivity shows a mixed picture. Whereas migrants from Latin America, southern European countries and Syria were mainly positively selected in terms of numeracy from the source country population, US-Americans and Western Europeans were less numerate than their compatriots at home.

tion, which is particularly strong in the São Paulo region due to the rise in the coffee industry. After 1880 around 70 percent of all immigrants arrived here (Merrick & Graham 1979). This uneven distribution of immigrant human capital spillovers might have contributed, at least partly, to the regional inequality in human capital that persists today. In 1924, the *Associação Brasileira de Educação* was established in Rio de Janeiro with members from different nations and played a very important role in bringing education on the national agenda (Paim 1981 as in Schwartzman and Brock 2003). Later human capital formation has, due to the country's colonial legacy, a long history, which we will now study in detail.

4.3 Human Capital Formation in Brazil 1700-1940: what do we know?

Baer (2001) argues that there was virtually no formal education in Brazil prior to the late 18th century, except for some regionally scattered efforts of Jesuits. They were expulsed from the country in 1759, after conflicts with the new minister *Marquis of Pombal*. The Portuguese crown was in need of resources to reconstruct Lisbon after the devastating earthquake in 1755, which is why it confiscated the assets of the fraternity (Oliveira Marques, 1986). Even later, the few existing schools had little impact on the cultural level of the population as a whole, because the schools were understaffed and the teachers poorly paid (Prado Júnior, 1959). Murilo de Carvalho (1982) stresses that Portugal, as a colonial power, refused systematically to allow the organization of any institution of higher learning in its colonies. In 1768, he further mentions the *Concelho Ultramarino* which was the administrative institution which dealt with overseas issues, denied a request of the captaincy *Minas Gerais* to build a school of medicine. Only after 1808, when the Portuguese court fled to *Rio de Janeiro* due to the Napoleonic invasion in Portugal, two medical and two military academies were allowed in Brazil. But not only was higher learning severely underdeveloped, also, and more importantly, mass education was lacking. Schwartzman and Brock (2003) state that by 1794 only 179 appointed professional teachers existed in Brazil, whereas in the small country of Portugal the number was around 748. Fritscher, Musacchio and Viarengo (2009) repeat that around a century later, in 1890, Brazil still had one of the lowest levels of literacy in Latin America. Only around 15 percent of the population were able to read and write. Schwartzman and Brock (2003) also argue that throughout the First Republic (1889-1930), primary and secondary education remained the responsibility of local and state governments, and only about 25% of the population, at most, were literate. Large problems in this context were regional disparities in education. In 1834, the responsibility for education was assigned to the provincial governments and for

several years the central government subsidized poorer provinces. However, in 1845 these subsidies were abolished and some provinces were unable to provide an adequate schooling system. As a consequence, the number of children enrolled in public schools declined (Barman 1994, p. 242). These financial difficulties fostered regional disparities within Brazil even more. Most of the improvement in education during the late 19th century took place in Rio de Janeiro and São Paulo or the more prosperous southern regions while the northeast fell behind (Lewis 2006, p.125). These regions are also the ones that received most immigrants during the turn of the centuries. Fritscher et al. (2009) find that educational development was associated with trade and commerce, as states that received export tax revenues invested them in schooling and education. However, they do not account for immigration at the state level. We find numeracy improvements exactly in those states to be associated with knowledge spillover effects of international immigration. Given the fact, that between 1772 and 1900 almost one fourth of all population growth is due to immigration, it is quite plausible that human capital formation in Brazil was influenced by immigrant's human capital. Baer (2001) even argues that public investment in immigration had been a sensible, potent substitute for investing in education.

4.4 Data

We use the age-heaping approach to proxy early human capital formation in Brazil. Age-heaping exploits the fact that people living in societies at early stage of human capital formation typically tend to round their age to convenient multiples of five instead of reporting their exact age (see Crayen and Baten 2009). The share of multiples of five to other numbers in the age distribution is expressed in the Whipple Index (A'Hearn, Baten, Crayen 2009). In this study, we apply the ABCC-Index, which is a linear transformation of the latter. It ranges between 0 and 100, where zero stands for an age distribution where everybody states a rounded age and 100 represents no heaping at all. Its coherent construction allows comparability over space and time and different data sources (Crayen and Baten 2009, Manzel, Baten and Stolz 2010). In recent studies, the indicator has proven very valuable in a wide range of topics in human capital research (see, for example Leunig and Humphries 2009, Stolz and Baten 2010, Prayon and Baten, 2010, for details, also see the internet appendix on the methodology age heaping). For a society at that early stage of human capital formation, the age heaping approach is a suitable indicator to proxy human capital. Usually, societies reach higher numeracy values before they reach better levels of literacy, as knowing one's exact age is a basic skill compared to reading and writing. One

of the major determinants of numeracy is, of course, schooling investments as a major input to mass education (Crayen & Baten 2009).

4.5 Brazilian Numeracy

To estimate the Brazilian numeracy trend, we use census data from five consecutive censuses, 1772, 1830, 1890, 1920 and 1950. For the first time, it is possible to extend the trend back to the very beginning of the 18th century, a period where quantitative data has been scarce so far. One census executed in the region of Sorocaba, São Paulo in 1772 provides data for the birth cohorts 1710-1740. This is the earliest enumeration of the region analyzed in a quantitative study, so far. It has survived in the *Ultramarino* Historical Archive of Lisbon. Almost the entire São Paulo region was enumerated. We took a sample of 6,278 observations from Sorocaba for this study. The census contains information at the household level. All household members were asked for their names and ages, and there is information if they actually belonged to the family or were "*agregados*". This word refers to people who were like servants in pre-industrial rural Northern Europe. They were free, lived in the master's household, performed a wide range of services, and received board and lodging, and sometimes also some money. Furthermore, the census provides information on the possessions of a household. Unfortunately, there is no information on age statements of slaves or ethnic background of the free population, so we do not know anything about the ethnic composition of the sample. However, knowing that only free people are included in this enumeration, we had to adjust it, to secure comparability with the other sources, which all contain data on free and enslaved people. To this end, we looked at the Sorocaba district in the later, 1830 evidence, containing data on slaves and free people and we calculated the difference of slave and free ABCC levels, weighted by the number of observations. With this value, we corrected the Sorocaba 1772 evidence to make it comparable to our other sources.

Next, the 1830 evidence refers to enumerations in the states of *Paraná*, *São Paulo* and *Minas Gerais*. For Minas Gerais, there are two very important series of nominative lists: one for 1831-32 and another one for 1838-40. These sets of documents were processed by researchers of the Center of Regional Development and Planning, of the Federal University of Minas Gerais (CEDEPLAR/FACE/UFMG), who codified and keyboarded all information.[171] They are the result of attempts at producing general censuses in the prov-

[171] This work was developed by the team coordinated by researchers Clotilde Andrade Paiva, Roberto Borges Martins and Maria do Carmo Salazar Martins, who kindly allowed me to use their database. Some results

inces and were organized by justices of the peace (the lowest judicial authority in post-colonial Brazil), and were commissioned by the Provincial Government. The lists of 1831-32 were used because they comprehend a larger share of the population of Minas Gerais at the time. They contain information about 234 out of the 410 districts (the territory under the jurisdiction of a justice of the peace) of the province, covering 57% of all localities. The data base contains 491,017 inhabitants, 356,267 (72.6%) of them were free and 134,750 (27.4%) were slaves, corresponding to approximately 71% of the estimated total population of the province for the first half of the 1830 decade. The nominative lists of inhabitants of São Paulo (that included the present state of Paraná) are part of a more substantial set of documents called *Maços de População*. These lists give us a large volume of information on the population of the province. The lists of 1836, which we use here, are the last link of a chain of assessments of data on the population, which started to be compiled in 1765. All the lists of the 38 municipalities of São Paulo have been assessed. The data base contains 257,751 inhabitants, 181,849 (70.6%) of them being free and 75,902 (29.4%) being slaves, corresponding to around 79% of the total population of the province. It contains regional data on both sexes and also on their status, i.e. slaves and free people, which is why we consider it to be representative at least for the southern areas of Brazil.

The later samples of the 1890 and 1950 nationwide censuses are representative of the entire Brazilian population per definition. Only the 1920 census might contain an upward bias due to enumerating strictly the district capitals and not the rural areas of the districts.[172]

To which degree are the early samples biased? Before the coffee boom and the large migration waves set in after the 1870s, the Southeast was not a region with a higher human capital compared to the Brazilian average (except the urban Federal District of Rio de Janeiro), unlike the South. We can compare the regional numeracy for the birth decade of the 1830s. In this decade, the ABCC indexes in São Paolo and Minas Gerais were some 3 percent below the national average, whereas the one on Paraná was some 3 percent above the national average. Assuming similar regional differences for the inhabitants in the 1830 population lists, the slight negative regional bias cannot be very pronounced (around minus 1 percent). For the 1772 census, in which only São Paolo is represented, it might be of the order of a 3 percent negative bias (i.e., the Brazilian national figure might be even higher). If we adjust the 1920 census which contains only provincial capital, downward by 4 per-

reached with the use of this material can be found in Paiva (1986, 1996), Paiva and Arnaut (1990), Rodarte (1999, 2008), Andrade (2001).

[172] All of the published volumes of these censuses can be found in the digital library of the Brazilian bureau of statistics (http://biblioteca.ibge.gov.br/).

cent, the values for the birth decades of the 1860s and 1890s would also be identical to the respective birth decades in the censuses of 1890 and 1950.

Our numeracy estimates are arranged by birth decade averages and allow us to shed light on human capital formation in Brazil from 1710 onwards. We use birth decades because basic numeracy is normally acquired during the first decade of life (Crayen and Baten, 2009). We only use the age ranges of the 23-62 year olds, to avoid age effects caused by selective mortality of the older age groups. Hence, every census provides us with four decades of numeracy values. We start to observe Brazilian numeracy in 1710 at a level of 60 ABCC points (Figure 3). This is similar to the value of Hungary at that time and better than other Latin American countries like Argentina or Mexico (Manzel, Baten and Stolz 2010). Between 1700 and 1780, the level of heaping stays at around the same magnitude, suggesting stagnation during those decades. The level of the 1830 birth cohort is significantly lower than the peak in 1790, suggesting a deterioration of human capital formation during the Napoleonic Wars and the following independence of Brazil in spite of being a quite peaceful event, by Latin American standards. However this stagnation in numeracy trends can also be observed in other Latin American countries (Manzel, Baten and Stolz 2010). From the 1840s onwards, Brazil started a sustained upward trend, which accelerated between 1870 and 1910, a period with very high immigration rates.[173]

4.6 Immigrant numeracy

The immigrant data was preserved in the National Archive of Rio de Janeiro. The archive contains passenger lists of ships that arrived in Brazil from the beginning of the 19th century. However, until 1822, Portuguese immigrants were not explicitly declared as Portuguese; because Brazil was still part of the Portuguese empire and the Portuguese king even reside there after Napoleon invaded Portugal in 1808. We took samples of various arrival years between 1812 and 1932. The largest part of the sample stems from the period between the 1870s to the 1920s. This is the period in which mass migration to Brazil really took off. The immigrants of our sample were recorded at the ports in *Rio de Janeiro* or *Santos*, which were the most important ports. According to Ferenczi and Willcox (1929), around 85 percent of all immigrants entered Brazil via those ports. Italians were the largest immigrant group, followed by the Portuguese, who have an exceptionally long migration

[173] The geographical, gender and ethnic distribution of human capital was, however, extremely unequal in Brazil. Colonial legacy had produced a society with a very unequal distribution of income and human capital (Engerman & Sokoloff 2000, Fritscher et al. 2009). Thus, one has to keep in mind that numeracy levels for slaves or women or people from rural areas are far lower than those of city dwellers and white or upper class men.

history with Brazil due to the colonial era. Third was Spain, followed by German speaking immigrants and migrants from Eastern European countries. There was also some immigration from Middle Eastern countries during the increasingly turbulent years before and after the fall of the Ottoman Empire (Klein, 1995). Our data set reflects this distribution quite well (Figure 4). Portuguese, Italians and Spaniards make up the biggest immigrant groups, followed by Germans and other central European nations. The Portuguese outnumber the Italians in our sample. In addition, the port and year of arrival, origin, ship and age were given. Numeracy estimates for immigrants are also arranged by birth cohorts. They show that except for Lebanese migrants, all immigrant groups are more numerate than native Brazilian birth cohorts (Figure 5).

4.7 How path dependent is human capital formation?

Our strategy to assess possible human capital spillovers econometrically follows a two-step approach. First, we investigate the impact of 1900 human capital on real GDP 2000 in a cross-country setting. This tests our hypothesis of a long-run path dependency of human capital formation on economic performance of an economy as a whole. After verifying the relationship, we calculate a counterfactual in which we simulate a human capital situation of Brazil in 1900 without the immigrant‘s human capital, allowing us to compare the migration versus non-migration scenario. In a second step, we run a regional cross-section of immigrant share increase across Brazilian states on ABCC increase. The hypothesis is that the immigrant’s share is positively correlated with ABCC increase after controlling for other influences such as educational expenditures at the state level. To this end, we use data from the period of 1890 to 1920, the time in which Brazilian mass immigration reached its peak.

In the first step, our argument is that economic growth is ceteris paribus determined by current human capital stock, which in turn is determined by human capital in the past. This argument points to the direction of the work of Acemoglu, Johnson and Robinson (2001) that stresses the importance of past events on economic performance today, although they focus on institutions rather than human capital. Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004), disagree on the fact that institutions should be at the focus of the analysis and point to the importance of human capital, being a “more basic source of growth”, which in turn shaped institutions and brought about the path of dependency for economic development. In a more recent study, Prayon and Baten (2010) test this relationship with an extended data base for human capital. They find that it is indeed the human capital channel that enhances institutions which affect economic output in the very long

run. Figure 6 shows the long-run relationship between real GDP per capita and numeracy performance in the year 1900. With the exception of some oil-exporting countries, we find a stable linear relationship. Those, who performed well in terms of human capital at the beginning of the twentieth century, are among the top-performers a century later.

What other variables should be of interest? Aggregate output is a function of physical capital, labor and human capital at a given technological level. Labor can be ignored, as we look at per capita values of GDP. What about physical capital? Baier, Dwyer and Tamura (2006) provide data on the physical capital endowment for a broad range of countries. Their work suggests that the levels of physical capital across world regions varied considerably in 1900.[174] Engerman and Sokoloff (2001) argue that geography and factor endowment play a crucial role in determining development outcomes across countries. Therefore, we include a set of geographical and institutional variables, such as a variable for the political system, civil war, longitude, latitude and a dummy variable for landlocked geographical situation.[175]

We regress log GDP per capita on ABCC values for 91 countries in the year 1900 (Table 1).[176] The numeracy level in 1900 is highly significant. An increase in numeracy in 1900 by one ABCC point leads to a two percent increase in GDP per capita in the year 2000. Institutional aspects and geography render significant results, as well. We explain almost 60 percent of the inter-country variance in real GDP with our model. There is an unexplained part, as, of course, today's economic policies and factor endowments should also play a crucial role in determining real GDP per capita today. Yet, we find considerable evidence for path dependency of human capital formation. This result implies that the countries, which had resolved their numeracy problem by 1900 like Norway or the US, are the states that are doing well in terms of economic performance in the year 2000, as well. Cases with higher GDP than expected are states like Indonesia, India, Kuwait, Bahrain, Egypt and Morocco.

How well does our estimation perform in predicting real GDP per capita in the year 2000 in Brazil? If we consult the Penn World Tables, we find that our estimation is quite

[174] In a regression of GDP 2000 on physical and human capital in the year 1900, which is possible only for a number of 23 countries, both determinants are, of course, highly significant. We tested this with the human capital indicator incorporated in the Baier et al. (2006) data set and also with our numeracy estimates.

[175] Geographical data is taken from www.cepiis.fr, political and institutional variables are from the PolityIV project.

[176] In the version with the world regional fixed effects, the only world region which displays a coefficient statistically different from zero is Sub-Saharan Africa. This result goes in line with previous studies from acknowledged scholars on economic growth which all detected a somewhat peculiar negative growth pattern for African countries (i.e. Sachs & Warner 1997, Englebert 2000).

close to the actual value of Brazil's real GDP per capita in the year 2000. Whereas we would predict 7584.65 US$, the PWTs display an actual value of 7455.9 US$.

An important issue is endogeneity. It is quite likely that wealthy economies can afford better educational systems and therefore achieve a more rapid human capital accumulation. However, the lagging of the human capital variable for one century should deal with this issue. The level of early human capital does not reflect other characteristics that are important for income today, which is why it should be uncorrelated with the error term.

In a next step, we calculate a counter-factual to answer the question whether the human capital incorporated in Brazil's immigrants had any long-term effects on Brazilian long-run economic performance. In 1900, the Brazilian population consisted of 17.4 million people. According to Klein (1995), seven percent of them were immigrants. The ABCC for the 1900 birth cohort of the whole population, including migrants is 89.7 percent. From our data, we are able to calculate the total average ABCC for immigrants, born in 1900, which yields 94.1 percent. This means that the average immigrant numeracy was around five percentage points higher than numeracy levels for the entire population. With these figures, we can calculate the numeracy value for native Brazilians only, conditional on their population share, which renders 89.32. Filling this into our estimated relationship between ABCCs in 1900 and real GDP per capita in the year 2000, we find that average real income per capita would be around 75 US $ less if the migrants had not come to Brazil and brought their numeracy skills with them. 75 US $ look like a small amount at first glance. How much is that worth in Brazil? 75 US $ equaled 134 Brazilian *Reais* in the Year 2000.[177] For comparison: the minimum wage was around 260 *Reais*, so this amount would have been roughly one half of a minimum wage. Or put differently, this equals more than the amount that ca. 44 million Brazilians receive with the "*Fome Zero*" program, which was introduced in 2002 by the Brazilian president *Inacio Lula da Silva*. It is aimed at the very poor of the Brazilian society and provides them with money or coupons to buy foodstuffs. Hence, this money would be a quite significant share of real income to many Brazilians, even today. Moreover, this is a lower bound estimate, because first, before 1900, the gap between natives and immigrants was larger. In the early 19th century, immigrants had an ABCC advantage over natives of around 20 percent (Figure 5b). Second, we measured only the level effect of one generation of migrants. Third, we do not capture any spillover effects to native numeracy in this regression.

[177] http://www.brasilien.li/artikel/wechselkurs-real-euro-1980, last accessed March 24, 2011.

4.8 Numeracy spillovers on a regional Level

How do we measure human capital spillovers? In this case, we look at the regional distribution of immigration and analyze whether the immigrant increase that Brazilian states experienced between 1890 and 1920 is correlated with the regional ABCC-increases during this period. Using data from the 1890 and 1920 censuses of Brazil, we generate a regional distribution of the evolution in numeracy between those two points in time (Figure 7).[178] This pattern suggests that immigration had something to do with the evolution of human capital in Brazil at the regional level. The biggest ABCC increases we find are in the states that were the main destinations for immigrants like Rio de Janeiro, São Paulo and Minas Gerais. In contrast, states with low immigration increase had also low ABC increase, such as Rio Grande do Norte and Piauí. We now analyze whether the ABCC increases are correlated with regional immigrant increase over the respective period. To control for educational expenditures, we use the data provided by Martinéz-Fritscher, Musacchio, Viarengo (2009). These authors provide a very useful data base on a regional level for Brazil around the turn of the century. We use their data on education expenditure. Martinéz-Fritscher et al. argue that an improvement in education was predominantly possible in states with higher income from commodity exports. They further state that immigration could not have played a role, because the immigrants were poorly educated and the great improvement in education took place only after the bulk of immigrants already had entered Brazil. We do not argue that their main argument of commodity export taxes improving state educational expenditures does not hold, but we argue that states which generated such income were also states that experienced significant immigration, because export economies were the labor markets targeted by the immigrants. Moreover, the immigrant communities themselves may have spread their knowledge in charitable institutions or might simply have also demanded more schools, because they were accustomed to governments that provided some education. Hence, we argue that in states that were open and integrated into the world market because they had a valuable commodity to trade, migration also played an important part in boosting education. Both, trade and migration take place at the same time, because they are both characteristics of open economies. This is also shown by a crude correlation between the growth rate of education expenditures between 1890-1920 and immigrant increase across states, which is as high as 0.55 (significant at the one percent level, p-value: 0.015). In our regressions, we run the ABCC increase between 1890

[178] The ABCC levels in 1890 are smallest for north-eastern Brazilian states like *Alagoas* or *Sergipe*. *Rio Grande do Sul* and *Santa Catarina* display very high numeracy values in 1890. *Rio de Janeiro* and *São Paulo* are situated somewhere in the middle. In 1920, however, these primary immigrant destinations display the highest numeracy values, together with the southern states of *Paraná* and *Rio Grande do Sul*

and 1920 and between 1890 and 1950 on the log of the absolute immigrant increase per state (Table 2 and 3). Additionally, we control for log population density, because more densely populated regions might also have a better infrastructure to provide education to their citizens. Furthermore, we control for initial ABCC levels, because the more numerate the people in the state already are, the less room might be available for improvement, since the ABCC index is an indicator of basic human capital. Of course, state educational expenditures are also important in this context. We control for this in two different ways, first, we take average state educational expenditures between our two time points and in another specification, we only control for the initial value of state expenditures in 1890, as the increase in this variable is correlated with the increase in immigrants throughout states. In our third regression, we additionally included a set of dummy variables that control for certain export commodities like coffee, cattle and cotton or mate.

We ran three different specifications, where we varied the measures of educational expenditures of the states and additionally included the export industry dummy variables. Throughout all specifications, the log of immigrant increase in the states has a significant effect of substantial and robust magnitude. If we calculate standardized beta coefficients, the effect is the most important in the first two specifications.[179] None of the commodity expenditure dummies becomes significant. The log of population density is only significant in the first two specifications, whereas the initial ABCC level is always significantly and negatively correlated with the ABCC increase of the respective state, a result we would have expected. To test the robustness of these results, we ran the same regression on the ABCC increase between 1890 and 1950 across states (Table 3). The results stay robust; immigrants seem to have produced an impact on human capital development across Brazilian states.

4.9 Conclusion

In this article, we studied the long-term human capital formation in Brazil. We constructed a numeracy trend for Brazil that extends back to the beginning of the 18th century. With a new sample on human capital endowment of 19th century mass migrants, we showed that international migration can produce human capital enhancing effects that in turn affect economic performance until today. Looking at regional data for Brazil, we showed that human capital spillovers took place predominantly in those states that received the most

[179] If we include the set of export dummy variables, the beta coefficient of average education expenditure becomes larger than the one on immigrants.

migration in absolute numbers. Increases in numeracy are significantly positively correlated with immigrant increases, even after controlling for educational expenditures on the state level. This suggests that human capital grew strongest in those states where most immigrants arrived. Could there be crowding-out effects in a way that Brazilians do not have education incentives, because the education-intensive jobs are increasingly filled with immigrants? Human capital intensive migration in contrast might stimulate firm creation and also provide teachers and institutions to provide education even for native Brazilians. Secondly, in today's world, migration hurdles often encourage to education investments, whereas migration hurdles were small in the 19th century world.

Hence, human capital in its various dimensions incorporated in international migrants can affect human capital endowment, which in turn determines economic development and performance of destination countries. This relationship is crucial in understanding today's human capital selective immigration policies of major immigrant countries like the US, Canada, or Australia.

4.10 References

Acemoglu, D., Johnson and S., Robinson, (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation" *The American Economic Review* 91, no. 5, pp. 1369-1401.

Andrade, Cristiana Viegas de (2001): *Domicílios mineiros oitocentistas: uma aplicação do método Grade of Membership (GoM)*. Belo Horizonte: UFMG, (Unpublished MA. Dissertation).

A'Hearn, B., Baten, J., and D. Crayen (2010): "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital" *Journal of Economic History* 47, no. 1, pp. 82-99.

Baer, W. (2001): *The Brazilian Economy: Growth and Development*. Sixth edition. Boulder.

Baier, S.L., Dwyer, G.P., and Tamura, R. (2006): "How important are Capital and Total Factor Productivity for Economic Growth?" *Economic Inquiry* 44, no. 1, pp. 23-49.

Balderas, J.U., and Greenwood M.J. (2010): "From Europe to the Americas: a comparative Panel-data analysis of migration to Argentina" *Journal of Population Economics* 23, pp. 1301-1318.

Barman, R.J. (1994): Brazil: The Forging of a Nation, 1798-1852. Stanford, Stanford University Press.

Barro, R.J. and Lee, J.W. (2001): "International data on educational attainment: updates and implications" *Oxford Economic Papers*. Oxford University Press. http://oep.oxfordjournals.org/content/53/3/541.short, last accessed, May 2nd, 2011.

Borjas, George .J. (1994): "The Economics of Immigration" *Journal of Economic Literature* 32, no. 4, pp. 1667-1717.

Campos Araújo, J.R. de; Cruz Paiva, O.da and Rodriguez, C.L. (2006): "Immigrants and Entrepreneurs in São Paulo, Brazil: Economic Development in the Brazilian 'Melting Pot'", In: Galbraith, C.S. and Stiles, C.H. (eds): *Developmental Entrepreneurship: Adversity, Risk and Isolation. International Research in the Business Disciplines*, Oxford, Vol. 5, pp. 297-322.

Capolupo, Rosa (2009): "The New Growth Theories and Their Empirics after Twenty Years", *Economics: The Open-Access, Open-Assessment E-Journal, Vol. 3*.

Carvalho Filho, I., and Colistete, R.P. (2010): "Education Performance: Was it all determined 100 years Ago? Evidence from São Paulo, Brazil." http://mpra.ub.uni-muenchen.de/24494/MPRA Paper No. 24494.

Cohen, Daniel and Soto, Marcelo (2007): "Growth and Human Capital: Good Data, Good Results" *Journal of Economic Growth* 12, pp. 51-76.

Cohn, Ray L. (2010): *Mass Migration under Sail. European Immigration to the Antebellum United States*. Cambridge University Press, Cambridge.

Crayen, D., Baten, J. (2009): "Global Trends in Numeracy 1820–1949 and its Implications for Long-term Growth", Explorations in Economic History Vol. 47, No. 1, pp. 82-99.

Drinkwater, S., Levine, P., Lotti, E. and Pearlman, J. (2003): "The Economic Impact of Migration: A Survey", Hamburg Institute of International Economics FLOWENLA Discussion Paper No 8. ISSN 1616-4814.

Engerman, Sokoloff (2000): "History Lessons. Institutions, Factor Endowments, and Paths of Development in the New World" *Journal of Economic Perspectives* 14, no. 3, pp. 217-232.

Englebert, P. (2000): "Solving the Mistery of the AFRICA Dummy" *World Development* 28, no. 10, pp. 1821-1835.

Ferenczi & Willcox (1929): *International Migrations. Volume I. Statistics*. In: Easterlin, Perlman and Thomas (eds): Demographic Monographs. A series of economics reprints. New York.

Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (2004): "Do institutions cause growth?" *Journal of Economic Growth* 9, no. 3, pp. 271-303.

Grossman, G.M., and Helpman, E. (1991): "Quality Ladders in the Theory of Growth" *Review of Economic Studies* 58, pp. 43-61.

Hatton, Tim J. and Williamson, Jeffrey G. (1998): *The Age of Mass Migration: Causes and Economic Impact*. Oxford University Press. New York.

Hatton, Tim J. and Williamson, Jeffrey G. (2008): *Global Migration and the World Economy: two Centuries of Policy and Performance*. MIT Press. Michigan.

Holloway, Thomas H. (1980): *Immigrants on the land: coffee and society in São Paulo, 1886-1934*. Chapel Hill, NC: University of North Carolina Press.

Humphries, Jane and Leunig, Tim (2009): "Was Dick Whittington Taller than those left behind? Anthropometric measures, migration and the quality of life in early 19th century London" *Explorations in Economic History* 46, no. 1, pp. 120-131.

Klein (1995): "European and Asian Migration to Brazil", In: Cohen (ed): *The Cambridge Survey of World Migration.* Cambridge University Press.

Lewis, W.A. (2006): *Tropical Development, 1880-1913*. London and New York, Taylor and Francis.

Lucas, R.E., jr. (1988): "On the Mechanics Economic Development" *Journal of Monetary Economics* 22, pp. 3-42.

Manzel, K.; Baten, J., Stolz, Y. (2010): "Convergence and Divergence of Numeracy: The Development of Age Heaping in Latin America, 17th to 20th Centuries" *Economic History Review*, forthcoming.

Martinéz-Fritscher, Musacchio, Viarengo (2009): "The Great Leap Forward in Education in Brazil: The Political Economy of Education in Brazil, 1890-1940" Harvard Business School Working Paper 10-075.

Murtin, Fabrice and Viarengo, Martina (2010): "American Education in the age of mass migrations 1870-1930" *Cliometrica* 4, pp. 113–139.

Merrick, T.W. and Graham, D.H. (1979): *Population and Development in Brazil 1800 to the Present*. Johns Hopkins University Press. Baltimore and London.

Messias, Erick (2003): "Income Inequality, Illiteracy Rate, and Life Expectancy in Brazil" *American Journal of Public Health* 93, no. 8, pp. 1294-1296.

Mokyr, J. (1983): *Why Ireland Starved*. London: George Allen & Unwin.

Murilo de Carvalho (1982): "Political elites and States Building. The Case of 19th Century Brazil" *Comparative Studies in Society and History* 24, no. 3, pp. 378-399.

Oliveira Marques, AH (1986): *História de Portugal*. Lisboa.

Paiva, Clotilde Andrade (1986): "Minas Gerais no século XIX: aspectos demográficos de alguns núcleos populacionais", in: COSTA, I. D. N. da (org.). *Brasil: história econômica e demográfica*. São Paulo: IPE/USP, pp. 173-187.

Paiva, Clotilde Andrade (1996): População e economia nas Minas Gerais do século XIX. São Paulo: USP, (Unpublished PhD. Dissertation).

Paiva, C. A., ARNAUT, L. D. H. (1990): Fontes para o estudo de Minas oitocentista: listas nominativas, V Seminário sobre a Economia Mineira, Anais, Belo Horizonte: CEDEPLAR/FACE/UFMG, pp. 85-106.

Prado Júnior, C. (1959): *História Econômica do Brasil*. São Paulo. Fifth edition.

Prayon, V. and Baten, J. (2010): Human capital, institutions, settler mortality, and economic growth in Africa, Asia and the Americas. University of Tuebingen Working Paper.

Rodarte, Mário Marcos S. (2008): O trabalho do fogo: Perfis de domicílios enquanto unidades de produção e reprodução na Minas Gerais. Belo Horizonte: UFMG, (Unpublished PhD. Dissertation).

Romer, P.M. (1994): "On the Origins of Endogenous Growth" *Journal of Economic Perspectives* 8, no. 1, pp.3-22.

Sachs, J.D. and Warner, A.M. (1997): "Fundamental Sources of Long-Run Growth" *The American Economic Review* 87, no. 2, pp. 184-188.

Schwarzman, S.; Brook, C. (2003): *The challenges of Education in Brazil*. Oxford University, Centre for Brazilian Studies.

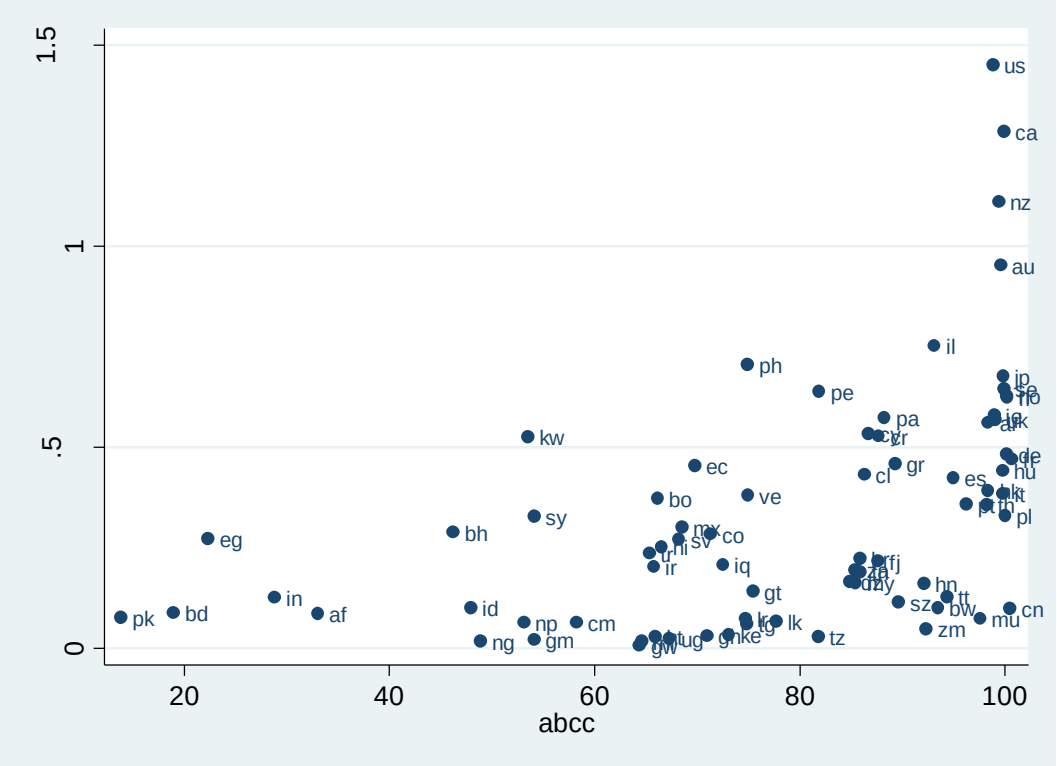
Stolcke, Verena (1988): *Coffee planters, workers and wives: class conflict and gender relations on São Paulo, 1850-1980*. Oxford: Macmillan, St. Antony's College Oxford.

Taylor, A. (1997): "Peopling the Pampa: On the impact of mass migration to the River Plate, 1870-1914" *Explorations in Economic History* 34, pp. 100-132.

Todaro, M.P. (1969): "A Model of Labor Migration and Urban Unemployment in Less Developed Countries" *The American Economic Review* 59, no. 1, pp. 138-148.

4.11 Figures

Figure 1: Numeracy 1900 versus higher educational attainment to total population in the year 2000 (Barro and Lee 2000, Prayon and Baten 2010, Crayen and Baten 2009)



Notes:

All country abbreviations are ISO 3166, (http://www.metatab.de/meta_tags/laenderkuerzel.htm)

Figure 2: Higher educational attainment 2000 (Barro and Lee, 2000) as a major determinant of log GDP per capita 2000 (Penn World Tables)

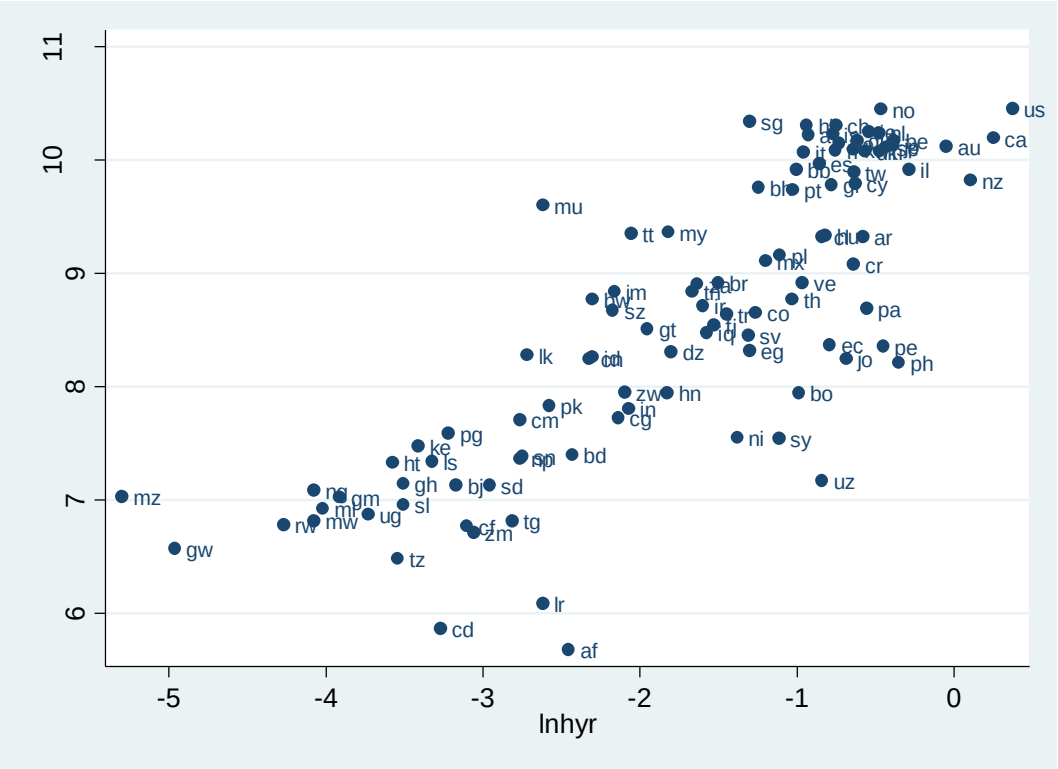
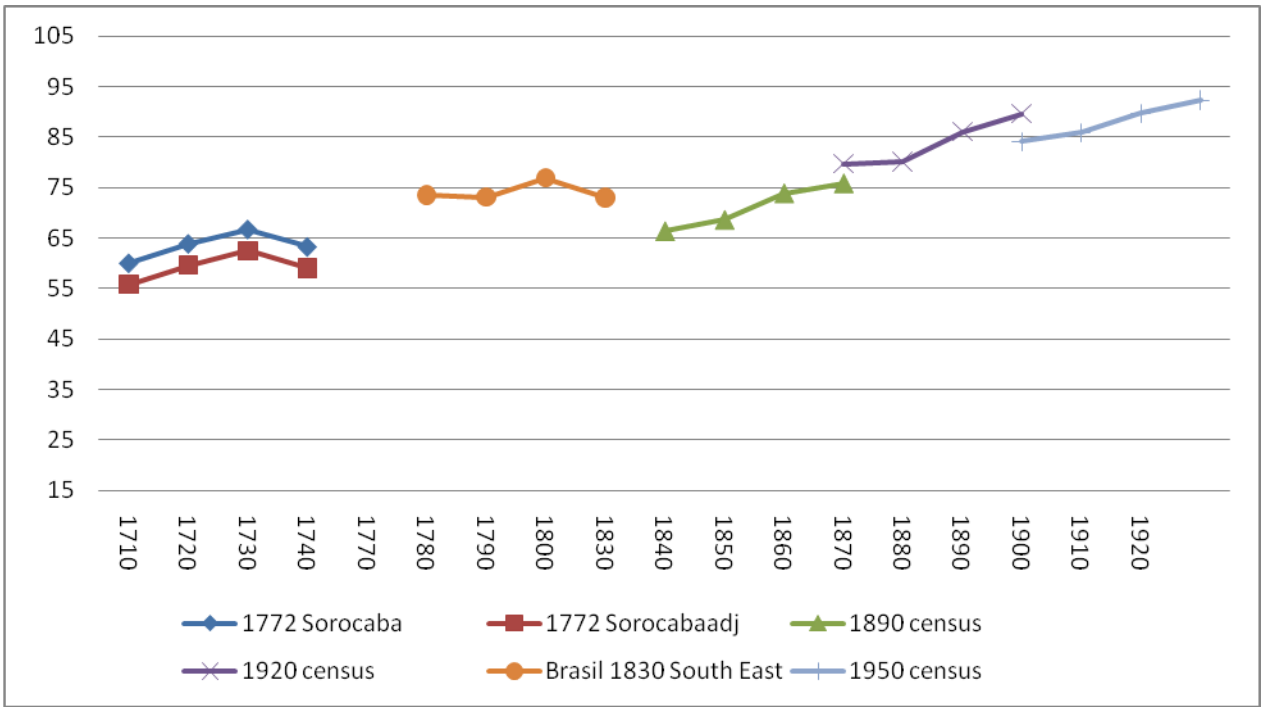
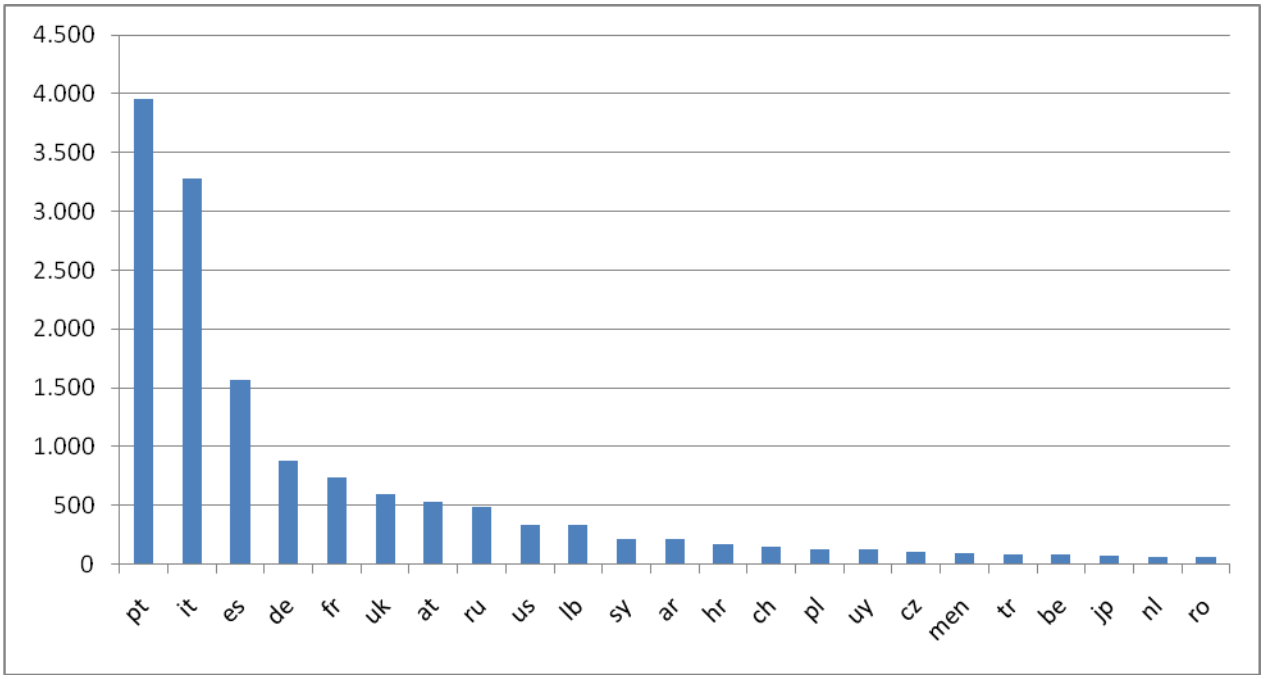


Figure 3: Numeracy estimates for Brazil 1700-1910. Sources see Text. Author's calculations.



Data: Author's calculations, sources, see text

Figure 4: Immigrants by Source Country n>50.



Notes:

Author's calculations. Data from archival records of the National Archive in Rio de Janeiro and http://www.an.gov.br/sian/inicial.asp. MEN stands for immigrants that were listed to be "arabe", "ottoman", "orient", "oriental" with no source country more specifically given. Country abbreviations according to DIN ISO 3166.

Figure 5: Immigrant Numeracy versus Brazilian Numeracy 1770-1910. (Selected Countries)

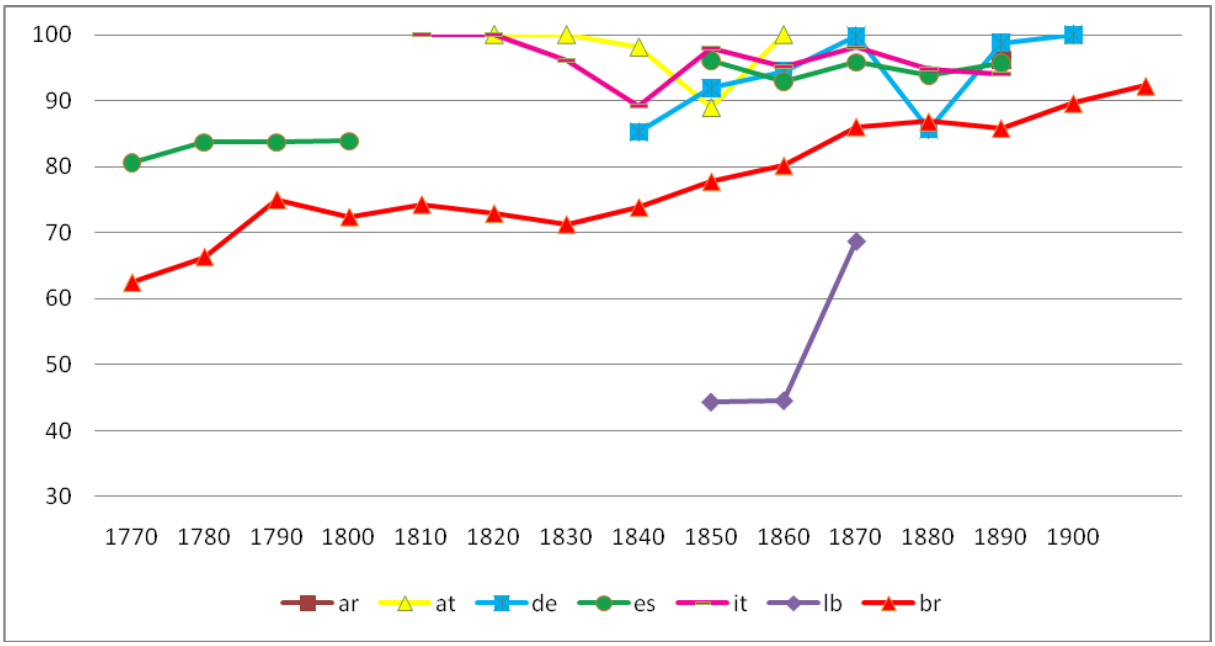


Figure 5b: Immigrant Numeracy versus Brazilian Numeracy 1770-1910. (Selected Countries)

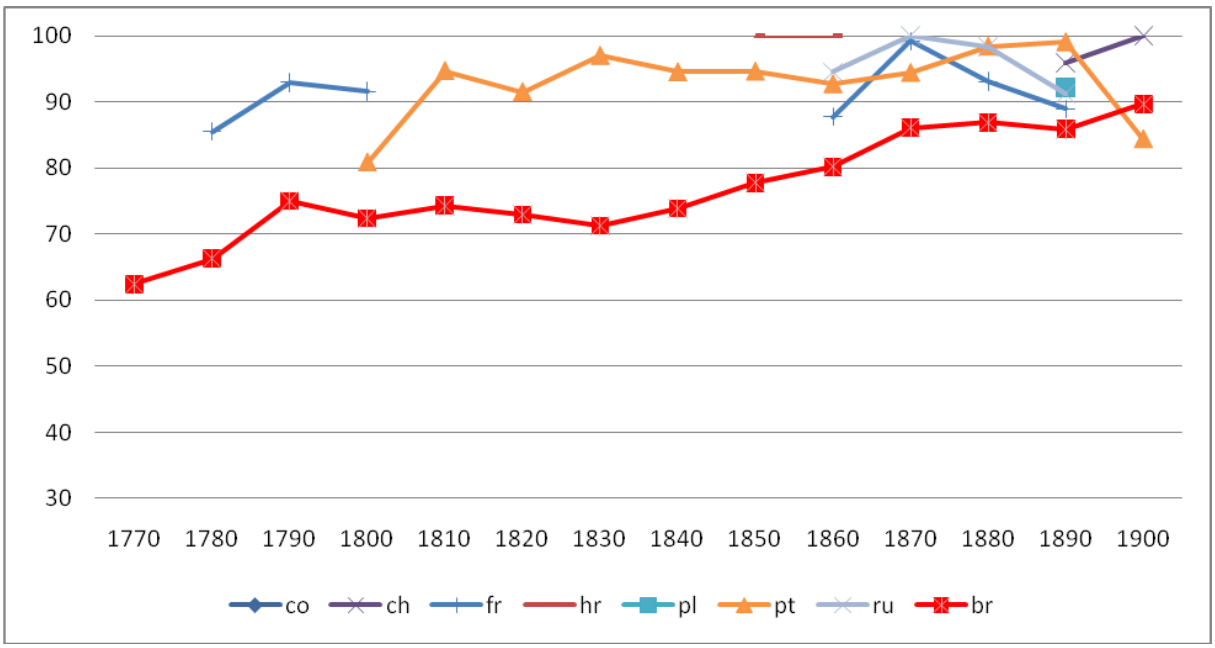
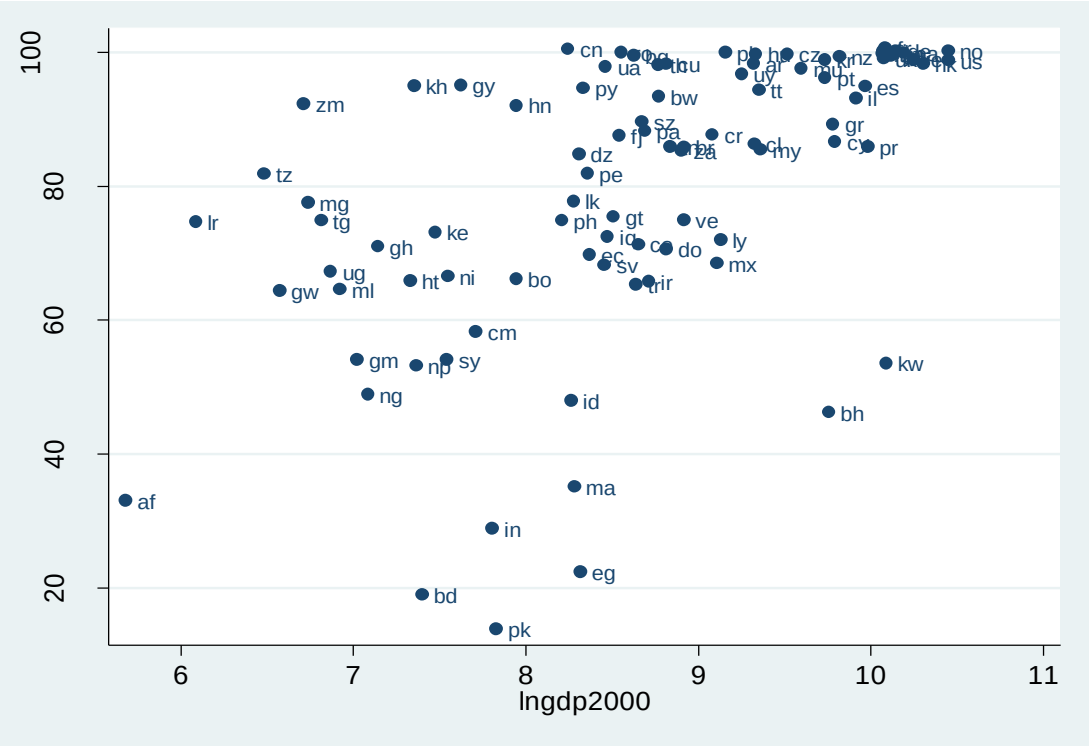
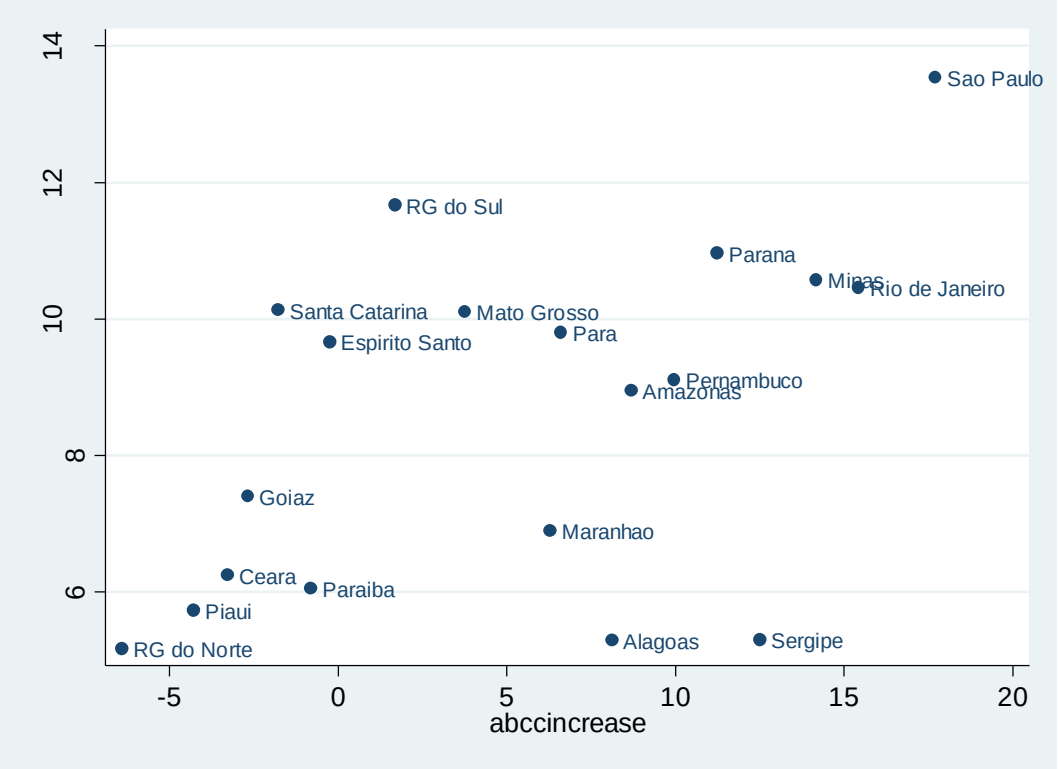


Figure 6: ABCCs in 1900 versus real GDP per capita 2000



Data: ABCCs: Crayen and Baten 2009, Prayon and Baten 2010. GDP per capita: World Penn Tables

Figure 7: Log immigrant increase versus increase in ABCC levels between 1890 and 1920.



Data: ABCCs: Crayen and Baten 2009, Prayon and Baten 2010, Log immigrant increase: Annuario estatistico do Brazil 1908-1912. Rio de Janeiro: Directoria Geral de Estatistica, v. 1-3, 1916-1927 and Anuário estatístico do Brasil 1938. Rio de Janeiro: IBGE, v. 4, 1939.

4.12 Tables

Table 1: Long term impact of human capital formation on economic performance

	(1)	(2)
	m1	m2
VARIABLES	lngdp2000	lngdp2000
Numeracy in 1900	0.03***	0.02***
	(0.000)	(0.001)
Democracy		0.16***
		(0.000)
Longitude		0.00
		(0.546)
Latitude		0.01***
		(0.010)
Landlocked		-0.50*
		(0.069)
Civil War		0.34
		(0.350)
Constant	6.31***	6.53***
	(0.000)	(0.000)
Observations	90	81
R-squared	0.29	0.59

in parentheses

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

Table 2: Regional Regressions of ABCC increase 1890-1920

	(1)	(2)	(3)
Log absolute immigrant increase per state	1.15***	1.58***	0.98*
	(0.006)	(0.000)	(0.084)
Log population density	1.46***	1.19**	-0.37
	(0.009)	(0.045)	(0.867)
Initial abcc	-0.52***	-0.63***	-0.59***
	(0.000)	(0.000)	(0.009)
Average educational expenditure	4.96***		8.20**
	(0.009)		(0.046)
Commodity export: cattle			6.48
			(0.151)
Commodity export: coffee			2.47
			(0.660)
Commodity export: cotton			4.68
			(0.174)
Commodity export: mate			3.99
			(0.344)
Commodity export: rubber			-6.16
			(0.425)
Commodity export: sugar			6.02
			(0.367)
Commodity export: tobacco			0.00
			(.)
Initial education expenditure in 1890		3.41*	
		(0.063)	
Constant	16.80*	29.22***	31.64*
	(0.059)	(0.006)	(0.090)
Observations	19	19	19
R-squared	0.85	0.83	0.89

Robust in parentheses

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

Table 3: Robustness-Test – Do the results hold also for the 1890 - 1950 ABCC increase?

	(1)	(2)	(3)
Log absolute immigrant increase per state	1.19***	1.65***	1.02*
	(0.006)	(0.000)	(0.084)
Log population density	1.52***	1.24**	-0.38
	(0.009)	(0.045)	(0.867)
Initial abcc	-0.50***	-0.62***	-0.57**
	(0.000)	(0.000)	(0.013)
Average educational expenditure	5.17***		8.55**
	(0.009)		(0.046)
Commodity export: cattle			6.75
			(0.151)
Commodity export: coffee			2.57
			(0.660)
Commodity export: cotton			4.88
			(0.174)
Commodity export: mate			4.15
			(0.344)
Commodity export: rubber			-6.41
			(0.425)
Commodity export: sugar			6.28
			(0.367)
Commodity export: tobacco			0.00
			(.)
Initial education expenditure in 1890		3.55*	
		(0.063)	
Constant	17.50*	30.43***	32.96*
	(0.059)	(0.006)	(0.090)
Observations	19	19	19
R-squared	0.84	0.82	0.89

Robust in parentheses

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

5. BRAIN DRAIN, NUMERACY AND ANTHROPOMETRIC INEQUALITY DURING THE ERA OF MASS MIGRATION: TESTING THE ROY-BORJAS MODEL WITH NEW INDICATORS

5.1 Abstract

Brain drain is a core economic policy problem for many developing countries today. Does relative inequality in source and destination countries influence the brain-drain phenomenon? We explore human capital selectivity during the 1820s to 1900s period. In a sample of 52 source and five destination countries we find selective migration determined by relative anthropometric inequality in source and destination countries. The results remain robust in OLS, IV and GMM approaches. We confirm the Roy-Borjas model of migrant self-selection. Moreover, we find that countries like Germany and UK experienced a small positive effect, because the less educated emigrated in larger numbers. We apply age heaping techniques to measure human capital selectivity of international migrants.

This chapter is based on a paper written jointly with Joerg Baten (University of Tuebingen). It is accepted by *Explorations in Economic History*. The concept of the paper was developed jointly; analyzes and writing equally shared.

5.2 Introduction

For countries with substantial emigration rates, brain drain is a core economic policy problem.[180] The migration of Germans to the U.S. or Switzerland in the recent past, for instance, has been discussed as brain drain, because the high U.S. skill premia attract a large number of highly skilled Germans. Chiswick (2005) summarized that often the "best and the brightest" would leave their home country to emigrate to more promising labor markets.[181] In African countries, brain drain is perceived as an important issue, as well (Docquier 2006). Although the health situation on the African continent is problematic, highly skilled African physicians leave and move in large numbers to the Western World because of higher returns to human capital. In a recent article in a leading medical journal, 'The Lancet', it was suggested that the recruitment of physicians from poor countries with high mortality ought to be treated as a criminal case because this would result in more people dying in the African source countries (Mills et al 2008). Consistent with these approaches, we define 'brain-drain' as the phenomenon where, relative to the remaining population, a substantial number of more educated people emigrate.

What determines the selectivity of migrants? Among other explanatory variables, relative inequality has been stressed in the theory of self-selection. If inequality is higher in the destination than in the source country, we would expect highly skilled individuals to migrate, as Borjas (1987) formulated on the basis of Roy's self-selection model (Roy 1951): higher inequality implies that the most educated receive higher relative wages. His views stimulated an excited debate because those who came to the U.S. from emigration countries with a high inequality like Mexico were expected to be negatively selected. That is, we would expect people with less education than the average person at home to migrate. We will call this theoretical approach "Roy-Borjas model" in the following, as Borjas applied the Roy Model to the process of migration (Borjas 1987). We introduce a methodological innovation to the migration literature by using anthropometric inequality measures. These are based on a large project in which evidence was collected on human stature as a

[180] This includes countries like Germany, which just recently was reported as net emigration country by a leading weekly journal ("die Zeit" 10/2010).

[181] Of course, in today's world of skill-selective immigration policies, incentives in source countries sometimes also impact on acquiring a good education in order to have the choice to migrate, even if the more educated individual does not migrate in the end. Furthermore, international migrants send remittances to their home countries that also have an important developmental effect. For example in the Philippines, remittances made up just over 10 percent of national income in 2007. In Mexico and India, the figures have even been higher.

welfare indicator (Fogel 1994, Steckel 1995, Komlos and Baten 1998, Blum and Baten 2011, see section 3).

The impact of relative inequality on the selectivity of international migrants remains so far unanswered. Brücker and Defoort (2006) find a positive correlation between inequality in the home country and educational selectivity of migrants in the OECD for the 1980-2000 period and develop a theoretical model that explains that better educated people can cope with migration policy hurdles. Furthermore, they find that inequality impacts positively on the human capital selectivity of migrants. Feliciano (2005) studies 32 immigrant groups in the US labor market and compares them with their source countries regarding their education and inequality level. Her results were not consistent with the Roy-Borjas model. Belot and Hatton (2009), however, find evidence for a modified Roy model for OECD immigration during the past decades.

We contribute the analysis of a new and unique data set to this debate, as we can include international migrants from 52 source countries who went to five destinations in the Americas and in Europe during the era of mass migration (1820s-1900s). A migrant is defined as somebody born outside the destination country, hence no citizenship or ethnic definition contaminates the study of migration decision. The overall number of underlying individual observations is 6.2 million. We aggregate them by decade and by source and destination country pairs (minimum: 50 cases). We obtain 127 country pairs with sufficient underlying cases. Our evidence provides a unique setting to investigate the question at hand, because migration flows were not yet mainly determined by immigration policies, which nowadays shape migrant selectivity significantly.[182] We include U.S. data until the 1900s, as the U.S. did not have strong immigration restrictions until 1919. Our Argentinean evidence covers only the migration until the decade of the 1880s, as Argentina was the first to impose strong immigration restrictions starting mainly in the 1890s (Timmer and Williamson 1996, Sanchez-Alonso 2008). The data studied here, therefore, provides relatively undistorted evidence of migrant self-selection. We include not only major transatlantic destination countries, but also European immigration targets such as the UK. Finally, we also study one destination country which had actually more emigration than immigration: Norway had significant immigration from Sweden, Finland, Denmark, Germany, Iceland, but also Italy, UK, US and Russia. The major transatlantic destination countries

[182] Germany, for example, attracted relatively low-skilled migrants during the 1960s and thereafter, because of the immigration policies at that time that aimed at providing unskilled labor for factory work, and the family unification allowances during the following period. Ireland on the other hand, attracted highly skilled labor in the recent decades which is partly due to its immigration policy, and partly due to large amounts of foreign direct investment before the economic crisis of 2009.

are represented in our sample by the U.S., Canada and Argentina. This study is the first general assessment of migrant selectivity during this most crucial period of human migration history: the age of mass migration.

We apply the age heaping approach which captures basic numeracy skills by looking at the share of people who are able to report an exact age. In previous studies, this measure has always been found highly correlated with other education indicators (see, for example Crayen and Baten 2009). It allows the calculation of the difference between migrants' numeracy and numeracy of the source country population. We use this differential as the dependent variable and regress it on a set of explanatory variables.

The paper is structured as follows: the next section briefly reviews the theory on human capital selectivity of international migrants and the results of earlier empirical studies. Then, we introduce the method, data and the model we estimate. In a next step, we discuss the findings and make a number of robustness tests. We address the potential endogeneity problem by proposing a set of instruments to measure relative inequality in source and destination countries. We end with a conclusion.

5.3 Theory: the relationship between skill selectivity and inequality

On the micro level, economic theory implies that utility maximising individuals base their migration decisions on the benefits and costs of migration. Provided, the skill set a migrant incorporates is sufficiently applicable in the destination country, the expected yield from such a decision is the income gap between destination and home country multiplied by the probability of not being unemployed.[183] Migration costs comprise all the psychological, physical and material costs of the journey and subsequent settlement in a different environment. Since migration always requires a certain amount of cash or "out-of-pocket"-money (Liebig and Sousa-Poza 2004: p. 128), and credit markets are often imperfect, a poverty constraint exists, as the poorest often simply cannot afford the cost. This restriction explains why, during the process of economic development, migration rises, when a country experiences initial economic growth. The poverty constraint loosens and more people can afford to migrate.

Migration costs increase with geographical and cultural distance, because travel and other costs (e.g. learning a language, religious differences etc) will be higher and the successful integration into the destination society might be more of a challenge. They decrease

with growing diaspora communities in the target country, because friends and relatives living abroad might send remittances and provide valuable information, employment or other support for the newly arrived migrant.[184]

The impact of all these determinants on migration decisions is relatively well-documented. Hatton and Williamson (1998) have prominently shown that economic incentives played an important role in determining migration flows throughout the mass migrations of the 19th century. What is less clear, however, is the question of what determines migrant selectivity. Borjas (1987) developed a framework based on the Roy Model to approach the issue of migrant selectivity (Roy 1951). The basic model was originally formulated to explain occupational self-selection and its impact on inequality, when an individual has the possibility chose between two options. Given that the skills are sufficiently correlated among occupations, the individual will select into the occupation that provides the highest expected earnings. Borjas (1987) adapted the model to migration decisions. Here, the migrant selects himself into migration to a certain destination country, when his skill set will realize more income in the destination labor market than in the domestic one. An underlying assumption is that the skills can be applied in both countries and are sufficiently valued in both labor markets. A second condition is a market with sufficient information so that migrants are able to respond to the incentive. The question is whether these assumptions are valid for the 19th century, our period of study. Previous migrants often informed their friends and relatives back home about the situation in the target country by writing letters. While their letters were sometimes more optimistic than the real situation, they provided some insight as to the comparative welfare of skilled and unskilled workers Moreover, a large number of migrants reversed their decision if the benefits were not as large as expected and returned home.

To sum it up, according to the model, whether a person with a given skill level actually moves or not depends *ceteris paribus* on the relative inequality of source and host country. Positive selection occurs when the destination displays a higher skill premium than the home country (see, for example German or African migration to the US in recent decades, or Russian Jews moving to 19th century U.S.). Negative selection occurs in the opposite case.

[183] During the late 19th century, labor markets were not much regulated; hence obtaining a job at low wage was typically possible.

[184] Besides those variables, the importance of population growth to explain migration rates has been stressed, which translates on labor markets into a relative labor abundance in certain sectors which puts wages under pressure and can therefore make emigration more attractive to the affected individuals.

Belot and Hatton (2009) develop a variant of the Roy model to explain educational selectivity of migration flows into 29 OECD countries over the past decades. They also include immigration policy and poverty constraints. After controlling especially for a poverty constraint – as the poorest are not able to migrate – they obtain significant results for link between inequality and selectivity the Roy model proposes. Moreover, they find cultural and geographic distance to be very important.

Other empirical studies, in contrast, did not confirm the Roy-Borjas model. Brücker and Defoort (2006), for example, find a positive correlation between inequality in the home country and educational selectivity of migrants. They argue that this is caused by higher abilities of the educated to jump over immigration restriction hurdles. Moreover, they find the same correlation for host country inequality. Feliciano (2005) finds no effect of income inequality on human capital selectivity for 32 immigrant groups in the US labor market, which also does not correspond with the Roy-Borjas model prediction. So far, there is no general agreement about the relationship between inequality and human capital selectivity of international migrants.

Moreover, the issue has not been investigated from a historical and a broad international perspective until now. Wegge (2002) and Abramitzky et al. (2009) provide valuable studies on country cases and Cohn (2009) studies the early skill composition of mainly English, German, and Irish migrants to the United States 1820-1860 using the occupational composition of migrants as a proxy. Cohn makes clear that it was the migrants themselves, who declared the occupations. They sometimes tended to make exaggerated statements about their social and occupational status at home. In a review of Cohn's book, Kampfhoefner (2009) suggested to complement this approach with the age-heaping method. Mokyr (1983) pioneered the technique for the Irish case (see also Ó Gráda (1986) for Baltic migrants to Dublin).

We extend those valuable historical studies by using the age-heaping indicator and by focusing on five destination countries and 52 source countries, offering additional systematic insights on this issue, taking a long-run, international approach.

5.4 Other determinants of migrant selectivity

We expect transport costs and poverty constraints to play an important role. The log distance from the source country capital to the destination country capital multiplied with the

decade-specific cost is included to proxy migration costs.[185] As the inhabitants of many poor countries and the poor in medium-income countries simply could not afford the transatlantic journey and many could not even afford migration within Europe, we need to control for poverty constraints. We subtracted GDP per capita from the maximum GDP per capita achieved in this period to obtain a measure of poverty (Maddison 2009).[186] As the poverty constraint might be less binding for a journey to a country which is closer, we multiply the logarithm of the distance with the measure for living standards to allow for varying intensity of this effect.

Other important components in the model are chain migration effects and remittances that earlier migrants might provide. Previously migrated friends or relatives sent home not only money, but also information about the destination country, which decreases the perceived risk of migration. Diaspora communities, too, provide valuable information and support in the form of money, employment, a shared language and identity, which makes the distance from home easier to bear. All these factors reduce the psychological and monetary costs of migration. Cohn (2009) found that the friends-and-relatives effect decreased human capital selectivity of transatlantic migrants between 1820 and 1860. Especially during the last decades migrants were less positively selected from the underlying source country population. Around mid-century less skilled individuals could also afford the cost of passage and ever greater numbers of migrants wanted to escape the catastrophe of "the hungry" 1840s in many European countries. Mokyr (1983) confirms that early migrants often reported occupations with high social status, but found that age heaping was significantly higher among Irish migrants than among the Irish population. While this is true for the whole pre-famine period, age-heaping on emigrant ships that arrived during the famine years was even higher. In some European countries the travel costs of the poor were even paid by the municipal communities which wanted to avoid the social transfers (von Hippel 1984, Bade 2008). This contributed to less positively selected migrants. Our data set concentrates on the immigrants of those mid-century decades and thereafter, continuing in the U.S. case until 1910.

Apart from economic incentives, political, cultural and religious factors might also play a role. In German historiography, the democratic revolution attempt in 1848 and its aftermath generated an outflow of highly educated individuals, who continued to play a

[185] The distance measure as well as data on colonial ties and common languages is taken from http://www.cepii.fr/anglaisgraph/bdd/distances.htm . On the decade-specific costs, see Sanchez-Alonso (2008).

role in American policies. We test whether the democracy situation in the destination country, relative to the source country, might have an impact on the selectivity of migrants.

Eastern European migration was significantly shaped by religious factors. The Jewish minority experienced strong discrimination in the Russian Empire during this period, which reached its maximum in the pogrom waves of the 1880s. During the 1880s, the mass exodus of more than two million Russian Jews began. But already before, a constant stream of highly skilled Jews left the country. This pronounced selectivity was not caused by economic incentives, but by political persecution. Therefore, we control for such occasions whenever possible in our regressions.

Finally, we control for common language and colonial ties. On the one hand, having to acquire a new language requires higher human capital of migrants than being able to use the mother tongue. On the other hand, advanced human capital can be more easily transferred between countries sharing the same language. This would suggest a positive effect on selectivity. While the sign of the effect is not clear, we would expect skill selection to differ for country pairs with the same language. Colonial ties often show the same features. A common culture and common institutions will make it easier for the migrant to adapt to the new environment. In the case of migration from India to Britain, the type of colonial migrant might have been quite often government officials who went to the colonies to work in the administration or military. Their families might later have returned to Britain, in which case we would expect them to be more numerate than the source country population.

5.5 Methodology: skill selectivity

Age heaping is a method that uses the share of persons who report their exact age, as opposed to those who round erroneously, as an indicator for basic numeracy (Mokyr 1983, Crayen and Baten 2009a and 2009b). This indicator has been widely applied recently (A'Hearn, Baten and Crayen 2009, de Moor and van Zanden 2008, Clark 2007, Humphries and Leunig 2009, Cinnirella 2008, O'Grada 2006). A'Hearn, Baten and Crayen (2009) have shown that within societies characterized by a lower level of human capital, the frequency of people stating their age erroneously is higher than in more developed societies. The tendency is to mention a convenient multiple of five instead of the exact age, which becomes evident in the frequency distribution of the age data. The ratio of the frequency of

[186] Where this was not available, we used imputations based on anthropometric values, see Baten and Blum 2010. This method exploits the fact that for historical data, the biological standard of living proxies living standards quite satisfactorily.

multiples of five in relation to the frequency of all mentioned numbers is defined as the Whipple Index.[187] The ABCC index employed in this paper is a simple linear transformation of the Whipple index. It represents the estimated percentage share of the population who reported an exact age (A'Hearn, Baten and Crayen 2009).

The ABCC index correlates strongly with literacy rates, schooling and other human capital indicators, a relation which remains relatively stable across time and space and is robust when applied to different types of data sources. Generally, the age heaping approach is considered a viable method to capture human capital in empirical studies. The great advantage of age heaping is the great variety of sources, where evidence can be drawn from and its coherent construction over space and time (Crayen and Baten, 2009). Our numeracy estimates are arranged by birth decade averages because basic numeracy is acquired during the first decade of life (Crayen and Baten, 2009). We only use the age group 23-72, to avoid age effects caused by selective mortality of the older age groups thereby obtaining up to a maximum of five cohorts with each census (those aged 23-32, 33-42, …, 62-72).[188]

Our dependent variable which measures human capital selectivity is constructed as the difference of the mean ABCC Index of migrants and the mean ABCC of the source country population per decade.

$$S_{ijt} = Mig_{ijt} - (Mig+Stayer)_{it} \quad (1)$$

This measure is different from other migration selectivity studies which use the share of secondary and tertiary educated persons or years of schooling for 20th century studies (Belot and Hatton 2009) or wage data to proxy for returns to human capital (Borjas 1987). These proxies are not available for the era of mass migration studied here. In our study, the source country numeracy is calculated as a weighted share of stayers and migrants if the migration rates reach a substantial number, since during the time before the migration decision, the migrants' human capital still was part of the source country environment.[189]

[187] The optimum is 100, i.e. an equal distribution of mentioned ages throughout the population, the extreme of 500 occurs, if everybody mentions a multiple of five only.

[188] Interestingly, while some specialized studies have used the occupational structure and age heaping of migrants as indicators, the literacy of migrants was not used before. Unfortunately, literacy of immigrants at arrival was only assessed in the U.S. starting in 1899, when the U.S. public grew concerned with the educational status of recent mass immigration from Southern and Eastern Europe, and those lists are not available as individual data sets. (See Internet Appendix C)

[189] We used the migration numbers in Ferenczi and Willcox to identify the countries in which the migration rate exceeded one percent per decade to a given target country (in most cases, there was only one target country with such substantial migration). For the periods before 1870, we used the stock of migrants in the target

The argument might arise that results are biased, if the census taking process in home and target country differ or if the states are differently institutionalized and therefore gather their citizens' ages with different frequency. However, Crayen and Baten (2009) have shown that the number of previous censuses taken as a proxy for institutionalized state-authority does not have a significant impact on the outcomes of the ABCC Index. Moreover, we will control for destination and source country fixed effects, which captures unobserved source and destination country specific effects.

Another possible concern relates to the numeracy of the migrants, which is based on questions posed years after migration, since in the meantime the migrant could have acquired further skills in the destination country. We did, however, counter-check our results with a sample of migrants that were obtained from ship lists, directly after arrival in the destination countries, and the correlation was very close.

5.6 Estimating Inequality

A second methodological question was the measurement of relative inequality. Although Borjas' original model looks at the standard deviation of wages, most of recent studies on the model use Gini coefficients of the income distribution, because they are available for a large number of countries since 1980. The underlying assumption is that wage variation and overall income Gini coefficients correlate.[190] For the 19th century, data on inequality has been scarce until recently (see van Zanden et al. 2009, Blum and Baten 2011), therefore we use an innovative approach to capture this independent variable with a newly available data set, that builds on the work of these authors.

Our core independent variable is inequality in the destination country minus inequality in the source country, constructed with an anthropometric method. Baten (2000) argued that the coefficient of variation of human stature is correlated with overall inequality in a society, and that it can be used as proxy measure, especially where income inequality indicators are lacking. The correlation has been confirmed in further analyzes, for example by Pradhan et al. (2003), Moradi and Baten (2005) and van Zanden et al.

countries, and compared overlapping numbers between Ferenczi and Willcox and census data in order to make sure that the differences in counting (Ferenczi and Willcox focus on migration statistics, hence an Irish migrant to Canada might have finally gone to the U.S; the census stock excludes those who died between migration and census taking). But the correspondence between both sources was quite good. For example, for the 1860s Ferenczi and Willcox list some 700,000 migrants from the UK (incl. Ireland) to the U.S., whereas the stock in the 1880s that we estimated to have migrated during the 1860s was 660,000. We then calculated the weighted average of numeracy of stayers and migrants. Only for very few cases we had to assume similar values to the ones of other migrants (for example, we assumed that Spanish migration to Brazil in the 1880s was similar to the one to Argentina in the 1880s etc.).

[190] Belot and Hatton (2009) use wage differences measured in the wages for occupations that normally require some skills versus some that do not

(2009). This method has been widely used in the economic history literature (Sunder 2003, Guntupalli and Baten 2006, Baten 2000). The idea is that heights reflect nutritional conditions during early childhood and youth. As wealthier people have better access to food and shelter and exhibit therefore less morbidity, they tend to be taller than the poorer strata of the population. Hence, the variation of height of a certain cohort may be indicative of income distribution during the decade of their birth.

Generally, anthropometric literature stresses that the approach offers a good complement to conventional inequality indicators and, in some respect, constitutes perhaps an even better indicator. If the distribution of food and medical goods in an economy becomes less equal, heights should also become more unequal. Yet, while a correlation with income does exist, this correlation is only partial, since some important inputs are not traded on markets but are provided as public goods. These lead to modest deviations between purchasing power-based and height-based inequality measures. This is a major argument in favor of height-based inequality measures: heights are an output indicator, whereas real income represents an input to human utility. Deaton (2003) and Pradhan et al. (2003) have highlighted the importance of measures of health inequality in general. A growing literature also points towards an important effect of health on economic growth (Bloom, Canning and Sevilla 2004, Weil 2007, Aghion, Howitt and Murtin 2010). Heights capture important biological aspects of the standard of living (Komlos, 1985; Steckel, 1995). Also in migration decisions a proxy that captures overall welfare is relevant, because individuals not only maximize their income but also health and longevity.

Anthropometric methods are particularly advantageous for studying developing countries of the 20th and the generally poorer countries of the 19th century. Moradi and Baten (2005) argue that the method generated new perspectives on inequality and is suitable as a useful countercheck for other indicators.

To illustrate the effect of social inequality on the mean height and height variance of the population, consider two different allocations of resources after birth (on the following, see Moradi and Baten 2005): All resources are perfectly equally distributed among the people in society versus a situation where there exists an unequal distribution of resources. In the first case, the height distribution only reflects genetic factors. Despite perfect equality, we observe a *biological variance* of (normally distributed) heights because individuals are differently tendowed with their genetic inheritance. In the latter case, the unequal allocation of resources allows some individuals to achieve their genetic maximum height because they have access to resources, while others who don't will stay shorter. In comparison with the first situation of perfect equality, the individual heights of the upper social

class shift therefore to the right, the heights of the lower social strata shifts to the left. Therefore rising social inequality reflects in a more unequal height distribution, however the correlation is not perfect as the genetic variation of heights is not affected. In practice, while most height distributions are normally distributed or very close to normal, the variance of the distribution is larger than in the case of social equality.

Baten (2000) compared height differences between social groups using the Coefficient of Variation (CV), an inequality measure that is the ratio of the mean and the standard deviation of a distribution. He showed that high CVs sufficiently reflect social and occupational differences without relying on classifications. Moradi and Baten (2005) have estimated the relationship between income inequality and height CV for 14 African countries and 29 five-year periods, controlling for the differences in income definition and population coverage. They found that height CV was significantly and positively correlated with the Gini coefficients of income. Furthermore, they recommended the following formula for translating height CVs into income Ginis (ibid: p.29):

(2) $$Gini_{it}=-33.5+20.5*CV_{it}$$

All in all, the relationship between Gini coefficient of income and height CV is well-established but was never before applied in the migration literature. We use the data set of Blum and Baten (2011) to generate the inequality measure.

5.7 Data

For measuring human capital selectivity of migrants, it is necessary to measure both the human capital of migrants and of the population of the source country. For the migrants, we use data sets from the IPUMS and the North Atlantic Population Projects that provide 100 percent census samples for the late 19th century for a number of countries, and smaller samples for other countries.[191] We only use information on individuals that are older than 23, because younger people still are more aware of their age. For numeracy of source countries, we use published national censuses of a great number of countries that were originally compiled by Crayen and Baten (2009). In Figure 1, we show the age distribution 43-82 of the UK population (census 1881) and of UK immigrants to the US in the Ameri-

[191] See notes to Table 2.

can census of 1880. They exhibit the typical spikes at ages that are multiples of five. These are more extreme among the migrants, reflecting a negative selectivity in this case. In the census data, the year of immigration is not noted. All previous migration studies found that a significant majority migrated when they were around age 15-35, except for some children and a small number of older persons. We therefore argue that the period of migration decision must have been mostly two decades after birth. This assumption has been counter-checked with lists created on ships, and we found it justified. The ages 15-35 are in majority by far. Even more importantly, the numeracy by decade and country is almost exactly the same when looking at ship lists (with known time of migration) and census data. Comparing all passenger lists of ships arriving to New York between 1860 and 1895 the correlation of ABCC values by country and decade with the census evidence is 0.6 (p= 0.00, N=105).[192]

Geographically, we cover 52 source countries in Europe, Latin America, Asia, the Asian-Pacific and Africa to the US, the UK, Canada, Argentina and Norway as destination countries (Table 1). With our five destination countries, our panel allows to observe 127 country pairs. For some countries, we observe both immigration and emigration. The global nature of our data set allows an in-depth analysis of international migration during the 19th century. The migration decades range from the 1820s up to the 1900s (Table 2). In this table, the average number of underlying observations is reported for each source country, decade, and destination country. Cases with less than 50 observations are excluded. The U.S. immigration before the 1880s is better documented than thereafter, because the NAPP project provided a 100% sample of the U.S. census in 1880, and smaller samples before and after.

5.8 Results

5.8.1 How did migrant selectivity develop during this period?

We first take a closer look at our dependent variable, which is defined as the numeracy of migrants minus the numeracy in the source country (both in percent). For the investigated period, average numeracy was almost equal in the source countries (90 percent), and in the destination countries (89 percent). The average numeracy of migrants was slightly lower,

[192] We included all ship lists which were provided by the transcriber's guild (New York arrivals: http://www.immigrantships.net/nycarrivals1_6.html). Unfortunately, the number of observations is much smaller than in the case of census data – only some 300,000 compared to 6.2 million that we study here based on the census data -- hence we did not perform the same analysis with the ship lists. The advantage of ship list evidence is the possibility to determine the human capital status (and age) directly at arrival. One disadvantage is that it includes temporary migrants or travelers who returned home after a few months, but still the

namely 87 percent (arithmetic mean by source country), or 86 percent (mean weighted by migrant numbers. On average, there was, thus, no numeracy brain drain, but rather a mathematical brain gain for the source countries, because migrants who left in the 19th and early 20th century were slightly less numerate than the remaining population. The difference is, however, small so that it is more meaningful to look at the variation of brain drain and brain gain between countries and over time and to study the determinants. In the following, we analyze some prominent examples of emigrant countries sending migrants to the US and UK. We arrange all numeracy values by migration decade.

The largest migrant flows to the United States in this period came from Germany and Ireland. These migrants were mainly negatively selected for the early cohorts of our sample (Figure 2).[193] We actually find 6-13 percent lower numeracy among those migrating during the 1820s-1850s compared to later cohorts. Irish migrants display a stronger negative selectivity, perhaps due to the Great Famine years, since remittances sent over by previous migrants were also used by the less educated to leave the country. Those who migrated in the "hungry 1840s" display a value that is 20 percent lower than those, who stayed in Ireland. Over time, this negative selectivity diminishes and eventually dissolves completely for the migration cohorts 1880-1900.[194]

Among the "new immigration areas" in Eastern and Southern Europe – and the Nordic countries such as Sweden – the development is quite different (Figure 3). The Swedes and Italians show a very modest negative selectivity over the whole period with no major changes. In contrast, the Russian immigrants initially are very positively selected. The earliest cohorts migrating in the 1840s are more than 20% more numerate than their compatriots staying at home.[195] This is partly due to the fact that large shares of Russian immigrants were Jews, who have a reputation for being better educated than the overall population. Additionally, the high costs of migration from Eastern Europe translated in highly skilled first-wave migrants. Afterwards, there are probably strong "friends and relatives"-effects at work, also supported by remittances, as illustrated by the fact that the strong positive selectivity of the first decades decreases among the later cohorts. We

comparison to census data provides valuable insights. We thank Oliver de Marco for his immense contribution to this point.

[193] We consider Ireland separately, although it was part of the British Empire, because the characteristics of Irish migrants were different.

[194] Except for the small dip in German selectivity, this might have been caused by the economic crisis of the early 1890s initiated by the Baring crisis.

[195] The immigration cohort of the 1830s would have been even more positively selected, but we removed it from the figure due to quite small sample size, in order not to provide an inadequate impression. Thanks to Ray Cohn for his important comment on this.

should note though that the first decades of Russian migration were characterized by small absolute numbers.

Looking at another world region, we find migration from Latin American countries positively selected. The absolute numbers here are small, which causes some volatility in the series (Figure 4). High migration costs could have caused brain drain for countries like Brazil, Peru and Chile to the US. The situation is different for Mexican migrants, who had lower migration costs due to the geographic proximity. The early Mexican migrants tend to be equally or slightly negatively selected in terms of human capital in comparison to the home country population.

As a second example of a migration destination, England is a particularly interesting case (Figure 5). Here, immigration is predominantly Irish in the first cohorts. These individuals are on average slightly positively selected (between 0 and 5 percent). Therefore, Ireland experiences some brain drain to England, but a brain gain migration to the US. Also, Poland and Russia, and to a lesser extent Canada suffer from brain drain effects due to migration to England. France and Germany, in contrast, did not experience brain drain with their modest migration flows to England.

In sum, although migrants are on average slightly negative selected, the variation between countries is large. Especially during the mid-19th century waves of migration, some of the main source countries display negative migrant selectivity partly caused by payments of source country government institutions who wanted to send away the poorest, and partly financed by remittances of earlier migrants (especially important for the Irish migration, see Cohn 2009, for the German case see Bade 2008). In contrast, Eastern European migrants are quite positively selected. Part of this migration is shaped by religious determinants. The Jewish minority experienced strong discrimination in the Russian Empire during the 19th century, which culminated in the persecution of the last decades of this century. The exodus consisted of individuals with much comparatively high human capital. Economic incentives might have played a minor role in this case because inequality in Russia was large.

5.8.2 What determines migrant selectivity?

We first estimate the factors determining migrant selectivity in an OLS framework using the following equation:

$$(3) \quad S_{ijt} = \alpha + \beta_1 \text{Inequ}_{ijt} + \beta_2 \log(\text{Mig}_{ijt-1}) + \beta_3 \text{PovConst}_{it} + \beta_4 \text{Log}(\text{Dist}_{ij}) + \beta_5 \text{PovConst}_{it} * \text{Log}(\text{Dist}_{ij}) + \beta_6 \text{Dem}_{ijt} + \beta_7 \text{Language}_{ij} + \beta_8 \text{Colony}_{ij} + \beta_9 \text{Civil War}_{it} + \beta_{10} \text{Relative Democracy}_{ijt} + \text{Fixed Effects}_i + \text{Fixed Effects}_j + \text{Fixed Effects}_t$$

Skill-selectivity of migrants S_{ijt} observed in destination country *j* from source country *i* in migration decade *t* is our dependent variable. The prominent explanatory variable is $Inequ_{ijt}$, which is the relative Gini coefficient from destination and source country in a given migration decade. This is the coefficient of interest, when analyzing the Roy-Borjas relationship. Next, we control for a set of standard migration variables that could also impact on migrant selectivity, such as the friends-and-relatives-effect, which we proxy with the log of the absolute immigrants observed in destination country *j* from source country *i* in t-1, $log(Mig_{ijt-1})$ that is, in the decade previous to the estimated migration decade. $PovConst_{it}$ is a variable that controls for the development level of the source country *i* as a possible poverty constraint. The log distance, $Log\ (Dist_{ij})$, between the capitals of source and destination countries is also included. The poverty constraint is also interacted with log distance, since we would expect the poverty constraint to be more binding in case of a transatlantic journey then in the case of intra-European migration because of the lower material and psychological migration costs in the latter case. Additionally, we control for common language, $Language_{ij}$ and colonial ties, $Colony_{ij}$.[196] And finally, we control for *Relative* $Democracy_{ijt}$ of source and destination country and $Civil\ War_{it}$ at the time of emigration in the source country.

Descriptive statistics are displayed in Table 3. The migrant selectivity variable is distributed between -27.9 and +52.0 numeracy points with an unweighted mean of -4.53, indicating that on average, migrants were slightly negatively selected. Relative Gini coefficients are distributed between -28 and 22. The raw migrant stock variable displayed some left skewness, which is why we take the logs.

In our estimations, we always employ destination and source country fixed effects to capture country specific political and socio-cultural characteristics as well as the income situation in destination and source countries.

As a result, relative inequality plays a consistent role in determining migrant selectivity (Table 4). The coefficients of this variable are positive and have the expected sign in all five specifications. The results confirm the relationship proposed by Roy (1951) and Borjas (1987). In the first regression, we include the Russian emigration, although it might

[196] http://www.cepii.fr/anglaisgraph/bdd/distances.htm, last accessed 15.10.2010

have been largely determined by religious factors, as explained in the previous section. In the second to fifth column, it is excluded and our results remain the same. In column 4, we tested a fixed effects model in order to control for unobserved heterogeneity (which is otherwise controlled with country dummies). The coefficient for relative inequality is robust in this specification, as well.[197] In column 4, we assess whether inequality matters only together with the friends and relatives effect, or poverty constraints, which is not the case.

Is the coefficient of relative inequality economically meaningful? One method of measuring economic significance is to consider the effects of one standard deviation of the explanatory variable. If we multiply the standard deviation of inequality (8.21) with its coefficient (0.16, col. 1 in Table 4), we obtain 1.31. This is roughly 16% of the standard deviation of the dependent variable (standard deviation: 8.16), which means that it can explain roughly one sixth of the standard deviation of the dependent variable. If we do the same with the coefficient of the IV regression below (Table 6, column 5: 0.47), we obtain 3.86, which is around 47% of the standard deviation of the dependent variable. This is a substantial share, indicating economic significance.

The other variables had much less consistent effects. The friends and relatives effect has always the expected negative sign, but is only statistically significant in specification 2, Table 4. This indicates some impact of already existing networks on the skill selectivity of new migrants. The provision of information and remittances might have encouraged also less positively selected individuals to migrate.

The poverty constraint variable renders no robust results. It mainly has a negative sign, but is highly insignificant. While Hatton and Williamson (1998) found that it was a determinant of migration flows, human capital selectivity does not seems to be significantly related to this variable.

We tested several distance measures, like the raw distance between the most populated regions in different countries, or a time variant measure of distance costs.[198] The latter is more intuitive, as it can be considered an estimate of "economic distance". The strong decline of transport cost with the arrival of the steamship innovation features prominently here. Nevertheless, this variable turns out to be insignificant (as does the raw distance measure). *Log ($Dist_{ij}$)* has a negative sign and becomes significant in one specifications. This result might seem counter-intuitive, but the result might be due to the fact that

[197] However, the Hausman test indicates that the random effects model applied in columns 1, 2, 4 and 5 is consistent and efficient (Prob>chi2 = 0.1609).

[198] We took the passenger cost estimates by Sanchez-Alonso (2008), and calculated the cost for distance unit for each decade. This is then multiplied with actual distances. (distance measures from http://www.cepii.fr/)

the majority of the variance in the distance variable stems from the difference in trans-Atlantic versus intra- European migration. We observe better selected individuals in European destination countries than in the US or Argentina, for example, perhaps because the risky environment of the New World deterred skilled migrants. Figures 2 and 4 indicate this particularly for the case of Irish emigrants. We also include an *interaction term between economic distance and poverty constraint*, but it turns out mostly insignificant.

Relative democracy is controlled for based on the estimates of democracy produced by the Polity IV project.[199] Since one might expect that the more educated were attracted by higher democracy values in the destination country, relative to the source country. This variable turns out to be insignificant, too. The politically motivated migration might have been too small in number during the 19th century, or it was probably not sufficiently restricted to the more educated strata.

Finally, we tested common language and colonial relationships, and found occasional positive effects for language. A common language might have been more useful for the more educated who usually have a comparative advantage with words and skills, rather than with brawn. Colonial ties do not seem to matter. Finally, in column 4 of Table 4 we test for a potential effect of civil war in the country of origin, which turns out negative, but insignificant.

A first test for robustness is omitting the Germans, Irish and English to test if our result were mainly driven by these very large immigrant groups (Table 5). The results do not differ very much the Roy-Borjas forces remain strong. If the Irish are omitted, the friends-and-relatives effects turns small and insignificant, this is neatly consistent with the literature that argued that this effect was particularly important for Irish migration (Cohn 2009).

An alternative robustness test is to weigh the observations with the number of migrants underlying each unit in a WLS regression which leads to more efficient estimates. Conventionally this is done with the square root of the number of underlying observations. The results in Table 6, column 1 to 4, are consistent with previous estimates: the relative inequality is a significant determinant of migrant selectivity. One potential disadvantage of weighted regressions is that a few source countries account for the majority of migrants; and thus they receive most of the weight in the estimates.

[199] Marshall, Monty G., and Jaggers, K. (2008): Polity IV Project: data set. http://www.systemicpeace.org/polity/polity4.htm

While in the OLS estimations, common language had a positive effect on human capital selectivity, the WLS regressions suggest the opposite (and so does the IV regression below, as well). Here the Irish, who were negatively selected and shared a common language with North Americans, gain a strong weight. A former colonial relationship also seems to be of importance in WLS: this variable turns highly significant and positive in all estimations, indicating that people, who migrated into a country to which they were linked by colonial ties, were more positively selected than migrants moving to other destination countries. This might have been caused partly by re-migration of the families of former colonial officials or similar special factors of the colonial administration.

We also assessed whether the difference between migrant numeracy and source country numeracy might depend on the level of source country numeracy. Those coming from a high education background might have been more likely to be negatively selected, even if we have seen many counter-examples in the Figures discussed above. We therefore include a term "ABCC level source country", which indeed turns significant but did not change the main results (Table 6, column 3). In a similar exercise to evaluate the properties of the dependent variable, we included only those in which the source country numeracy deviates from the optimum of 100 percent (Table 6, column 4). This removes some 35 cases, relative to column 2, but again the coefficients do not change.

A comprehensive test of the properties of time series indicated that the main series as well as the residuals do not display unit root problems. The Fisher test for unbalanced panels, as well as the Hadri-LM-test for the three largest source countries in a balanced panel, were calculated and suggest that our series do not suffer from non-stationarity.

5.8.3 Instrumental Variable Approach

One of the underlying conditions to perform OLS analysis is that the error term is stochastically independent from the explanatories. If this is not the case the problem of endogeneity arises. In this case, a shock in the error term would influence both the explanatory and the explained variable in the equation simultaneously. OLS would then be systematically biased. We would then have to search for an instrument that correlates with the potentially endogenous explanatory variable, but is unlikely to influence the dependent variable except via the potentially endogenous variable. In our case, we are interested in the relationship between human capital selectivity and relative inequality. Theoretically, one could imagine that a massive exodus of a large and highly selected share of a population would influence relative inequality of the source country, since relative inequality should ceteris paribus decline, if, for example, a large share of unskilled workers leaves. In most countries, however, the requirement of a large share of the population leaving is ful-

filled, since emigration rates were normally below 5 percent per decade. Exceptions are Ireland where in some decades more than 10 percent left, and Italy right before WWI (Hatton and Williamson 1998). Hence, endogeneity only arises if the degree of selectivity and the migrant share of the population are large enough, and so it depends on scale. We propose two instruments to fit a two stages least square estimation, the first one rests on the scale argument.

The first instrument rests on the assumption that it is the scale effect that matters for endogeneity. We constructed an instrument that uses the relative inequality of world regions instead of country pairs to instrument for the latter. One could, for example, imagine that selective Irish migration does have an impact on inequality in Ireland, but not on inequality of Europe as a whole, because the numbers of Irish migrants were small relative to the large population of this world region and migration decades. Therefore, we calculated the inequality measure for all the world regions instead of country-destination pairs from which our migrants came. While the largest share of country-decade observations came from Europe, also Asia, Latin America, and North America contributed.

The second instrument is taken from the political sphere. In the second half of the 19th century, the Prussian politician Bismarck introduced the first social insurance laws in Germany. The initiator was a conservative who was concerned about the success of socialist movements in his country; improving the living standard and security of the working masses seemed a sensible strategy to reduce inequality and hence the motivation of workers' associations. The social insurance laws consisted of sickness and injury insurances. Particularly the people at the bottom of the income distribution benefitted from these reforms. After Germany and Austria-Hungary started in the 1880s, other countries followed their example during the decades thereafter, while still others waited until the mid-20th century (Flora 1983, Cutler 2002). This might have had an effect on migrant selectivity, but mainly via the potentially endogenous variable, inequality.

The results are shown in Table 6, Column 5, confirming earlier results of a significant relationship between relative inequality and skill selectivity, and thus the Roy Model.[200] The magnitude of the effect however increases substantially, indicating a potential lower-bound estimate of OLS. Post-estimation analysis shows that our proposed instruments have a joint non-zero impact on the dependent variable, the F-statistic is above the conventional level of 10 ($F(2,338) = 11.08$). We tested for overidentifying restrictions and

[200] The Durbin and Wu-Hausman test of endogeneity rendered both the result that a null hypothesis of exogenous instruments could not be rejected. Furthermore, we regressed the residuals of the second stage on the instruments and they were insignificant.

can reject the hypothesis of weak instruments in our case (Sargan statistic: 1.224, p-value =0.2686, hence we cannot reject the H0 that our instruments are valid).

Finally, we tested whether our results also hold when a Generalized Method of Moments estimator is applied (Arellano and Bond, 1991). While this method is conventially used in dynamic settings to account for the likely endogeneity of lagged dependent variables, it basically generates a large number of instrumental variables from lagged first difference values of the dependent variable. Arellano and Bond (1991) suggested the consistent GMM estimator for this model in the presence of correlation between the unobserved error and the independent variables. This estimation in first differences is of advantage, because it allows us to make sure that trend correlation is not a problem here. The equation is estimated in first differences to eliminate the individual effects. Again, the relative inequality coefficients turn out robust (Table 7).

To conclude, a wide range of econometric techniques suggests that relative inequality had an effect on migrant selectivity as measured by relative numeracy with the age heaping method. There is some evidence -- although more limited -- on friends-and relatives-effects, colonial relationships and common language, whereas counteracting forces might have rendered the effects of economic distance and democracy mostly insignificant.

5.9 Conclusion

In this study, we assessed the selectivity of migrants in the era of mass migration. We focus not only on the main transatlantic migration destinations, but also on two European destination countries, the UK and Norway. No less than 52 source countries could be included, with 127 country pair flows. The underlying data set is based on 6.2 million individual migrants.

The main model tested is the Roy Model of self-selection (1951) that Borjas applied to the process of migration (1987). It states a relationship between skill selectivity of migrants and relative inequality of source and destination countries. We confirm the influence of these economic migration incentives after controlling for a large number of other variables such as “friends and relatives effects”, poverty constraints, economic distance, relative democracy, common language and colonial relationships. This study has been the first general assessment of migrant selectivity during this most crucial period of human migration history, using large samples that included a variety of different source and destination countries.

It is crucial to understand the brain-drain processes between source and destination countries, because the stock of human capital determines future growth capabilities. Brain

drain effects have not been systematically studied for the era of mass migration of the mid-to-late 19th century with large international samples before. In the case of mid-19th century mass migration history, we also noted some arithmetic brain gains for the source countries, since those who left Scandinavia or central Europe around mid-century were often less numerate than the remaining population. There could have been, for instance, marginal positive human capital growth effects in Germany or in some Scandinavian countries, because the average numeracy should have increased with to migration, due to negatively selected emigration. In contrast, Eastern Europe lost a large number of the numerate population, and the migration effects might have been *ceteris paribus* negative. Clearly, also a large number of other factors were at work, which is why these effects should not be seen in isolation.

5.10 References

Abramitzky, R., Boustan, L.P. and Eriksson, K. (2009): "Measuring Selectivity and Returns in the age of Mass Migration" NBER Working Paper 15684.

Aghion, Philippe, Howitt, Peter, and Murtin, Fabrice (2010): "The Relationship Between Health and Growth: When Lucas Meets Nelson-Phelps" NBER Working Paper 15813.

Arellano, M., and S. Bond. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations" *Review of Economic Studies* 58, pp. 277-297.

Bachi, R., (1951): "The tendency to round off age returns: measurement and correction" *Bulletin of the International Statistical Institute* 33, pp. 195-221.

Bade, K.J. (ed.). (2008): *Enzyklopädie Migration in Europa: vom 17. Jh. bis zur Gegenwart*. Stuttgart.

A'Hearn, Brian, Baten, Joerg and Crayen, Dorothee (2009): "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital" *The Journal of Economic History* 69, no. 3, pp. 783-808.

Baten, Joerg, (2000): "Economic Development and the Distribution of Nutritional Resources in Bavaria, 1797-1839", *Journal of Income Distribution* 9, pp. 89-106.

Baten, Joerg and Blum Matthias (2010), Global Height Trends and its Determinants, 1810-1984: An Overview, see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/baten_blum_ht_100331a.pdf, last accessed March 31st, 2011.

Böhm, Andreas and Baten, Joerg (2010): "Trends of Children's Height and Parental Unemployment: A Large-Scale Anthropometric Study on Eastern Germany", 1994-2006.", G*erman Economic Review* 11-1. pp. 1-24.

Belot, M.V.K., Hatton, Tim J. (2009): "Immigrant Selection in the OECD" Center for Economic Policy Research Discussion Paper No. 571.

Bloom, David E., Canning, David, and Sevilla, Jaypee (2004): "The Effect of Health on Economic Growth: A Production Function Approach" *World Development* 32, no. 1, pp. 1-13.

Blum and Baten (2011): "Estimating Inequality with Anthropometric Indicators" *Review for Economics* – German "*Jahrbuch für Wirtschaftswissenschaften*", forthcoming.

Borjas, George J. (1987): "Self-Selection and the Earnings of Immigrants" *The American Economic Review* 77, no. 4, pp. 531-553.

Brücker H, Defoort, C. (2006) "The Self-Selection of International Migrants Reconsidered: Theory and New Evidence" IZA Discussion Paper Series IZA DP 2052.

Chiswick, Barry R. (2005): "High Skilled Immigration in the International Arena" IZA Discussion Paper Series IZA DP 1782.

Cinnirella, Francesco (2008): "Optimists or pessimists? A reconsideration of nutritional status in Britain, 1740–1865", *European Review of Economic History* 12, pp. 325-354.

Clark, Gregory (2007): *A Farewell to Alms: A Brief Economic History of the World*. Princeton University Press.

Cohn, Ray L. (2009): *Mass Migration under Sail: European Immigration to the Antebellum United States*. Cambridge University Press. Cambridge.

Crayen, Dorothee and Baten, Joerg (2009): "Global Trends in Numeracy 1820–1949 and its Implications for Long-term Growth" *Explorations in Economic History* 47, no. 1, pp. 82-99.

Crayen, Dorothee and Baten, Joerg (2010): "New Evidence and New Methods to Measure Human Capital Inequality before and during the Industrial Revolution: France and the US in the Seventeenth to Nineteenth centuries" *Economic History Review,* forthcoming. Published Online: Jul 29 2009 10:50AM. DOI: 10.1111/j.1468-0289.2009.00499.x.

Cutler, David M. and Richard Johnson (2002): "The Birth and Growth of the Social Insurance State: Explaining Old Age and Medical Insurance Across Countries" Working Paper Harvard U /Kansas Fed.

Deaton, Angus, (2003): "Health, Inequality and Economic Development" *Journal of Economic Literature* XLI, pp. 112–158.

De Moor, Tine. and Van Zanden, Jan-Luiten. (2008): "Uit fouten kun je leren. Een kritische benadering van de mogelijkheden van 'leeftijdstapelen' voor sociaal-economisch historisch onderzoek naar gecijferdheid in het pre-industriële Vlaanderen en Nederland" *Tijdschrift voor Economische en Sociale Geschiedenis* 5-4, pp. 55-86.

Docquier, F. (2006): "Brain Drain and Inequality Across Nations" IZA Discussion Paper Series IZA DP 2440.

Feliciano, C. (2005): "Educational Selectivity in U.S. Immigration: How Do Immigrants Compare to those left behind?" *Demography* 42, no. 1, pp. 131-152.

Ferenczi, Imre and Willcox, Walter F. (1929): *International Migrations, Vol I&II.* New York.

Flora, Peter. (1983): *State, Economy and Society in Western Europe: 1815-1975. A data handbook in two Volumes*. Frankfurt, M. Campus.

Fogel, Robert (1994) "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy" *American Economic Review* 84, pp. 369-95.

Guntupalli, Aravinda M. and Baten, Joerg, (2006): "The development and inequality of heights in North, West and East India, 1915-1944" *Explorations in Economic History* 43, no. 4, pp. 578-608.

Hanson, Gordon, H. (2007): *Emigration, Labor Supply, and earnings in Mexico. In: Mexican Immigration to the United States*, edited by George J. Borjas, University of Chicago Press.

Hatton, Tim J. and Williamson, Jeffrey G. (1998): *The Age of Mass Migration: Causes and Economic Impact*. Oxford University Press. New York.

Hatton, Tim J. and Williamson, Jeffrey G (2008): *Global Migration and the World Economy: two Centuries of Policy and Performance*. MIT Press. Michigan.

Hippel, W. von (1984): *Auswanderung aus Südwestdeutschland. Studien zur Württembergischen Auswanderung und Auswanderungspolitik im 18. und 19. Jahrhundert*. Stuttgart.

Humphries, Jane and Leunig, Tim (2009): "Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London" *Explorations in Economic History* 46, no. 1, pp. 120-131.

Liebig, T., and A. Sousa-Poza, (2004): "Migration, Self-Selection and Income Inequality: An International Analysis", *Kyklos* 57, no. 1, pp. 125-147.

Kamphoefner, W. (2009): "Mass Migration under Sail: European Immigration to the Antebellum United States. By Raymond L. Cohn. Book Review" *The Journal of Interdisciplinary History* 40, no. 4, pp. 621-622.

Komlos, John (1985): "Stature and Nutrition in the Habsburg Monarchy: The Standard of Living and Economic Development in the Eighteenth Century" *The American Historical Review* 90, no. 5, pp. 1149-1161.

Komlos, John and Joerg Baten (eds.) (1998) *The Biological Standard of Living in Comparative Perspectives: Proceedings of a Conference Held in Munich January 18-23*, 1997. Stuttgart: Franz Steiner Verlag.

Maddison, A. (2009): *The World Economy: a millennial Perspective*. OECD Publishing 2001.

Manzel, Kerstin and Baten, Joerg (2009): "Gender Equality and Inequality in Numeracy: The Case of Latin America and the Caribbean 1880–1949" *Revista de Historia Económica/Journal of Iberian and Latin American Economic History*, Second Series, 27, pp. 37-73.

Marshall, Monty G. and Keith Jaggers (2008): *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2008*. Center for Systemic Peace and Center for Global Policy at George Mason University. Available at http://www.systemicpeace.org/polity/polity4.htm last accessed 29.03.2011

Myers, R., (1954): "Accuracy of age reporting in the 1950 United States census" *Journal of the American Statistical Association* XLIX, pp. 826-831.

Moradi, Alexander and Baten, Joerg (2005): "Inequality in Sub-Saharan Africa: New Data and New Insights from Anthropometric Estimates" *World Development* 33, No. 8, pp. 1233-1265.

Milanovic, B. (2009): "Global Inequality and Global Inequality Extraction Ratio: the story of the last two Centuries" World Bank Policy Research Working Paper No. 5044.

Mills, E.J., Schabas, W.A., Volmink, J., Walker, R. , Ford, N., Katabira, E., Anema, A., Joffres, M., Cahn, P. and Montaner, J. (2008): "Should active recruitment of health workers from sub-Saharan Africa be viewed as a crime?" *The Lancet* 371, pp. 685-88.

Mokyr, J. (1983): *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850*, London and Boston.

O'Grada, Cormac (1986). "Across the Briny Ocean: some thoughts on the Irish emigration to America 1800-1850" *Migrations across time and nations: population mobility in historical contexts*. New York, pp. 79-94.

O'Grada, Cormac (2006): "Dublin Jewish Demography a Century Ago" *The Economic and Social Review* 37, no. 2, pp. 123-147.

Pradhan, M., Sahn D.E., and Younger, S.D (2003): "Decomposing World Health Inequality" *Journal of Health Economics* 22, no. 2, pp.271-293.

Rodriguez Galdo, M.X., Cordero Torrón, X. (2007): "Emigración e Mercado de Traballo. Espanois en Arxentina 1882-1926" *Revista Galega de Economía*, 16. Numero extraordinario, pp. 93-116.

Roy, A. (1951): "Some thoughts on the distribution of earnings" *Oxford Economic Papers* 3, pp. 135–46.

Sanchez Alonso, Blanca (2008): "The Other Europeans: Immigration into Latin America and the International Labor Market, 1870-1930" *Revista de Historia Económica*/ *Journal of Iberian and Latin American Studies* XXV, no. 3, pp. 395-426

Schmitt, L.H. and Harrison, G.A. (1988): "Patterns in the Within-Population Variability of Stature and Weight" *Annals of Human Biology* 15, no. 5, pp. 353-364.

Singer, J. David and Melvin Small (1972): *The Wages of War, 1816-1965: A Statistical Handbook.* New York. Or see http://www.correlatesofwar.org, last accessed March 31st, 2011.

Steckel, Richard (1995): "Stature and the Standard of Living" *The Journal of Economic Literature* 33, pp. 1903-1940.

Sunder, M. (2003): "The Making of Giants in a Welfare State: The Norwegian Experience in the 20th Century" *Economics and Human Biology* 1, no. 2, pp. 267-276.

Timmer, A.S., Williamson, Jeffrey G. (1996): "Racism, Xenophobia or Markets? The Political Economy of Immigration Policy Prior to the Thirties" NBER Working Paper W5867.

Van Zanden, Jan Luiten (2009): "The skill premium and the "Great Divergence"", *European Review of Economic History* 13, pp. 121-153.

Van Zanden, Jan Luiten., Baten, Joerg, Foldvari, Peter and Leeuwen, Bas v., (2009): "World Income Inequality 1820-2000" Working Paper University of Tuebingen, Dept. Economic History No. 87.

Wegge, Simone A. (2002): "Occupational Self-selection of Nineteenth-Century German Emigrants: Evidence from the Principality of Hesse-Cassel" *European Review of Economic History* 6, no. 3, pp. 365-394.

Weil, David N. (2007): "Accounting for the effect of Health on Economic Growth" *Quarterly Journal of Economics* 122, no. 3, pp. 1265-1306.

5.11 Tables

Table 1: Underlying number of cases by source country

Country	Cases	Country	Cases	Country	Cases
Ireland	1877232	Mexico	45828	Barbados	1841
Germany	1719228	Netherld.	42674	India	1208
UK	978433	Austria	32834	Iceland	1159
Canada	467201	France	29777	Uruguay	1050
Sweden	205227	Russia	26044	Greece	845
Norway	138013	Portugal	11362	Brazil	844
Belgium	101223	Luxembg	10902	Hong Kong	812
China	86092	Spain	9274	Turkey	812
Switz.ld	75371	Hungary	8589	Romania	751
Czech	60458	Finland	8021	Jamaica	670
Denmark	53816	Cuba	4683	Japan	548
Italy	51385	Australia	2229	Bermuda	430
US	47985	Chile	1978	Bolivia	244
Poland	46183				

Sources: see notes to Table 2.

Table 2: Average number of underlying cases for each decade, destination and source country, by decade and destination country

Destination	1820	1830	1840	1850	1860	1870	1880	1890	1900
Argentina		109	265	428	721	655	623		
Canada	197	10664	8723	7228	5220	4352	381	421	
Norway	527	878	1490	2511	2120	2002	1798	1655	
UK	1451	1208	1611	515	411	376			
US	915	13006	24900	32064	35651	30703	989	941	655

Notes: For example, 109 was the average number of cases of all source countries that provided migrants to Argentina in the 1830s.

Census evidence was available for Argentina (1869, 1895) – sample; Canada (1871, 1881-100%, 1901); Norway (1865, 1875, 1900); England (1851, 1881); US (1850, 1860, 1870, 1880-100%, 1890, 1900, 1910).

Sources: On the U.S. except 1880: Ruggles, Steven, Matthew Sobek, and Trent Alexander, et al. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004. On Argentina: Somoza, J. and Lattes, A. (1967): Muestras de los dos primeros censos nacionales de población, 1869 y 1895. Documento de Trabajo No 46, Instituto T. Di Tella, CIS, Buenos Aires. On all other samples: North Atlantic Population Project and Minnesota Population Center. NAPP: Complete Count Microdata. NAPP Version 2.0 [computer files]. Minneapolis, MN: Minnesota Population Center [distributor], 2008. [http://www.nappdata.org];

Table 3: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Migrant selectivity	303	-4.32	8.16	-27.91	52.00
Relative inequality	303	-3.88	8.21	-28	22
Ln migrant stock	303	0.60	2.14	-4.98	4.61
Poverty constraint	303	0.85	0.34	0.00	2.1
Ln distance	303	3.01	1.07	0.48	4.60
Ln dist*pov. constr.	303	2.69	1.56	0.00	8.27
Relative democr.	303	1.34	3.50	-5.70	10.00
Common language	303	0.23	0.42	0.00	1.00
Colonial r'ship	303	0.15	0.36	0.00	1.00

Note: only the cases are included for which all explanatory variables (Table 4, Col 1) did not contain missing values. Sources: see Table 2 and 3. See also notes to Table 4.

Table 4: Baseline regressions of human capital selectivity (numeracy migrant in % - numeracy source country in %)

	(1)	(2)	(3)	(4)
Estimation Method	***RE***	***FE***	***RE***	***RE***
Relative Inequality dest - source	0.16***	0.16***	0.12**	0.14**
	(0.003)	(0.003)	(0.04)	(0.012)
Friends & relatives (Ln stock mig)	-0.43	-1.11*		-0.13
	(0.23)	(0.081)		(0.67)
Poverty constraint	-7.67*	1.53		
	(0.094)	(0.71)		
Ln economic distance	-2.11**	-6.74***		
	(0.039)	(0.002)		
Ln economic distance * poverty constraint	1.55	-1.75		
	(0.31)	(0.25)		
Relative democracy	-0.44			
	(0.43)			
Common language	2.90*			
	(0.097)			
Colonial relationship	1.24			
	(0.52)			
Civil war				-1.81
				(0.11)
Country fixed effects	YES		YES	YES
Destination fixed effects	YES		YES	YES
Time fixed effects	YES		YES	YES
Country pair fixed effects		YES, FE		
Constant	2.6	19.90***	-10.70***	-11.51**
Observations	303	300	376	300
Number of codeno	70	69	78	69
R-squared		0.15		

r2_b	0.852	0.0288	0.861	0.849
r2_w	0.157	0.15	0.113	0.204
r2_o	0.591	0.0838	0.535	0.561

P-values based on robust standard errors are included in brackets. Migration decades 1820s-1900s are included. Column 1, 2, 4 and 5 are estimated with random effects models (but country dummies included), col. 3 is based on fixed effects estimates. Hausman test P-Value (0.1618) suggests, however, that a random effect estimator is consistent and efficient.

Sources: on the migrant numeracy, see Table 2. Numeracy in the source countries are from Crayen and Baten (2009). Inequality is from Blum and Baten (2011), Estimating Inequality with Anthropometric Indicators, forthcoming for an online version, see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/Baten_Blum_skpr100331a.pdf, last accessed March 31st, 2010. The stock of migrants was calculated with migrant data sets cited in the notes to Table 2. Poverty constraints are based on Maddison (2009), and for those countries for which values were lacking we used the imputations first done by Baten and Blum (2010), see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/baten_blum_ht_100331a.pdf last accessed March 31st, 2010. The distance measure as well as data on colonial ties and common languages is taken from http://www.cepii.fr/anglaisgraph/bdd/distances.htm last accessed March 31st, 2010. The distance was then multiplied with the passenger cost estimates by Sanchez-Alonso (2008) to account for the decline in distance costs. Relative democracy data is from the Polity IV project, see Marshall, Monty G., and Jaggers, K.(2008): Polity IV Project: data set. http://www.systemicpeace.org/polity/polity4.htm last accessed March 31st, 2010. Civil War data is from the Correlates of War Project, see Singer, J. David and Melvin Small (1972): The Wages of War, 1816-1965: A Statistical Handbook. New York. Or see http://www.correlatesofwar.org last accessed March 31st, 2010.

Table 5: Robustness of human capital selectivity regression: excluding the largest source countries

Omitted Source Country	Germany	Ireland	United Kingdom
Estimation Method	RE	RE	RE
Relative Inequality dest - source	0.15**	0.13**	0.14**
	(0.015)	(0.031)	(0.017)
Friends & relatives (Ln stock mig)	-0.69*	-0.12	-0.54
	(0.072)	(0.703)	(0.150)
Poverty constraint	-4.22	-6.58*	-6.03
	(0.351)	(0.093)	(0.133)
Ln distance	-2.00*	-0.77	-1.67*
	(0.067)	(0.445)	(0.090)
Ln distance * poverty constraint	0.82	0.74	0.97
	(0.593)	(0.585)	(0.482)
Common language	5.08**	5.32***	3.41
	(0.021)	(0.004)	(0.113)
Colonial relationship	1.33	-1.90	0.18
	(0.578)	(0.413)	(0.943)
Source fixed effects	YES	YES	YES
Destination fixed effects	YES	YES	YES
Time fixed effects	YES	YES	YES
Constant	0.19	-4.84	0.30
	(0.979)	(0.427)	(0.964)
Observations	272	281	277
Number of codeno	64	66	65
r2_b	0.861	0.880	0.864
r2_w	0.224	0.160	0.186
r2_o	0.595	0.608	0.587

P-values based on robust standard errors are included in brackets. Migration decades 1820s-1900s are included. Russia excluded. Sources: see Table 2 and 3.

Table 6: Regression of human capital selectivity, weighted by number of underlying observations, and IV estimation

	(1)	(2)	(3)	(4)	(5)
Estimation method	LSDV	LSDV	LSDV	LSDV	IV
Relative Inequality dest - source	0.15**	0.15**	0.16***	0.17**	0.47***
	(0.016)	(0.018)	(0.003)	(0.013)	(0.009)
Friends & relatives (Ln stock mig)	-0.05	-0.05	0.13	-0.48	-0.24
	(0.863)	(0.861)	(0.612)	(0.118)	(0.311)
Poverty constraint	-2.49	-2.39	-1.71	0.34	-1.62
	(0.525)	(0.544)	(0.679)	(0.944)	(0.798)
Ln distance	-2.55***	-2.53***	-1.63*	-2.38**	-4.30***
	(0.005)	(0.005)	(0.082)	(0.013)	(0.000)
Ln distance * poverty constraint	1.02	0.98	0.02	0.15	1.05
	(0.447)	(0.464)	(0.989)	(0.923)	(0.557)
Relative democracy	0.23				
	(0.564)				
Common language	-0.99	-0.99	-1.18	-0.42	-7.30***
	(0.496)	(0.494)	(0.384)	(0.794)	(0.000)
Colonial relationship	6.13***	6.15***	6.36***	6.93***	4.82***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ABCC level source country			-0.73***		
			(0.000)		
Source fixed effects	YES	YES	YES	YES	
Destination fixed effects	YES	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES	
Constant	4.31	3.04	75.12***	5.46	7.51**
	(0.392)	(0.583)	(0.000)	(0.369)	(0.029)
Observations	303	312	312	267	312
R-squared	0.64	0.64	0.71	0.65	0.06
chi2					71.22

P-values based on robust standard errors are included in brackets. Migration decades 1820s-1900s are included.. Russia excluded. Sources: see Table 2 and 3.

Instrumental variables: Dummy variable “Social Insurance reforms”, and relative inequality by world region. Migration decades 1820s-1900s are included. Russia excluded. Tests of overidentifying restrictions (IV in col. 5): Sargan (score) chi2(1) 1= .22405 (p = 0.2686), hence we cannot reject the hypothesis that our instruments are valid.

Table 7: Arellano Bond dynamic panel regressions

	(1)	(2)
Estimation Method	GMM	GMM
Lagged selectivity	0.21	0.09
	(0.107)	(0.469)
Relative Inequality dest - source	0.13***	0.11***
	(0.002)	(0.007)
Friends & relatives (Ln stock mig)	0.17	-0.38
	(0.808)	(0.608)
Ln distance		-4.63*
		(0.089)
Poverty constraint (max LGDP - LGDP)		0.89
		(0.907)
Ln distance * poverty constraint		-0.69
		(0.769)
Constant	-3.50***	11.14
	(0.000)	(0.175)
Observations	228	228
Number of codeno	57	57

Migration decades 1820s-1900s are included. Russia excluded. We use the entire lag structure for instrumentation, i.e. starting from the (t-2) lag of the difference for the levels equation, and the (t-1) lag of the level for the difference equations. Arellano-Bond test for AR(2) in first differences. Prob > z: 0.22. The Sargan test of overidentifying restrictions yielded a chi2 of 47.26 (Prob > chi2 = 0.23). Sources: see Table 2 and 3.

5.12 Figures

Figure 1 Panel A : Age distribution of 23 to 82 year old population in the UK, 1881 census

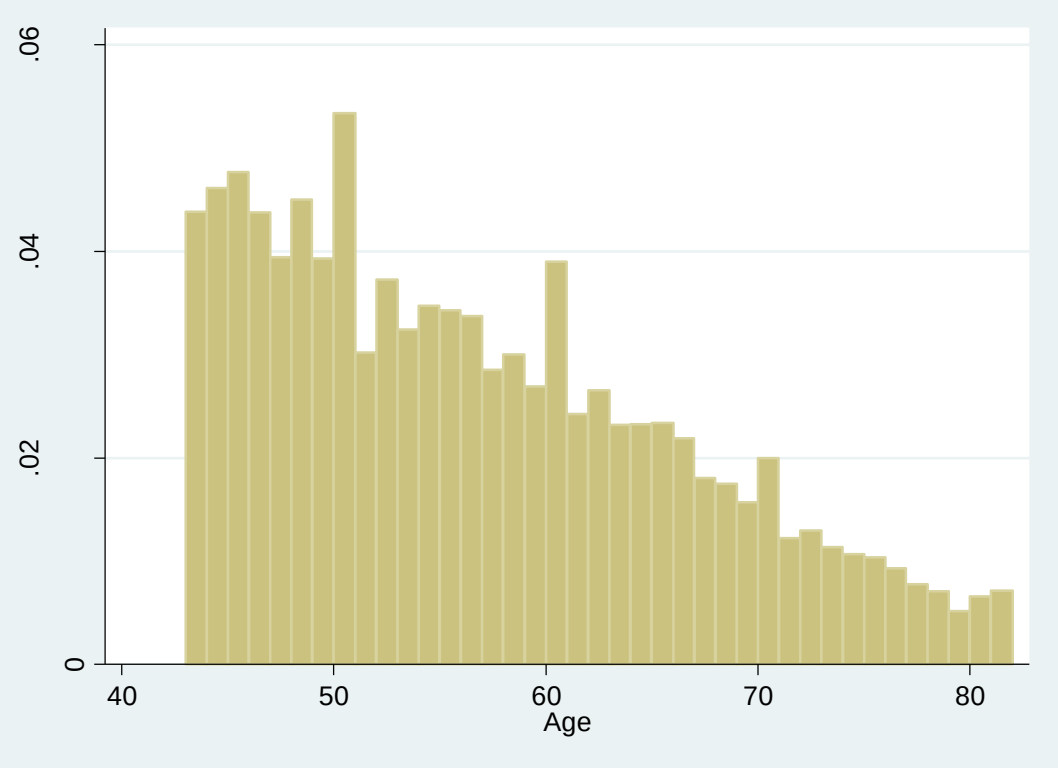
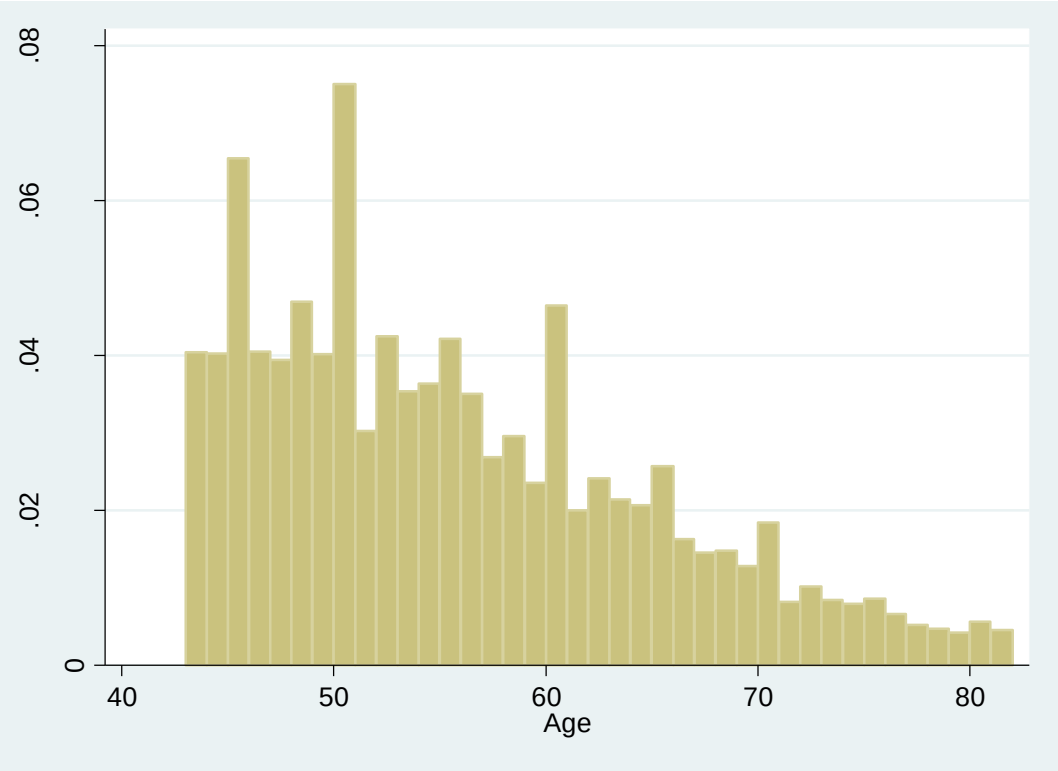


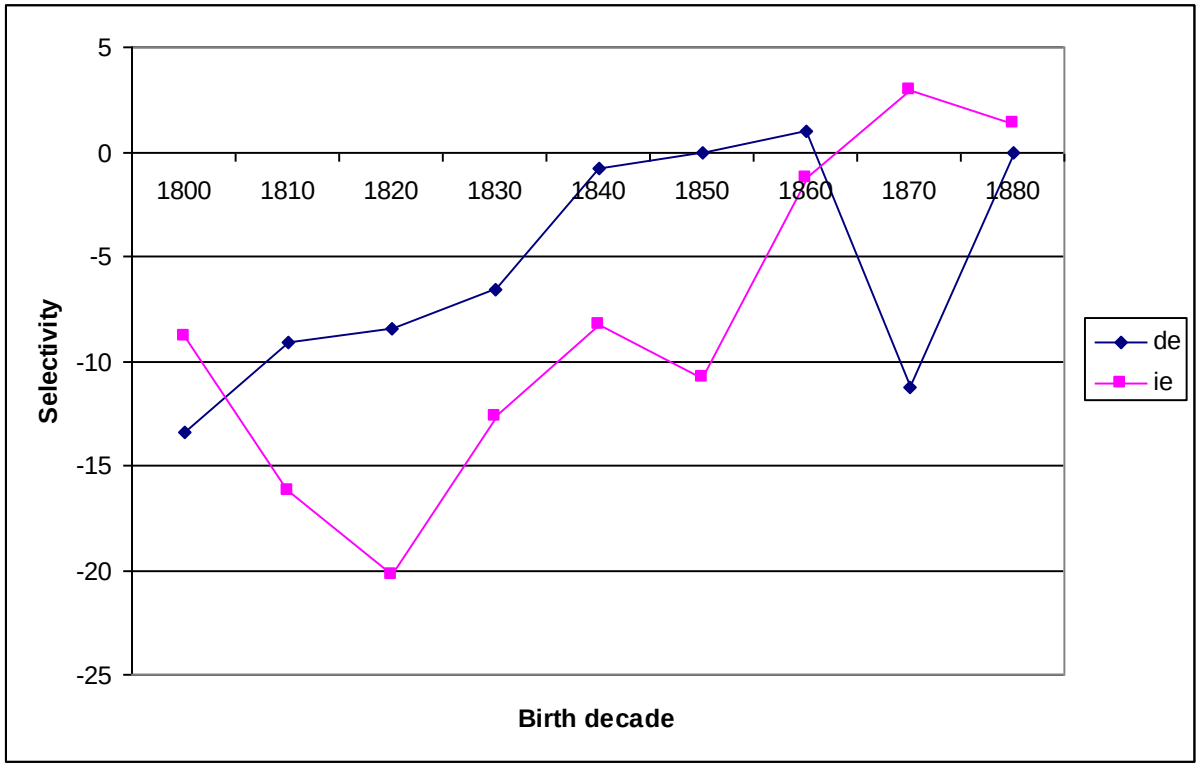
Figure 1 Panel B: Age distribution of 23 to 82 year old immigrant population from the UK, living in the US, 1880 census



Notes to Figure 1 Panel A and B: we performed logit regressions of the migrant status on numeracy by decades. The migrant variable always rendered a negative, highly significant coefficient. Hence, migrants in this panel have a statistically significant lower chance of being numerate than the source country population of the UK.

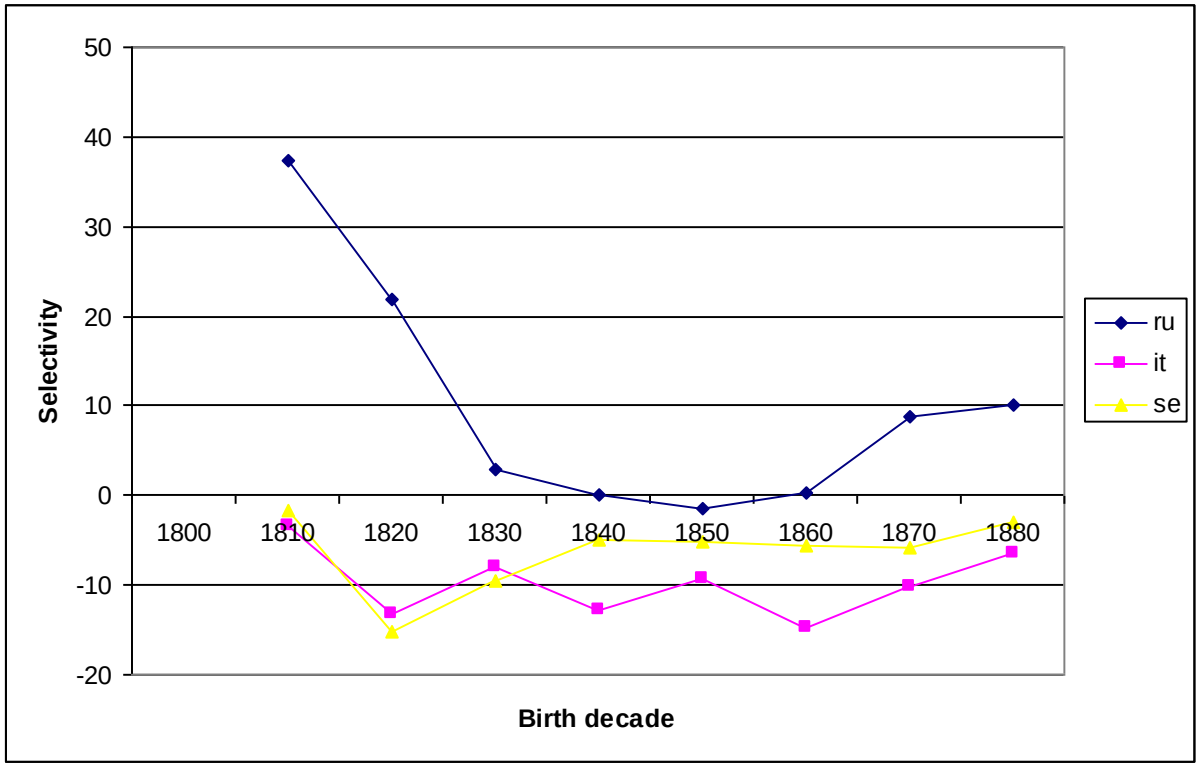
Sources: US census of 1860, Ruggles, Steven, Matthew Sobek, and Trent Alexander, et al. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004

Figure 2: Selectivity among migrants from Germany and Ireland ("old migration countries") to the U.S. by migration decade.



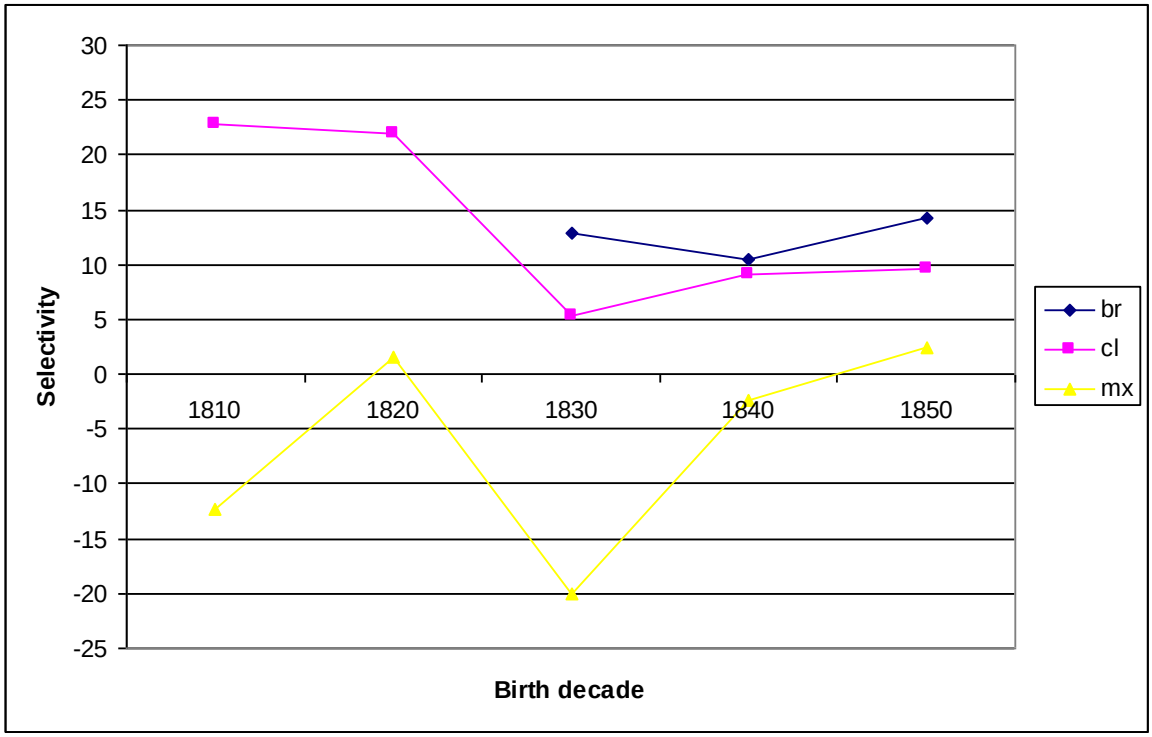
Sources: see Table 2.

Figure 3: Selectivity among migrants from Russia, Italy and Sweden to the U.S. by migration decade.



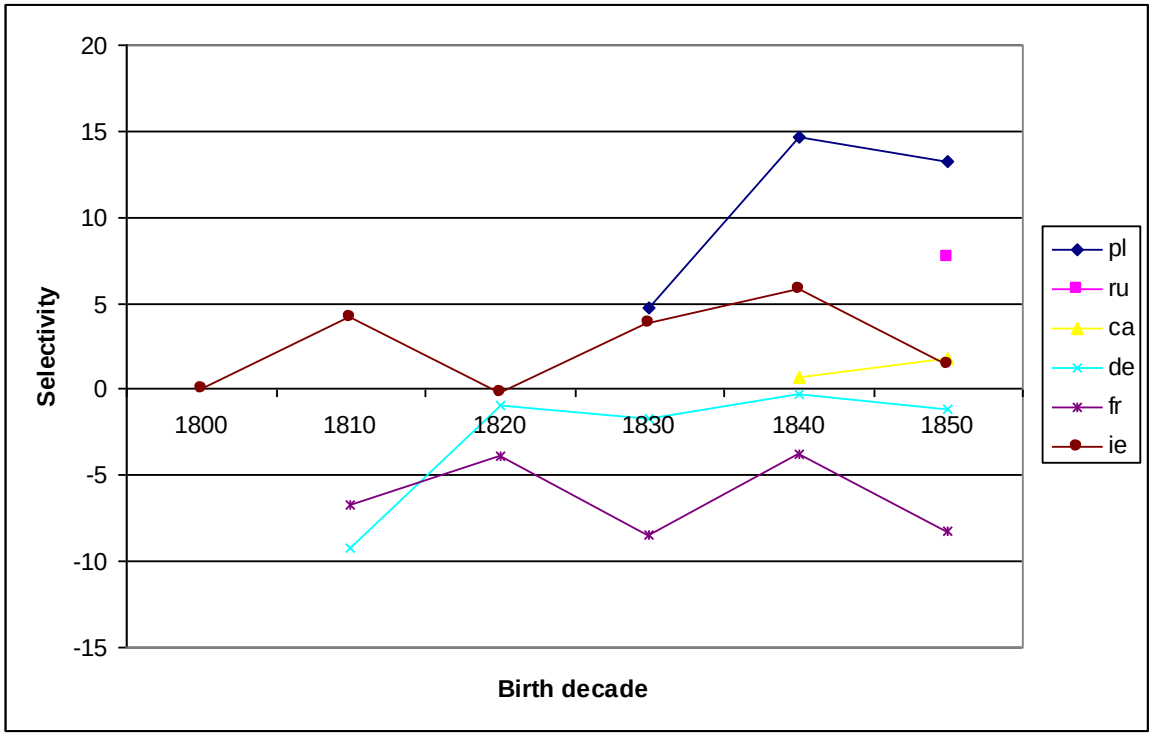
Sources: see Table 2.

Figure 4: Selectivity among Latin American migrants (Brazil, Chile and Mexico) to the U.S. by migration decade.



Sources: see Table 2.

Figure 5: Selectivity among Polish, Russia, Canada, Germany, France, and Irish Immigrants to England by migration decade.



Sources: see Table 2.

5.13 Internet Appendix A: Autocorrelation. Appendix Table 1: Feasible GLS regressions, assuming an AR(1) process

	m1	m2	m3	m4
VARIABLES	GLS	GLS	GLS	GLS
ginidestpremium	0.15**	0.12***	0.13**	0.16***
	(0.014)	(0.002)	(0.029)	(0.003)
lnms	-0.12	-0.03	-0.16	-1.59***
	(0.614)	(0.846)	(0.535)	(0.000)
povconstr	-8.43*	-2.75	-9.22*	0.21
	(0.051)	(0.285)	(0.053)	(0.966)
lndist	-2.20**	-0.94	-2.58**	-4.81**
	(0.023)	(0.119)	(0.022)	(0.021)
lndistpov	1.84	-0.27	2.27	-1.33
	(0.168)	(0.755)	(0.122)	(0.398)
lang	3.34**	2.16**	3.54**	-0.44
	(0.029)	(0.026)	(0.029)	(0.872)
colo	1.35	2.49**	1.51	2.27
	(0.386)	(0.040)	(0.356)	(0.483)
Source FE	YES	YES	YES	
Destination FE	YES	YES	YES	
Time FE	YES	YES	YES	YES
Country Pair FE				YES
Constant	-0.45	-0.76	0.14	18.06**
	(0.925)	(0.827)	(0.980)	(0.022)
Observations	300	300	288	288
Number of codeno	69	69	57	57

P-values based are included in brackets. Migration decades 1820s-1900s are included. Russia excluded. Sources: see Table 2.

5.14 Appenix C: Literacy as an alternative measure of skill selectivity

Interestingly, while some specialized studies have used the occupational structure and age heaping of migrants as indicators, literacy of migrants was not used before. Unfortunately, literacy of immigrants at arrival was only assessed in the U.S. starting in 1899, when the U.S. public grew concerned with the educational status of recent mass immigration from Southern and Eastern Europe, and those lists are not available as individual data sets.

Literacy was also recorded in the censuses between 1850 and 1910, but the comparison between the literacy of immigrants in the U.S. and the population in the source country is difficult for a number of reasons. Firstly, literacy in source countries was recorded using a number of different definitions. Some sources recorded literacy of the adult population, whereas the majority recorded those aged 15 and older, 10 and older or even six years and older.[1] Many statistics report just one number for the whole population which makes it impossible to calculate literacy of age groups or to obtain time series by birth cohorts. Secondly, literacy of individuals coming from different linguistic backgrounds is always difficult to measure. Even if census takers were instructed to record literacy in any language and not only in the official language of the destination country, migrants from different language families could still have declared themselves illiterate when they were asked by census takers. We compared literacy and age heaping from the census data of the different migrant groups in the United States directly. Migrants with a Romanic-language background, namely Italy and Portugal, displayed average numeracy values. However, they had significantly lower literacy rates than one would expect according to their average numeracy. Thirdly, although the vast majority arrived as young adults, a part of the migrants came as children and teenagers to the United States. Already for the mid-19th century, Cohn (2009) reports roughly one quarter arriving as children. When we look at the literacy of persons with migration background in the census some years later, we therefore have to be aware that many of them acquired literacy when they already lived in the United States. So the literacy performance is not only influenced by selective migration but also by age structure and schooling possibilities for migrants. To make things even more complicated, the U.S. was often the destination for migrants coming from countries with lower schooling (Eastern and Southern Europe), but also from countries with better schooling than the U.S., such as Sweden, Norway and so on. The children of those migrant families might have "lost" some of the schooling they would have obtained in their source countries if they had not migrated. Therefore, there exist various biases of different directions which are difficult to quantify. For these reasons, the study of U.S. migrant selectivity based on

literacy is too difficult at the present stage of knowledge. Fortunately, the age heaping techniques provides a feasible alternative to study this important issue.

6. MALE, YOUNG, AND SINGLE? A GLANCE AT FEMALE HUMAN CAPITAL SELECTIVITIES IN 19TH CENTURY MASS MIGRATIONS

6.1 Abstract

Since the 1960s, at least 50 percent of the migration streams around the globe have been composed of women. Based on census data for five major immigrant countries, this study shows that during the 19th century migration, women also made up 47 percent of migrants. So far, the experience of women has been somewhat neglected by scholars. This article investigates human capital selectivity for female migrants from 42 source countries to five destination countries during the 19th century. The model of self-selection (Roy 1951, Borjas 1987) is tested. In the empirical analysis, the inequality – human capital selectivity relationship holds for women, even after controlling for marital status. In a second step, it is shown that determinants for married women are different than those for singles: while for married women, relative inequality is a selectivity driver, for single women, distance and costs are the most important determinants.

6.2 Introduction

Since the 1960s, at least 50 percent of the migration streams around the globe have been female (Pfeiffer et al. 2007). US immigration has already been dominated by women from the 1930s on (Houstoun et al. 1984). However, recent migration research has only begun to revise the view of migration as a non-gendered or predominantly male subject. King and Zotini (2000) even go as far as attesting "gender-blindness" to the field of recent migration research. Hence, "feminization" is one of the key trends of research on international migration (Castles and Miller 1993). Also, Economic History studies have so far paid little attention to the female side of migration during the first global era. Most articles sketch the typical migrant as male, young and single (Hatton and Williamson 1998, Sanchez-Alonso 2008). While it is certainly true that an overwhelming share of migrants was indeed male, young and single, a significant female share of the migration streams has been overlooked mostly (an exception is the Irish case, see Fitzpatrick 1986).

There are different female experiences, depending on the socio-economic and relationship status of women, since female migration ranges from women migrating with husband and family to single women who migrate on their own. In a Portuguese passport list[201], we find, for example, married women, like Maria Pereira aged 34, who travelled with her husband and daughter. Also single women, like Felicidade de Melo, 48, who, according to the very same list, left Portugal in 1893, heading for Rio de Janeiro, Brazil. She was a tailoress, illiterate and travelled with a cousin, perhaps, Ms. Florinda Fernandes, 31, an ironer, who was also single. When married, women are often perceived as "tied movers" which gave reason to some researchers to not pay attention to the female side of the story (Chiswick, 1999). This article provides data on human capital selectivity of both married and single women.

Why is *human capital selectivity of international migrants* important? Due to different individual human capital characteristics, individuals of the same country have different probabilities to migrate. Therefore, it is crucial factor determining who selects to migrate and who does not (Massey et al. 1993). If individuals are positively selected in terms of human capital, the source country might experience a brain drain, since it is the most motivated and skilled citizens who emigrate. This can impact on future growth capabilities of source (Hatton and Williamson 1998, Stolz and Baten 2011) and host countries (Taylor 1997, Stolz, Baten and Botelho, 2011). On the other hand, if individuals from the lower tail of the skill distribution select to migrate, the home country might even experi-

ence a brain gain. A very basic labor economics model explaining occupational self-selection is the Roy-Model (Roy 1951) which was later applied to the migration process (Borjas 1987). The model states a relationship between inequality in source and destination countries and the resulting human capital selectivity of international migrants. It predicts that if inequality in the destination is higher than at home, above average educated individuals select into migration (positive selection) and the opposite vice versa. This is the case, because a more unequally dispersed income allows a higher skill premium at the upper end of the skill distribution as opposed to a society, where high incomes are taxed and redistributed to people with fewer skills.

Female human capital selectivity impacts on human capital stock via two channels. First, women often pertain to different labor markets than men, particularly in historical times. When we take a look at the Portuguese passport list again, we mainly find occupations related to the textile sector like weaver, tailor or sewer, and a lot of domestic servants. Hence, labor markets were highly segmented: not only along the line migrant – native lines, but also according to gender (on segmented labor market theory: see Massey et al. 1993). Second, women historically, and to this day, have to divide their time between productive and reproductive activities. That is, work related to child bearing and service to husband and family. During the reproductive period, it might seem as that human capital of women is not important to the human capital stock of a country, as she is not in the labor market. However, research in the field of development economics has shown that the human capital endowment of women is the crucial starting point for human capital achievements of the next generation. Rosenzweig and Wolpin (1994) use information on ability and achievement test scores of sibling children to test the hypothesis that children's human capital is determined by the schooling attainment of mothers. They find that postponing childbearing by two years in favor of continuing school education increases test scores of their children by almost five percent. In another study, Verner (2005) found that literacy of the parents, in particular the mother, also leads to greater literacy in the next generation. Hence, even when engaging in reproductive work, the human capital endowment of women is important to human capital formation of a society as a whole.

What was the human capital endowment of female migrants in comparison to non-migrant women? What role does marital status play in the process of self-selection? What are the determinants of female human capital selectivity? These are the issues to be investigated in this article. To answer these puzzles, census data for five major immigration

[201] www.museu-emigrantes.org/passaportes-emigrantes.xls, *last accessed February, 22nd, 2011.*

countries and 42 source countries is used and a unique data set of female human capital selectivity is constructed.

The following section displays the relevant literature on female self-selection and derives some testable hypotheses. Then the data is introduced and discussed. Section five provides a first glance at female selectivity trends. Section six displays the empirical results and section seven provides a conclusion.

6.3 Determinants of Female Self-Selection: Testable Hypotheses

Economic theory implies that at the micro-level, a rational actor bases his - or her - migration decision on optimized expected returns from migration (wage gap between source and destination country) after considering the psychological and material costs of such a move (Sjaastad 1962, Todaro 1969, Borjas 1989, Borjas 1994). The probability to migrate increases with expected earnings and decreases with the costs. As a certain amount of "out-of-pocket-money" is necessary to pay for the journey, a budget constraint hinders the very poor from migration even if they have to gain the most by it (Liebig and Sousa-Poza 2004). The human capital endowment of the potential migrant plays a crucial role, because expected earnings and the probability of employment are a positive function of the migrants' human capital endowment. Therefore, before the migrant can evaluate his or her expected earnings, he – or she – has to first consider his or her human capital characteristics and only in the next step, he or she can make an assumption of expected earnings. Furthermore, migration costs may be correlated with human capital characteristics, as any potential migrant with a higher human capital endowment would already earn more in his source country and would therefore more likely be able to afford migration costs. He or she also might find it easier to process information about the destination country beforehand.

At macro-level, the individual migration decisions aggregate to human capital selective migration flows. These can affect the human capital stock of source and destination countries, and therefore have an impact on further growth capabilities of the respective economies, which is why human capital selectivity of the individual migrant is of interest to policy makers in immigration and emigration countries. But what are the drivers of selectivity? In the following, some testable hypotheses are developed which will be tested econometrically.

As determinants of selectivity, firstly, a set of distance and migration cost proxies are included. *Geographical distance* should be of importance, as the move is more costly, the farther away the destination country is. Therefore migration flows are expected to be more

positively selected in the case of long-distance migration than in the short-distance case (Feliciano 2005). The sign of selectivity in the presence of *colonial relationship* and *common language*, both proxies for *cultural distance*, is not clear, however. One would imagine that a common language makes migration easier for people with less education because they do not have to learn a new language. But people with a human capital intensive education might also benefit from going to a country where the same language is spoken, since their very specific skills are better transferable to the destination country's labor market.

Network effects could also play a role via the cost channel, because friends and relatives, already residing in the destination country might provide valuable information and maybe even remittances to finance the journey (Massey et al. 1993, Pedersen, Pytlikova and Smith 2008, Cohn 2009). This leads to an expected negative impact of networks on human capital selectivity.

Destination and source country characteristics should play a role, such as the relative level of economic development. If, for example, the source country is poor in relation to the destination country, more people have an incentive to migrate, but as a *poverty constraint* exists, only those, who can actually afford it, will make the move (Hatton and Williamson 1998). So in the beginning of the migration cycle, a poverty constraint should lead to more positively selected migrants.

Furthermore, the *educational level of the source country* might impact on the degree to which immigrants differ from their source country‘s population in terms of human capital. Migrants from source countries with a generally higher human capital level may be less positively selected, because simply the odds of them being better educated than the average in their homeland is small (Feliciano 2005).

Gender, too affects the individual migration decision in various ways, as prospective earnings and labor market opportunities in destination countries differ for women. For nineteenth century migration, human capital endowment of women should be significantly lower than for men, as women did not have full access to educational facilities, and a more traditional view of gendered division of labor was still common. Gary S. Becker (2002) even states that women's education was generally neglected in historical times.

For women, another decisive factor in the migration decision is *marital status*. Oded Stark (1985) states that marriage and migration are interrelated in such ways that issues like marital stability, fertility and investment in human capital are better understood if migration and marriage are not separately analyzed. The direction of the effect of marital status on human capital selectivity, however, remains unclear. Stark argues that being married

could, in the case of weakly developed market institutions[202], facilitate migration by providing financial support and insurance benefits. Therefore, a married woman migrating with family could be less positively selected than a single woman. Married couples or families with children, however, would face a greater risk, because the family might have to be supported with one wage-earner, only. This would imply that married migrants should be more positively selected than singles, because to make up for a higher risk, there must be more to gain. And only those with a relatively high human capital endowment would have reason to assume a successful integration in the destination country's labor market that would render enough to support a family. Mincer (1978), finally, proposes a model of family migration decisions stating that family ties tend to hinder migration, because they reduce employment probabilities and income of migrating wives. This would strengthen the argument of more positively selected married migrants. To sum up, scholars do not agree on the impact of marital status on human capital selectivity of women. Therefore, marital status is controlled for in the estimations below.

An open discussion is the impact of *relative inequality* on the selectivity of international migrants. Economic theory predicts a relation between the inequality of a source country relative to that of the destination country and the selectivity of their migrants – the larger the perceived inequality in the destination, the higher the probability for skilled people to emigrate– as they will probably experience better returns to skills at their destination (Roy 1951, Borjas 1987, Stolz and Baten 2011). In an empirical study, Brücker and Defoort (2006) find a positive correlation between inequality in the home country and educational selectivity of migrants in the OECD for the period 1980 to 2000 and develop a theoretical model that explains why better skilled people can cope with migration policy hurdles. Feliciano (2005) studies 32 immigrant groups on the US labor market and compares them to their source country's education and inequality level. Her results were not consistent with the Roy model. Belot and Hatton (2009), however, find evidence for a modified Roy model for OECD immigration during the past decades. So did Stolz and Baten (2010) for a large new dataset on the era of mass migrations. Cobb-Clark (1993) studied the inequality – selectivity relationship for women employing a population survey from 1983. She only studied married women and confirms the Roy-Borjas model. Looking only at married women, however, poses a potential problem in so far, as the migration decision of married women is probably not made by them alone but rather in conjunction with their spouses. Hence, a confirmation of the inequality – selectivity relation only for married women

[202] Capital and insurance institutions such as in today's LDCs or in many 19th century societies

might also indicate, that the family focuses on labor market opportunities of the strongest earner. This might probably be the husband, not the wife. This study will therefore test this relationship, for the first time, using a dataset of married and single women to obtain more general results.

6.4 The dependent variable: gendered data on human capital selectivity

Many studies employ earnings as a proxy for skills, as it is the realized returns to skills (Borjas 1987, 1994). This implies that, on the macro level, the skill distribution and the wage distribution in a society must be closely correlated, which is illustrated by the Mincerian human capital earnings function (Mincer 1974). Yet, when it comes to women, the income distribution might not proxy the skill distribution perfectly, because first, women engage in reproductive activities and could therefore not be active in the labor market. Thus, assuming that the majority of women in the 19th century who engaged in reproductive activities were in wedlock, expected earnings are also a function of marital status. Second, if in the labor market, women often specialise in different jobs than men. Hence, labor markets are segmented by gender. Third, within a patriarchal society, even women who engage in the same jobs as men, often still earn less than their male counterparts, even with comparable human capital endowment. Therefore, women in historical studies are difficult to capture with data on earnings or wages, in most cases this kind of data is only available for male laborers (Allen 2001). In this article, human capital selectivity of migrants is measured directly, proxied with an indicator capturing basic human capital endowment, using the age-heaping technique (Crayen and Baten 2009). This has the advantage that data on both single and married women can be constructed.

Numeracy is a relatively new research field in Economic History that employs data on self-reported ages to construct a human capital indicator (A'Hearn, Baten and Crayen 2009, Crayen and Baten 2009, Manzel, Baten and Stolz 2011). Mokyr (1983) pioneered these techniques for the case of Irish migrants. The idea is that in less developed countries of the past, only a certain share of the population was able to report the own age exactly when census-takers or ship personnel asked for it. The remaining individuals reported a rounded age, for example, 30, when they were in fact 29 or 31. The typical result is an age distribution with spikes at ages ending in a five or a zero and an underrepresentation of other ages, which does not reflect the true age distribution. The ratio of rounded or "heaped" ages to the normal age distribution yields the Whipple Index. In this study a linear transformation of which is used: the ABCC Index. Here, 100 stands for all ages are uni-

formly distributed and zero expresses that everybody stated a rounded number. Even today, in underdeveloped countries like Bangladesh or Pakistan, a pattern of heaping at multiples of five can be traced in census data. Hence, the age-heaping approach is a suitable human capital indicator for societies and individuals at a very basic stage of human capital accumulation, which used to be the case in the 19th century. The indicator correlates highly with other human capital proxies, e.g. literacy. One of its main determinants is schooling (Crayen and Baten 2010). The age-heaping approach has recently been successfully applied to the question of migrant selectivity (Stolz and Baten 2011), where the authors confirmed the Roy Model (1951) of self-selection in a panel for five major immigration countries and 52 source countries.

Human capital selectivity is constructed as a differential between the average numeracy level of female migrants and the average numeracy of female source country population per decade:

$$S_{ijt} = \text{Female Migrants}_{ijt} - \text{Female (Migrants + Stayers)}_{it}$$

Where S_{ijt} is the selectivity of all woman observed in destination country j from source country i in migration decade t. In the second part of the analysis, the selectivity variable distinguishes women in the destination country by marital status. This average is related to the average human capital of women in source country i in migration decade t. The source country data does not allow distinguishing women by marital status. It is assumed that there is no systematic human capital bias regarding the marriage decision, which will be explained in further detail, below. Further, it is assumed that migration took place in the second decade of life.

Census data to cover the destination countries is taken from the North Atlantic Population Project (NAPP) and IPUMS Project. Both projects provide census data on Canada, Great Britain, Germany, Norway and the US.[203] For Argentina, data from the population censuses 1869 and 1895 is used. Comparable sources for five immigration countries and 42 source countries were collected from which the information on age, gender and marital status were drawn. All in all, observations by destination and source country pairs and decades are constructed from more than 7.9 Million underlying individual observations. The biggest share stems from US immigration as the NAPP team provides 100 percent samples of US censuses. From all other countries, also sufficient observations for the

[203] Minnesota Population Center. North Atlantic Population Project: Complete Count Microdata. Version 2.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center, 2008. And: Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

study could be generated. Within the data at hand, women make up 47 percent of the immigrants who went to the UK, and more than 45 percent of the immigrants who went to the US (42 to Canada). Argentina received the most masculine immigrant flows with only 39 percent of women in the data. Nevertheless, all in all, women made up roughly between a third and almost one half of the total immigrant population. This stresses the importance of research in a gender-aware perspective.

What about marital status? One has to bear in mind that the data on migrants is taken from population censuses within the destination country. Therefore, the evidence does not show whether marriage took place before or after migration. Nevertheless, a look at the census data reveals that 53 percent of all migrants, men and women, are married. In the US it is 71 percent, Canada 70 percent, Norway 67 percent and the UK 62 percent. It seems that the comparably low number for Argentina stems from the huge share of "*Golondrina*" immigration that is single males who went back and forth between Argentina and Italy as seasonal workers (Hatton and Williamson 1998). A look at the 1880 US census shows, that even among the youngest age group (23-32 years) nearly 55 % of the immigrants were married which implies that at least a high probability exists for immigrants to having had married before starting the journey.

The source-country data stems from a data set generated by Crayen and Baten (2009), which was further enlarged by Manzel, Baten and Stolz (2011) and Prayon and Baten (2011). These authors provide numeracy estimates from census data for a broad range of countries.

Is the numeracy indicator robust to capture human capital selectivity of female migrants? Peter Földvári, Bas van Leeuwen and Jieli van Leeuwen-Li (2011) have recently raised some doubts on the applicability of the numeracy concept for women. They argue that married women would adjust their age statements to the heaping of their spouses, thereby systematically biasing numeracy estimates of married women upwardly. This is unlikely for several reasons. First, basic numeracy – age awareness - is usually acquired in the first decade of life while marriage takes place in the second or third decade of life. Hence the authors neglect the chronology of these events. Human capital endowment can, of course, influence success in the marriage market, but it is unlikely that marriage influences the age awareness of women. Secondly, the authors rule out that women who are successful in the marriage market often display better human capital endowment than those who do not succeed in finding a spouse. Their argument, therefore, ignores a literature pioneered by Gary S. Becker (1972) who states, that success on the marriage market might depend, ceteris paribus, positively on the human capital endowment of the spouses. Empir-

ical studies on the topic find human capital characteristics of the spouses to be highly correlated (Alstrom 1961). This does not necessarily mean that married people display better human capital or education per se. Boulier and Rosenzweig (1984) tested the economic theory of marriage using micro data on the Philippines and found that for women, additional schooling attracted a spouse with more earning potential.

Now turning to historical literature, Manzel and Baten (2009) find that in 19th century Latin American societies, education of women was often neglected, or only important as a means to better fulfil their conjugal duties. Still, women had to learn how to manage a household and maybe advise personnel and do the groceries. For example, in German 19th century middle-class, husbands often also expected some "educated competence" of their wives in public (Nipperdey 1994: 53). Fitzpatrick (1986) investigated the case of Irish women in the 19th century. He finds that the structural change in the Irish economy reduced labor market possibilities for women. At the same time an increasingly more competitive marriage market led to a worsening bargaining position of women versus men in the Irish society. The female response to that development was twofold: first, women took advantage of the spread of schooling institutions and invested in better education. The second option was emigration. Often, the acquisition of literacy was seen as a way of improving emigration possibilities. As a result, by the beginning of the 20th century, women were even marginally more literate and mobile than their male compatriots (Fitzpatrick 1986: 218).

Hence, ex-post married women would be those who were ex-ante attractive for men in a vector of market and non-market features. Human capital is one dimension within this vector, however not the only one. Nevertheless, now turning to the data at hand, additional light on the question is shed by looking at female human capital selectivity for some countries by marital status. Figure 1a and 1b display female migrants from Eastern Europe in the US by marital status. The graph does not indicate any specific selectivity pattern for singles and married women. Both, married and single women display positive selection (Russia, Poland) or negative selection in some cases (Hungary, Czech Republic). A look at a box plot shows, however, that the mean of the selectivity variable is lower for singles than for married women (Figure 1c). The standard deviation of the distributions is 10.24 for married female migrants and 10.31, respectively. Both married and single emigrants display worse numeracy than the average female non-migrant population in the source countries. Thus female migrants are, on average, negatively selected. Married female emigrants however, are a little less negatively selected than their single compatriots.

6.5 Female human capital trends in the US, Canada and Argentina

A look at the trends for human capital selectivity of women reveals that, now taking married and single women together, female immigrants to the US from South Europe were negatively selected from the source country population, except for Spanish immigrants in the very first period of the Spanish immigration wave to the US. Figure 2 is organized by migration decades, assuming that most migrants migrated in their 20s.[204] The Spanish, also the Italians, and the French and, to a lesser extent, the Portuguese reveal at first more positively selected immigrants. Later, immigrants are less favorably selected from their home country population. Particularly during the birth decades of the 1840s to the 1860s, the Portuguese and the Italians show a very negative selectivity. Due to the utility maximization of the individual, one would expect positively selected migrants in the first migration waves, as first movers cannot rely on already existing family-and-friends-networks which might provide valuable information and help and might diminish migration costs. Negative selection around the 1840s birth decade is particularly interesting. Their negative selection might be a reflection of the "hungry forties" all over Europe, a period with bad harvests that caused many people to suffer and die from hunger in the Old World and caused, particularly in Ireland, a great emigration wave (O'Gráda, Vanhaute and Paping 2007). For Portugal, the period of agricultural disasters went for additional two decades, causing a severe retardation of Portuguese living standards in comparison to the European standard (Stolz, Baten and Reis, 2011).

Figure 3 shows selectivity trends for Canada and exhibits some similar facts: higher selectivity for the first migrants, and then a decline, in some cases a drop in the 1840s decade (note the Irish, for example). French immigrants are positively selected, while UK immigrants are negatively selected. Noteworthy are the Russian immigrants who are strikingly more positively selected than the other immigrant groups. This could be due to the pogroms prevailing in Russia after the assassination of the Czar, which lead many Jews to migrate into the New World (Chiswick 1991). Chiswick (1991) highlights the labor market success of turn-of-the-century Jewish immigrants in the US. He finds this related to their good human capital endowment and general favorable attitude towards education.

Argentina displays different trends (Figure 4): all but German women are negatively selected in comparison to the source country population. Intra-Latin American migration seems worst selected. A fact that could be explained by low migration costs from Brazil to neighbouring Argentina. The stark negatively selected Italians are well documented in the

literature for men (Hatton and Williamson, 1998). But also Italian women show a negative selectivity, deteriorating over the course of time, thereby providing evidence, that not all of the *Golondrina* migration was male young and single (Hatton and Williamson 1998, Sanchez-Alonso 2008).

6.6 Empirical Results

The empirical equation underlying the estimations is the following:

$$S_{ijt} = \alpha + \beta_1 Inequ_{ijt} + \beta_2 Numeracy_{it} + \beta_3 log\,(Mig_{j,\,t-1}) + \beta_4 PovConst_{it} + \beta_5 Log\,(Dist_{ij}) + \beta_6 PovConst_{it} * Log\,(Dist_{ij}) + \beta_7 Languag_{ijt} + \beta_8 Colony_{ijt} + \beta_9 PovConst_{it} * Log(Mig_{j,\,t-1}) + \beta_{10} Dem_{ijt} + \beta_{11} Sinlge\ Woman + Fixed\ Effects_j + Fixed\ Effects_i + Fixed\ Effects_t\ [+ CountryPair\ FE]$$

For the explanatories, the following data were used: *Relative inequality* is the variable that is important when it comes to the determination of the self-selection model of Roy (Roy 1951), is proxied using an anthropometric approach. Inequality is proxied with Gini-coefficients. These are partly estimated using height Coefficients of Variation (CVs) for which a new data set has become available recently (Blum and Baten 2011).[205] The methodology is described in detail in Stolz and Baten (2011). The *educational level of the source country* is controlled for by including the ABCC-level of the source country (Crayen and Baten 2009, Prayon and Baten 2011). *Network effects* – a possible friends-and-relatives-effect is introduced quantitatively by calculating the share of migrants of a particular source country over all migrants observed in the destination country out of the census data (IPUMS, NAPP). This is then lagged by one decade to capture the dynamics of the effect.[206] A *poverty constraint* is constructed by subtracting the source countries per capita GDP from the global maximum GDP per capita during the respective decade. Data is taken from Maddison (2009) and Baten and Blum (2010).[207] The variable should turn significant, if human capital selectivity of migrants was income constrained. Also an *inter-*

[204] This is reasonable to assume, as most studies find an average aged first global era migrant to be around 24 years of age (Sanchez Alonso 2008).

[205] Special thanks to the authors for sharing this data with me.

[206] Curran and Rivero-Fuentes (2003) suggest distinguishing the gender composition of migrant networks, because men and women are differently affected by male and female migrant networks. Therefore a female migrant stock variable was calculated as the share of female migrants over the total migrants from a specific source country in migration decade-1. This variable, however, did not render any significant results, which is why it was dropped from the analysis.

[207] See also: Stolz and Baten (2011). Baten and Blum (2011) use a dataset of heights to estimate Gdp Values. These estimates were taken wherever Maddison did not provide data.

action term with distance is introduced, as a transatlantic journey would require a higher travel budget than intra-European migration, only. On 19th century migration costs, Sanchez-Alonso (2000: p. 321), for example, states that around the end of the 19th century the ticket price from Galicia to Argentina was around 321 pesetas. Wegge (2009: p. 12) finds a price of 50 or 60 Thaler during the 1850s from Hesse-Cassel, Germany. Keeling 2009 provides some estimates of transoceanic fares and migration costs. All in all, however, the studies on overseas migration are scattered and difficult to compare, on the case of Intra-European migration, not much is known. *Geographical distance* is therefore proxied with the log distance between the two capitals of source and destination countries. To control for *cultural distance,* a dummy variable that turns one if the same language is spoken at home as in the receiving country, is included. Another dummy turns one if a colonial relationship existed at some point.[208] The *distance variable is also interacted with the migrant stock* because the existence of a diaspora community in the destination country might affect long- and short distance migration differently. *Relative democracy* is a political variable that controls for potential political push factors.[209] Furthermore, a variable that turns one when a *civil war* was going on in the source country is included.[210] The information on *marital status* for the female migrants is also retrieved from the individual census data and then converted into a dichotomous variable that turns one, if a woman in the data set was not married.

To control for further *unobserved source or destination country specific effects* and also for unobserved effects that are inherent in specific country pairs, a full set of fixed effects is included. Alternatively, a set of source-destination country combinations is used to control for specific unobserved country combinations. This approach does not change the main results.

The descriptives of all variables are displayed in Table 1. A covariance matrix shows that there exist no collinearity problems. The next chapter will provide a first look at the trends of female human capital selectivity in three major destination countries in the Americas.

Firstly, a general model is estimated with all women, controlling for marital status (Table 2). As the Russian migration flow was probably positively selected due to political and religious reasons, columns 2 to 6 are estimated excluding the data on Russia mi-

[208] The data is taken from http://www.cepii.fr/, last accessed April, 18th, 2011.
[209] Polity IV project, see Marshall, Monty G., and Jaggers, K. (2008): Polity IV Project: data set. http://www.systemicpeace.org/polity/polity4.htm, last accessed February, 22nd, 2011.
[210] Correlates of War Project, see Singer, J. David and Melvin Small (1972).

grants.[211] The last column applies country pair dummies instead of controlling for source and destination country unobserved effects separately (Table 2, column 6). Next, separate models for single and married women are estimated to sort out differences in selectivity drivers according to marital status (Tables 3 and 4).

Table 2 displays the results for the *general model of the determinants of human capital selectivity of women*. Throughout all specifications, relative inequality becomes significant, although only at the five or ten percent level. This variable has a positive sign and suggests that, if inequality, measured in Gini-coefficients, is higher in the destination country than at home, women display –ceteris paribus- a higher numeracy than the female non-migrant population. This result goes in line with recent findings of Baten and Stolz (2011), who already confirmed this relationship for men and women jointly.

Marital status is significant in all specifications. Single women who select to migration are, ceteris paribus, around 5 ABCC-points less numerate than married women. Put differently, migrating wives exhibit better human capital characteristics than their single compatriots. The result confirms what was already obvious from the raw distribution of the selectivity variable (Figure 1c). Due to the way the selectivity variable is constructed, it is not clear, if the effect stems only from migrant selectivity or also from an underlying selectivity of married women. A probability exists, that married women display better human capital characteristics than singles per se, because they are more successful on the marriage market. At least part of the effect, however, certainly stems from human capital selectivity of migration, which is reasonable, given that also other migration variables also explain some of the variation in the selectivity variable. The result would indicate that family ties hinder migration in such a way that only those who are better endowed with human capital actually migrate (Mincer 1978). If this is due to better labor market opportunities or simply a better capability to process information beforehand, can only be speculated about.[212]

[211] See Bade (2008) or Stolz and Baten (2011): due to pogroms of the Jewish population, many highly educated Jews left the Russian empire.

[212] The effect of marital status on human capital selectivity is similarly large and significant in a panel for men only – this shows that not only married women but also married men are more positively selected than singles who migrate:

	(1)
VARIABLES	Selectivity
Migration Control Variables	YES
Single	-3.68***
	(0.000)
Constant	55.39***
	(0.000)
Time FE	YES
Source FE	YES

The numeracy level of the source country is also significant in all estimations, indicating, that the level of selectivity is negatively associated with the educational level at home. This confirms results of Feliciano (2005) who states that migrants from source countries where high levels of schooling exist are likely to be less positively selected.

Common language becomes significant in two of the specifications. When destination and source country have the same language migrants are around 5 ABCC-points more numerate than the home country population (columns (2) and (5)). All of the other variables do not render significant results in the general model.

Next, tables 3 and 4 display the results for married and single women in two separate panels: now, it becomes obvious that self-selection into migration is a different situation for married than for single women.

Married women's selectivity is determined by relative inequality (Table 3). The variable plays a consistent role in all specifications, confirming the results of Cobb-Clark (1993), who also finds this for a panel of married women. The Roy-Model of self-selection holds in this data set. The numeracy level of the source country also exhibits significant results, indicating that women from a source country with higher numeracy on average are less positively selected. The interaction term between the poverty constraint and log migrant stock turns significant in one of the three specifications. Thus a higher living standard at home in combination with a larger diaspora community in the destination country lead to less positively selected married female migrants. Relative democracy is weakly significant in one specification, indicating that a more democratic regime in the destination country in comparison to the home country attracts less positively selected married women.

For *Single women* the results are different: they are more positively selected when the same language is spoken in the destination country than at home (Table 4). The magnitude of the variable is large: women are about 8 ABCC points higher than the average woman at home, when the same language is spoken in the destination and in the source country (ceteris paribus). Furthermore, the distance variable becomes significant, although with a negative sign, implying that long-distance migration attracted less favorably selected single female migrants. This could be due to the fact that the majority of the variance in

Destination FE	YES
Observations	528
Number of cdsno	110

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

the distance variable stems from intra-European versus trans-Atlantic migration. Actually, intra-European migration seems to be better selected than long-distance migration in the data set provided, here. This is somewhat surprising, as one would expect that the people who could afford the transatlantic journey would be those from the upper part of the income distribution, since long-distance migration requires more resources. Assuming that income is closely correlated with human capital, they should also be better endowed with human capital than the average population.

The results of these empirical analyzes provide some support to the reasoning of Mincer (1978) who argues that family ties hinder migration. This conclusion is drawn because family migration seems, according to the significant impact of relative inequality on the human capital selectivity of wives, more sensitive to labor market incentives and hence to relative skill premia. It seems that, only those wives who could actually afford the journey and had a positive perspective of being successful in the receiving country actually migrated. If labor market incentives are a significant determinant of married women, that implies that returns to skills are at their focus when deciding whether to migrate or not. I.e. they only migrate, when they assume that returns to skills in the destination country are higher than in the source country. This would indicate that family migration is perceived as a high risk and only those, who are sure to be able to succeed in the destination country's labor market, actually make the move. This could also mean that the returns to skills of the husband are actually at the focus of the migration decision, provided that skills of spouses are correlated. It is not clear in how far this result is driven by the abilities of the husband, then. Since family migration decisions are taken collectively and also the relative inequality variable is a macro variable that does not distinguish gender, it could be that married women seem to react upon relative inequality because they maximize their husband's prospective income. In most cases, husbands might have been the major provider of the family. As migrating married men also exhibit higher numeracy than singles (see footnote 242), the conclusion can be drawn that, notwithstanding the fact that migrants on average are negatively selected, migrating families are less so, probably because migrating as a family is riskier a endeavour than migrating as a single person. As a single woman, mainly distance and common language are of importance, i.e. variables that capture migration costs. This indicates that for singles, mostly the immediate costs of migration are at the focus, maybe because singles are more flexible and not responsible for other people and assume therefore to adjust rapidly in the new environment.

6.7 Conclusion

This article constructed a new and unique dataset of 19th century women who migrated to five different destination countries in Europe and the Americas. All in all, evidence on 42 source countries was provided. For the first time, empirical data on the female side of mass migration was compiled and analyzed here, including human capital selectivity of both transatlantic and intra-European migration flows. It could be shown that a significant share of 19th century migration streams was female. Thus this paper fills a research gap on the female side of the first global era. Female human capital selectivity was measured applying the age-heaping approach. On average, female migrants were less numerate than the source country population. This negative selectivity is more pronounced for married women than for singles. The determinants of human capital selectivity were estimated and the results show that selectivity drivers for married and single women differ. While for married women, relative inequality renders results consistent with the Roy-model of self-selection (Roy 1951, Borjas 1987, Cobb-Clark 1993), for single women the picture is different. Here, mainly cost-proxies like distance and common language are of importance. The analysis confirms Stark's (1985) view that marriage and migration are interrelated and should therefore be investigated jointly.

6.8 References

A'Hearn, Brian, Baten, Joerg Crayen, Dorothee (2009): "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital" *Journal of Economic History* 68, no. 3, pp. 783-808.

Allen, Robert C. (2001): "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War" *Explorations in Economic History* 38, pp. 411-447.

Alstrom, C. H. (1961): "A Study of Inheritance of Human Intelligence" *Acta Psychiatrica et Neurologica Scandinavica*.

Bade, Klaus, J., Emmer, Pieter C., Lucassen, Leo and Jochen, Ottmar (eds): *Migration in Europa. Vom 17. Jahrhundert bis zur Gegenwart*. Paderborn.

Baten, Joerg and Blum, Matthias (2010): "Global Height Trends and its Determinants, 1810-1984: An Overview", see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/baten_blum_ht_100331a.pdf, last accessed March 31st, 2010.

Belot, M.V.K., Hatton, Tim J. (2009): "Immigrant Selection in the OECD" Center for Economic Policy Research. Discussion Paper No. 571.

Blum and Baten (2011): "Estimating Inequality with Anthropometric Indicators" *Review for Economics – German "Jahrbuch für Wirtschaftswissenschaften"*, forthcoming.

Borjas, George J. (1987): "Self-Selection and the Earnings of Immigrants." *The American Economic Review* 77, no. 4, pp. 531-553.

Borjas, George J. (1994): "The Economics of Immigration", *Journal of Economic Literature* 32, no. 4, pp. 1667-1717.

Boulier, Bryan L. and Rosenzweig Marc R. (1984): "Schooling, Search and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior" *The Journal of Political Economy* 92, no. 4, pp. 712-732.

Brücker H, Defoort, C. (2006): „The Self-Selection of International Migrants Reconsidered: Theory and New Evidence" IZA Discussion Paper Series IZA DP 2052.

Castles, S. and Miller, M.J. (1993): *The Age of Migration: International Population Movements in the Modern World*. London: Macmillan; New York: Guilford Books.

Chiswick, Barry, R. (1991): "Jewish Immigrant Skill and Occupational Attainment at the Turn of the Century", *Explorations in Economic History* 28, pp. 64-86.

Chiswick, Barry, R. (1999): "Are Immigrants favorably self-selected?" *The American Economic Review* 89, no.2, pp. 181-185.

Cobb-Clark, Deborah, A. (1993): "Immigrant Selectivity and Wages: The Evidence for Women" *The American Economic Review* 83, no. 4, pp. 986-993.

Cohn, R.L. (2009) *Mass Migration under Sail: European Immigration to the Antebellum United States.* Cambridge University Press. Cambridge.

Coppel, J., Dumont, J.-C., and Visco, I. (2001): "Trends in immigration and economic consequences" OECD Economics Department Working Paper, (284).

Crayen, Dorothee and Baten, Joerg (2009): "Global Trends in Numeracy 1820–1949 and its Implications for Long-term Growth", *Explorations in Economic History* 47, no. 1, pp. 82-99.

Crayen, Dorothee and Baten, Joerg (2010): "New Evidence and New Methods to Measure Human Capital Inequality before and during the Industrial Revolution: France and the US in the Seventeenth to Nineteenth centuries" *Economic History Review*, forthcoming. Published Online: Jul 29 2009 10:50AM. DOI: 10.1111/j.1468-0289.2009.00499.x

Curran, Sara R. and Rivero-Fuentes, Estela (2003): "Engendering Migrant Networks: The Case of Mexican Migration" *Demography* 40, no. 2, pp. 289-307.

Feliciano, Cynthia (2005): "Does selective migration matter? Explaining ethnic disparities in educational attainment among immigrant children" *International Migration Review* 39, no. 4, pp. 841-871.

Fitzpatrick, David (1986): 'A share of the honeycomb': Education, emigration and Irishwomen, *Continuity and Change* 1, no. 2, pp. 217-234.

Földvári, Peter, van Leeuwen, Bas and van Leeuwen, Jieli (2011): „How did Women count? A note on gender-specific age heaping differences in the 16th to 19th centuries" *Economic History Review*, forthcoming.

Gary S. Becker (2002), "Human Capital," The Concise Encyclopedia of Economics, http://www.econlib.org/li brary/enc/HumanCapital.html, last accessed April, 28th 2011.

Hatton, T.J., Williamson, J.G. (1998): *The Age of Mass Migration: Causes and Economic Impact*. Oxford University Press. New York.

Houstoun, Marion F., Kramer Roger G., and Joan Mackin Barrett (1984): "Female Predominance in Immigration to the United States since 1930: A First Look", *International Migration Review* 18, No. 4, Special Issue: Women in Migration, pp. 908-963.

Keeling, Drew (2011): "Costs, Cycles and Causes of Migration across the North Atlantic, 1870-1914", University of Zuerich Working Paper.

Liebig, Thomas and Sousa-Poza, Alfonso (2004): "Migration, Self-Selection and Income Inequality: An International Analysis" *Kyklos* 57, pp. 125-146.

Maddison, Angus (2007) World population, GDP and per capita GDP, 1-2001 AD, [WWW document]. URL http://www.ggdc.net/maddison/, [accessed on March 31st, 2011].

Manzel, Kerstin, Baten, Joerg (2009): "Gender Equality and Inequality in Numeracy: The Case of Latin America and the Caribbean, 1880–1949" *Revista de Historia Económica (Second Series)* 27, pp. 37-73.

Manzel, Kerstin, Baten, Joerg, and Stolz, Yvonne (2011): "Convergence and Divergence of Numeracy: The Development of Age-Heaping in Latin America, 17th to 20th Centuries" *Economic History Review*, forthcoming.

Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor (1993): "Theories of International Migration: A Review and Appraisal", *Population and Development Review* 19, no. 3, pp. 431-466.

Mincer, Jacob A. (1974): "The Human Capital Earnings Function" NBER Chapters, in: *Schooling, Experience, and Earnings*, National Bureau of Economic Research, Inc. pp. 83-96.

Mincer, Jacob A. (1978): "Family Migration Decisions" *The Journal of Political Economy* 86, no. 5, pp. 749-773.

Mokyr, Joel (1983): *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850* (London and Boston).

Nipperdey, Thomas (1994): *Deutsche Geschichte 1866-1918. Band I: Arbeitswelt und Bürgergeist.* Verlag C.H. Beck. München.

O'Gráda, Cormac, Vanhaute, Eric and Paping, Richard (2007): "The European subsistence crisis of 1845-1850. A comparative perspective" In: *When the potato failed. Causes and effects of the 'last' European subsistance crisis, 1845-1850*, pp. 15-26.

Pedersen Peder J., Pytlikova, Mariola, and Smith, Nina (2008): "Selection and network effects – Migration flows into OECD countries 1990-2000" *European Economic Review* 52, no. 7, pp. 1160-1186.

Pfeiffer, Lisa, Richter, Susan, Fletcher, Peri and Taylor, Edward J. (2007): "Gender in Economic Research on International Migration and Its Impacts: A Critical Review" *The International Migration of Women*, World Bank, pp. 11-51.

Prayon, Valeria and Baten, Joerg (2011): "Human Capital, Institutions, Settler Human Capital, and Economic Growth in Africa, Asia and the Americas" University of Tuebingen Working Paper.

Rosenzweig, MR, Wolpin, KI (1994): "Are there Increasing Returns of the Intergenerational Production of Human Capital? Maternal Schooling and Child Intellectual Achievement" *The Journal of Human Resources* XXIX, no. 2, pp. 670-694.

Roy, A. (1951): "Some thoughts on the distribution of earnings." *Oxford Economic Papers 3*, pp. 135–46.

Sachnez-Alonso, Blanca (2000):" European emigration in the late nineteenth century: the paradoxical case of Spain" *Economic History Review* 53, no. 2, pp. 309-330.

Sanchez-Alonso, Blanca (2008): "The Other Europeans: Immigration into Latin America and the International Labor Market, 1870-1930" *Revista de Historia Económica/ Journal of Iberian and Latin American Studies* XXV,no. 3, pp. 395-426.

Sjaastad Larry A. (1962): "The Costs and Returns of Human Migration Part 2: Investment in Human Beings" *The Journal of Political Economy* 70, no. 5, pp. 80-93

Stark, Oded (1985): "On Marriage and Migration" *European Journal of Population* 4, pp. 23-37.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Stolz, Yvonne and Baten, Joerg (2011): "Brain Drain, Numeracy and Anthropometric Inequality during the Era of Mass Migration: Testing the Roy-Borjas Model with New Indicators" University of Tuebingen Working Paper.

Stolz, Yvonne, Baten, Joerg and Botelho, Tarcísio (2011): "Growth effects of 19th century mass migrations: "Fome Zero" for Brazil?" University of Tuebingen Working Paper.

Stolz, Yvonne, Baten, Joerg and Reis, Jaime (2011): "Portuguese Living Standards 1720-1980: Heights, Income, and Human Capital" University of Tuebingen Working Paper.

Taylor, Alan M. (1997): "Peopling the Pampa: On the Impact of Mass Migration to the River Plate, 1870-1914" *Explorations in Economic History* 34, no. 1, pp. 100-132.

Todaro, Michael P. (1969): "A Model of Labor Migration and Urban Unemployment in Less Developed countries" *The American Economic Review* 59, no. 1, pp. 138-148.

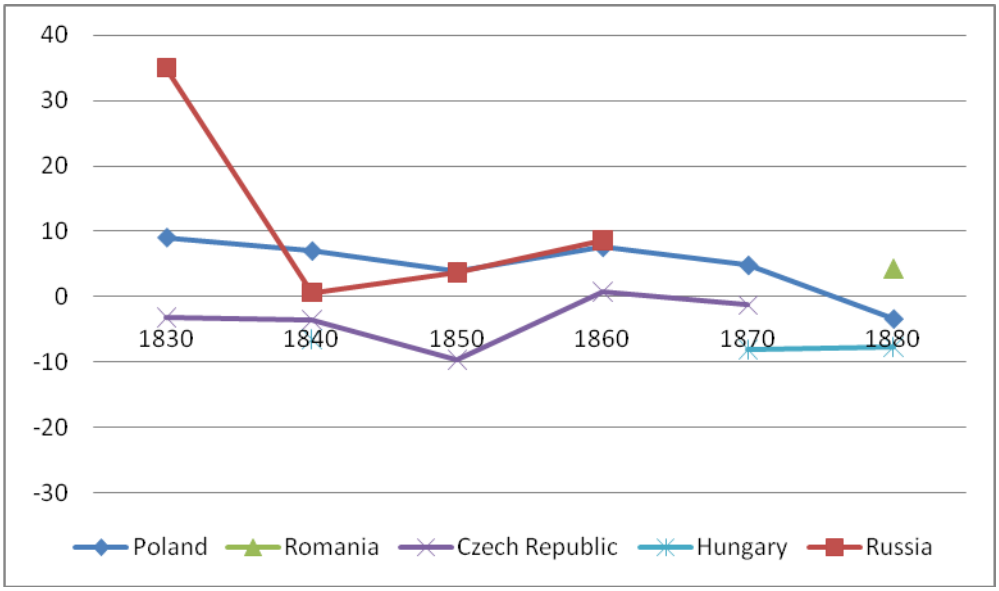
Van Zanden, Jan Luiten., Baten, Joerg, Foldvari, Peter, Leeuwen, Basv., (2009) "World Income Inequality 1820-2000" Working Paper University of Tuebingen, Dept. Economic History No. 87.

Verner, Dorte (2005): "What factors influence world literacy? Is Africa different?" Policy Research Working Paper Series 3496. The World Bank.

Wegge, Simone (2009):" Do migrant origins matter? Migrants vs. Non-Migrants: Self-Selection of Migrants Out of the Home Population" City University of New York Working Paper.

6.9 Figures

Figure 1a: Human capital selectivity of married immigrant women in the US.

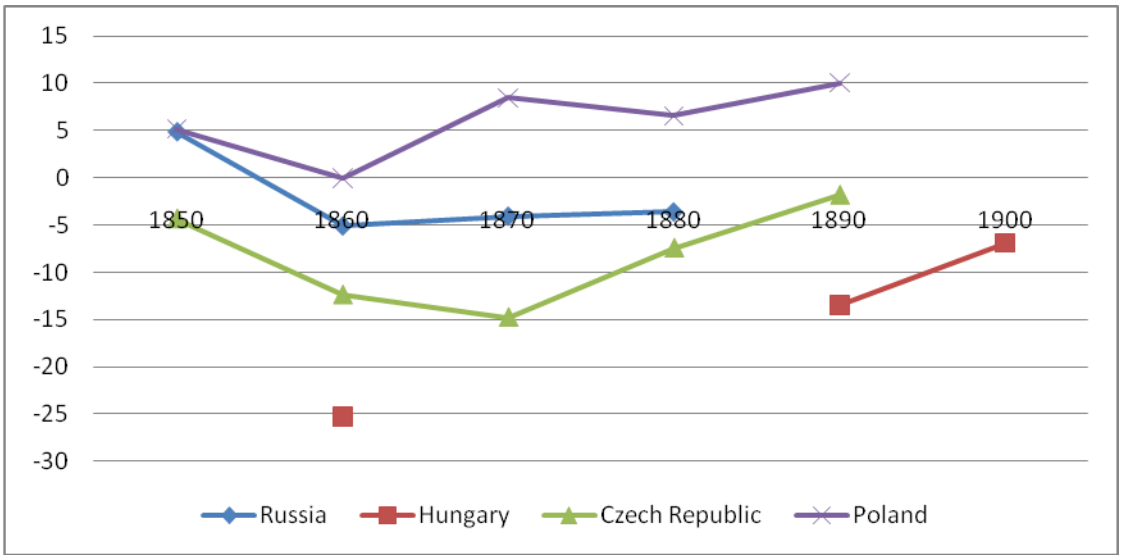


Notes:on the vertical axes it the selectivity of the immigrants. On the horizontal axes the migration decade (Birth decade + 20 years). Selectivity is measured in numeracy differentials, authors own calculations, construction: average numeracy of female immigrants from source country i observed in destination country j in decade t. Census evidence was available for Argentina (1869, 1895); Canada (1871, 1881, 1901); Norway (1865, 1875, 1900); England (1851, 1881); US (1850, 1860, 1870, 1880, 1890, 1900, 1910).

Sources: On the U.S. except 1880: Ruggles, Steven, Matthew Sobek, and Trent Alexander, et al. Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004. On Argentina: Somoza, J. and Lattes, A. (1967): Muestras de los dos primeros censos nacionales de población, 1869 y 1895. Documento de Trabajo No 46, Instituto T. Di Tella, CIS, Buenos Aires. On all other samples: North Atlantic Population Project and Minnesota Population Center. NAPP: Complete Count Microdata. NAPP Version 2.0 [computer files]. Minneapolis, MN: Minnesota.

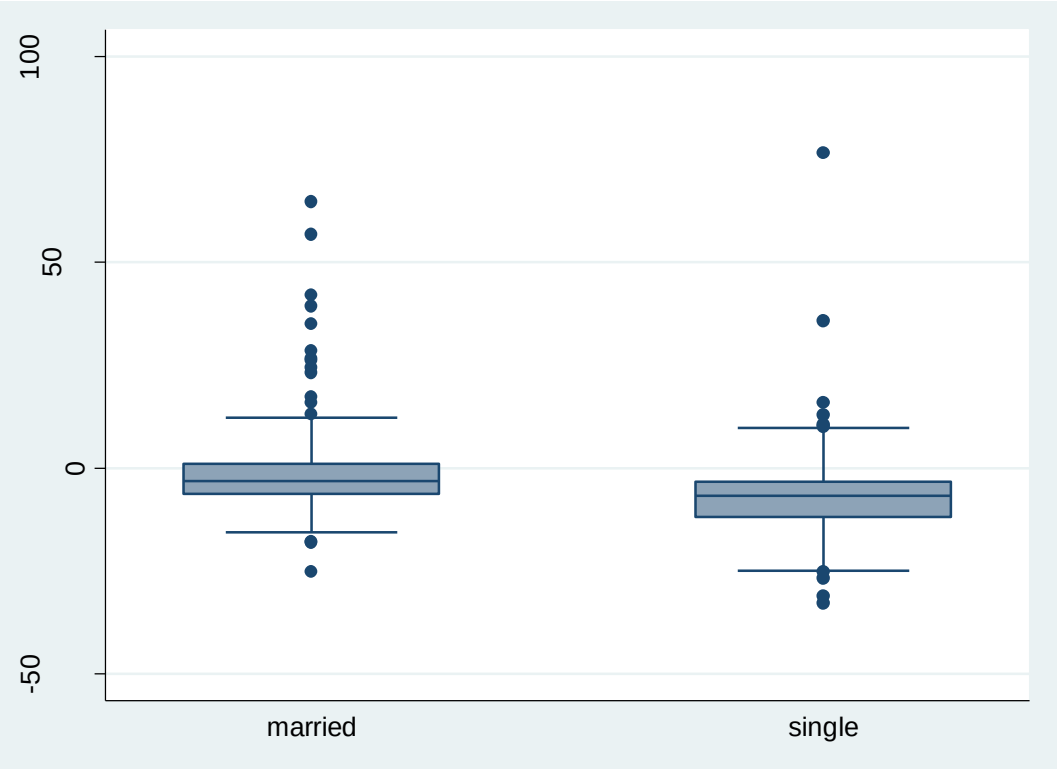
Data for source country numeracy estimates are taken from Crayen and Baten (2009).

Figure 1b: Human capital selectivity of single immigrant women in the US.



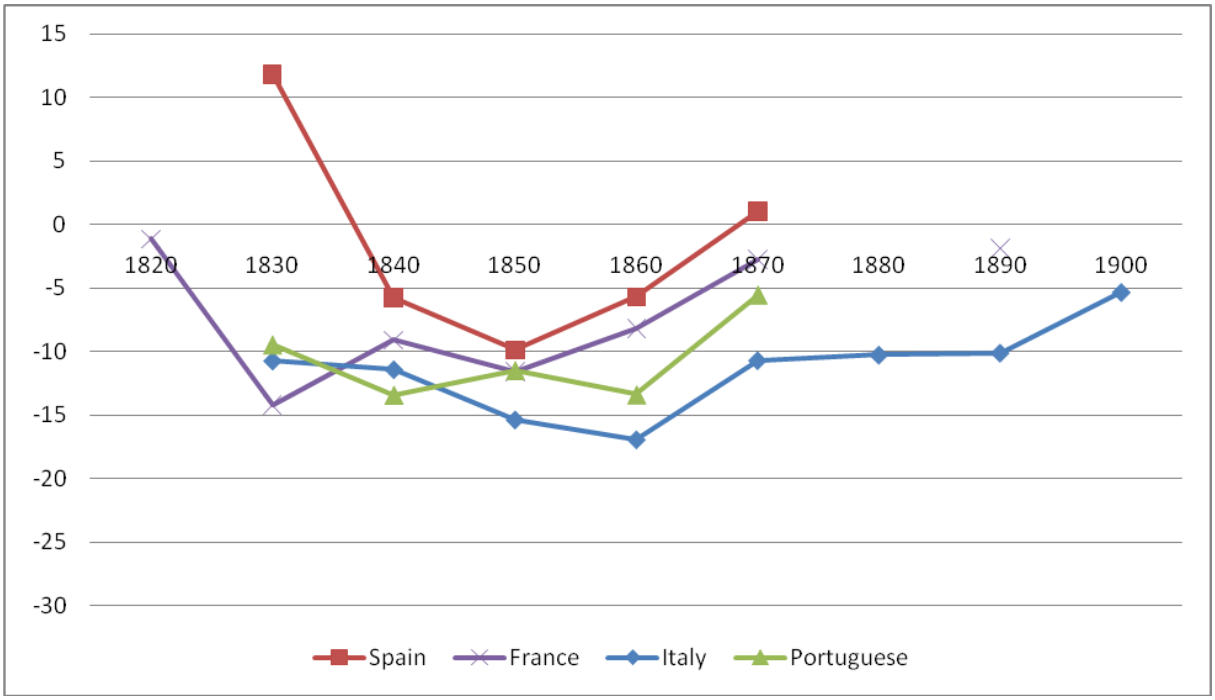
Notes and Sources: see Figure 1a.

Figure 1c: Human capital selectivity of married and single immigrant women in the US.



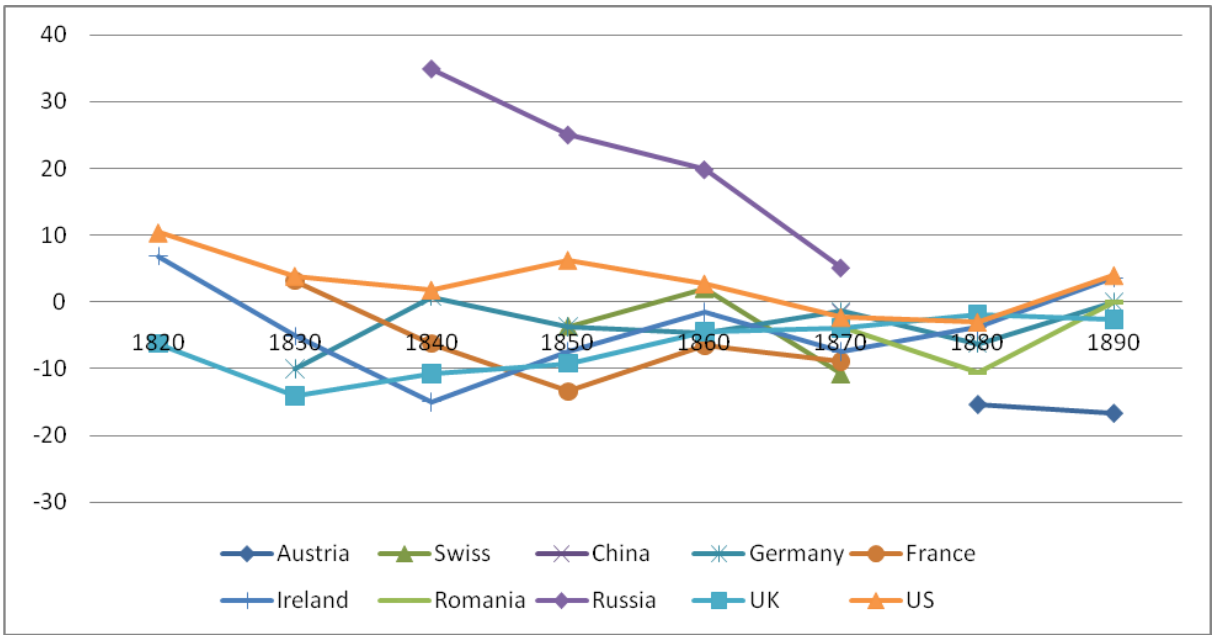
Notes and Sources: see Figure 1a.

Figure 2: Female Human Capital Selectivity of Southern European Migrants to the US



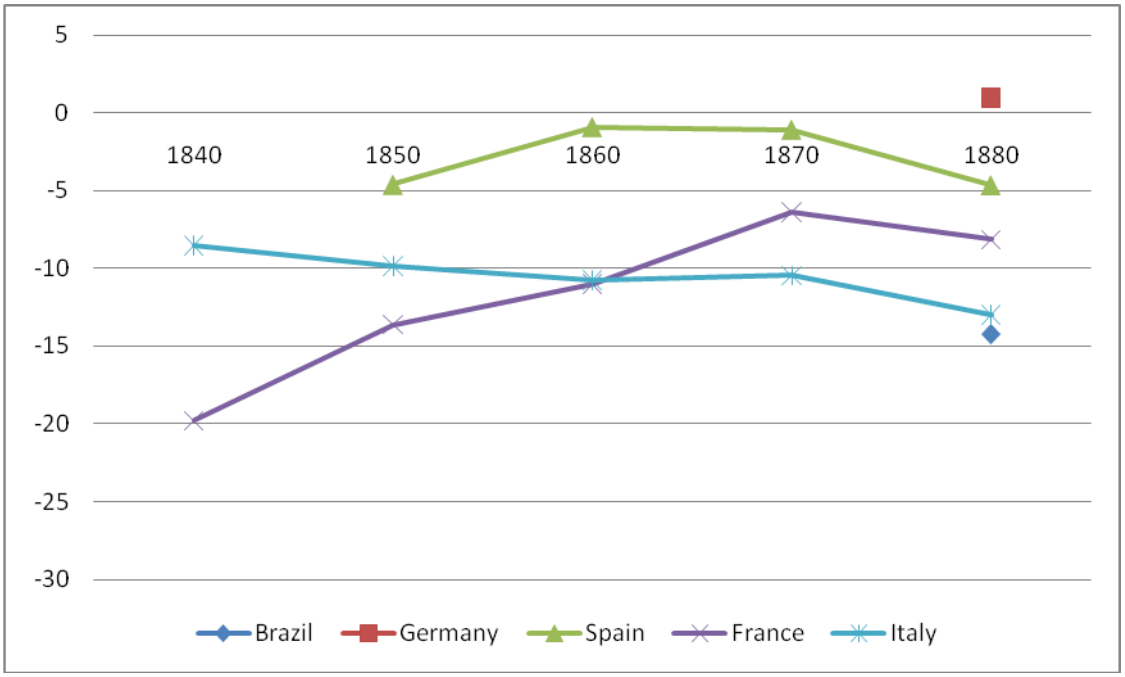
Notes: selectivity measured in numeracy differentials, authors own calculations, construction, see text.
Sources: see Figure 1a.

Figure 3: Female Human Capital Selectivity of Canadian Immigrants



Notes: selectivity measured in numeracy differentials, authors own calculations, construction, see text.

Figure 4: Female Human Capital Selectivity, Argentinean Immigrants



Notes: selectivity measured in numeracy differentials, authors own calculations, construction, see text.

6.10 Tables

Table 1: Descriptive Statistics of the dataset

Variable	Observations	Mean	Std. Dev.	Minimum	Maximum
Selectivity	505	-4.92	8.35	-36.73	35.36
Relative inequality	505	-3.04	0.12	-0.42	0.33
Numeracy level source country	505	93.68	8.39	48	102
log migrant stock	505	1.10	2.04	-4.98	4.61
Poverty constraint	505	0.82	0.34	0	2.10
Log distance	505	3.01	1.04	0.55	4.60
Log distance * poverty constraint	505	2.63	1.52	0	8.27
Relative democracy	505	1.14	3.21	-5.7	10
Common language	505	0.27	0.45	0	1
Colonial ties	505	0.18	0.38	0	1
Marital status	446	0.56	0.50	0	1

Notes: all calculations and estimations were done using the stata 11 programme. In the estimations, marital status was coded as a dummy variable (0=married, 1=single) and the reference category is a married person.

Sources: on the migrant numeracy, see Table 2. Numeracy in the source countries are from Crayen and Baten (2009). Inequality is from Blum and Baten (2011), Estimating Inequality with Anthropometric Indicators, forthcoming for an online version, see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/Baten_Blum_skpr100331a.pdf, last accessed March 31st, 2011. The stock of migrants was calculated with migrant data sets cited in the notes to Table 2. Poverty constraints are based on Maddison (2009), and for those countries for which values were lacking we used the imputations first done by Baten and Blum (2010), see http://www.wiwi.uni-tuebingen.de/cms/fileadmin/Uploads/Schulung/Schulung5/Joerg/baten_blum_ht_100331a.pdf last accessed March 31st, 2011. The distance measure as well as data on colonial ties and common languages is taken from http://www.cepii.fr/anglaisgraph/bdd/distances.htm last accessed March 31st, 2011. The distance was then multiplied with the passenger cost estimates by Sanchez-Alonso (2008) to account for the decline in distance costs. Relative democracy data is from the Polity IV project, see Marshall, Monty G., and Jaggers, K.(2008): Polity IV Project: data set. http://www.systemicpeace.org/polity/polity4.htm last accessed March 31st, 2011. Civil War data is from the Correlates of War Project, see Singer, J. David and Melvin Small (1972): The Wages of War, 1816-1965: A Statistical Handbook. New York. Or see http://www.correlatesofwar.org last accessed March 31st, 2011.

Table 2: Estimating skill selectivity of women: the general model

VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
		Russia excluded	Russia excluded	Russia excluded	Russia excluded	Russia excluded
Relative inequality	0.12**	0.12**	0.10*	0.10*	0.09*	0.12**
	(0.032)	(0.037)	(0.082)	(0.081)	(0.099)	(0.043)
Numeracy level source country	-0.78***	-0.61***	-0.65***	-0.64***	-0.60***	-0.70***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log migrant stock	0.18	0.05	-0.55	0.36	0.01	-0.38
	(0.698)	(0.922)	(0.508)	(0.480)	(0.986)	(0.529)
Poverty constraint	-1.38	-0.86	0.19	2.16	-0.77	4.54
	(0.758)	(0.844)	(0.969)	(0.709)	(0.858)	(0.340)
Log distance	-0.14	-0.24		0.21	0.00	1.16
	(0.938)	(0.893)		(0.904)	(0.998)	(0.536)
Log distance x poverty constraint	0.82	0.56	-0.44	0.01	0.50	-1.51
	(0.602)	(0.718)	(0.787)	(0.997)	(0.745)	(0.346)
Relative democracy	-0.30	-0.12				0.06
	(0.599)	(0.838)				(0.914)
Common language	3.97	5.32*		5.08	5.42*	4.73
	(0.210)	(0.098)		(0.134)	(0.094)	(0.473)
Colonial ties	-5.19	-5.69		-4.74	-5.61	-7.57
	(0.237)	(0.203)		(0.333)	(0.210)	(0.235)
Civil war					-1.69	
					(0.106)	
Poverty constraint x log migrant stock				-0.07		
				(0.240)		
Single Women	-5.21***	-5.03***		-5.04***	-5.00***	-5.27***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Country Pair FE						YES
Time FE	YES	YES	YES	YES	YES	YES
Source FE	YES	YES		YES	YES	
Destination FE	YES	YES		YES	YES	

Constant	51.08***	36.63**	58.13***	43.86**	41.58**	62.90***
	(0.001)	(0.028)	(0.001)	(0.017)	(0.021)	(0.000)
Observations	446	429	500	439	439	446
Number of cdsno	90	86	104	88	88	90
r2_b	0.683	0.657	0.130	0.643	0.652	0.912
r2_w	0.193	0.167	0.179	0.180	0.178	0.201
r2_o	0.511	0.477	0.0680	0.476	0.479	0.616

Notes: Robust in parentheses, *** p<0.01, ** p<0.05, * p<0.1, data sources, see text, all specifications are estimated with robust standard errors. Dependent variable is selectivity measured in ABCC differentials The Hausman specifications test implies that a random effects estimation is efficient and consistent.

Table 3: Selectivity Drivers for Married Women

VARIABLES	**(1)**	**(2)**	**(3)**
		Russia excluded	Russia excluded
Relative inequality	0.17**	0.14**	0.18**
	(0.016)	(0.039)	(0.020)
Numeracy level source country	-1.02***	-0.79***	
	(0.000)	(0.000)	
Log migrant stock	0.18	-0.16	0.11
	(0.754)	(0.849)	(0.848)
Poverty constraint	-2.53	-1.86	-2.37
	(0.594)	(0.453)	(0.693)
Log distance	2.59		1.91
	(0.165)		(0.336)
Log distance x poverty constraint	0.53		0.50
	(0.765)		(0.807)
Relative democracy	-1.65*		-0.26
	(0.085)		(0.792)
Common language	0.38		1.81
	(0.927)		(0.660)
Colonial ties	-3.58		-3.50
	(0.557)		(0.568)
Poverty constraint x log migrant stock			-0.34**
			(0.011)
Country Pair FE		YES, FE	
Source FE	YES		YES
Time FE	YES	YES	YES
Destination FE	YES		YES

Constant	68.26***	76.01***	-18.37
	(0.000)	(0.000)	(0.167)
Observations	250	245	240
Number of cdsno	53	52	51
r2_b	0.759	0.297	0.704
r2_w	0.329	0.261	0.123
r2_o	0.583	0.176	0.436

Notes: Robust in parentheses, *** p<0.01, ** p<0.05, * p<0.1, data sources, see text, all specifications are estimated with robust standard errors. Dependent variable is selectivity measured in ABCC differentials.

Table 4: Selectivity Drivers for Single Women

VARIABLES	**(1)**	**(2)**	**(3)**
		Russia excluded	Russia excluded
Relative inequality	0.04	0.02	0.06
	(0.645)	(0.786)	(0.496)
Numeracy level source country	-0.36	-0.36	
	(0.177)	(0.298)	
Log migrant stock	-0.67	-1.03	-0.65
	(0.287)	(0.492)	(0.340)
Poverty constraint	3.41	2.38	
	(0.560)	(0.387)	
Log distance	-3.02**		-4.03***
	(0.036)		(0.001)
Log distance x poverty constraint	-0.33		0.78
	(0.869)		(0.375)
Relative democracy	-1.19		-0.77
	(0.154)		(0.290)
Common language	8.06**		8.46**
	(0.021)		(0.013)
Colonial ties	-1.61		-1.08
	(0.731)		(0.814)
Poverty constraint x log migrant stock			-0.25
			(0.113)
Country Pair FE		YES, FE	
Time FE	YES	YES	YES
Source FE	YES		YES
Destination FE	YES		YES
Constant	34.01	25.99	11.96

	(0.131)	(0.419)	(0.164)
Observations	196	194	189
Number of cdsno	37	36	35
r2_b	0.908	0.00321	0.969
r2_w	0.274	0.271	0.247
r2_o	0.610	0.0388	0.601

Notes: Robust in parentheses, *** p<0.01, ** p<0.05, * p<0.1, data sources, see text, all specifications are estimated with robust standard errors. Dependent variable is selectivity measured in ABCC differentials.

7. SUMMARY AND DIRECTIONS FOR FURTHER RESEARCH

Human capital may be perceived by some as an ugly word. What it depicts, however, is not ugly at all: as Gary Becker puts it, it is the "knowledge, skills, health, or values"[213] of people. On the micro-level, human capital is important to individuals because "what you know shapes your life".[214] The education of people determines success on the job market, earning possibilities, how they participate in civil society and social life, the way they interact with others and generally the way the process information. On the macro-level, human capital is a determinant of economic performance and productivity, and living standards more generally. In this thesis, the nexus of living standards, human capital and selective migration was at the center of attention. By providing new data on human capital applying the age-heaping method as a proxy, empirical evidence on geographical regions and time periods, where not much was known until now, was generated.

With the data at hand, first, the empirical relationship between biological living standards and human capital in the very long run was established in chapter two. Here, evidence of the long-term development of Portugal from the 18th to the 20th centuries was compiled. A new dataset allowed tracing the biological standard of living from the early 18th century on. It showed that the Portuguese, who today are the shortest of Europe's people, have not been always such an outlier in terms of stature. Among other determinants, both delays in real wage convergence and human capital formation in Portugal in comparison to other European regions was a crucial factor for the country's anthropometric backwardness.

Second, human capital selectivity of migrants was at the focus of the analyses. The impact on destination countries' human capital formation was assessed (Chapters 3 and 4) and next the determinants of human capital selectivity of migrants were investigated (Chapters 5 and 6).

Chapter three compiled long-run data on human capital for seven Latin American countries from the 17th to 20th centuries. The region, which has been growing sustainably during the past ten years or so, was actually on a path of convergence with Western Europe already during the early 18th century. After independence, however, the development stagnated in some Latin American countries. Regional disparities in human capital formation began to intensify. Human capital formation was boosted by European immigration in the major receiving nations Argentina, Uruguay and, to a lesser extent, Brazil.

[213] http://www.econlib.org/library/Enc/HumanCapital.html, last accessed April, 15th, 2011.

Chapter four looked at the case of Brazil more closely. Data on human capital formation since the early 18th century could be provided, because a very early enumeration of São Paulo could be gathered in a Portuguese Archive. Additionally, new data on Brazilian immigrants was retrieved from an Archive in Rio de Janeiro. The evidence allowed estimating the long-run impact of human capital on GDP in the year 2000. Human capital formation was shown to be a highly path dependent process. In a regional panel, increases in numeracy on the state level are significantly positively related to the absolute increase in numbers of immigrants entering the state. This suggests that human capital grew strongest in those states where most immigrants arrived. These states are today still among the top performers in the Brazilian economy.

Chapter five was dedicated to an assessment of the determinants of migrant selectivity. A data set on human capital selectivity comprising immigrants from 52 source countries observed in five destination countries in the Americas and Europe was generated. Relative inequality in source and host countries, measured with an innovative anthropometric approach turned out to be a very robust and important predictor of human capital selectivity of migrants. The results confirmed the Roy model of self-selection (Roy 1951). On average, migrants turned out to be slightly less numerate than the non-migrant populations at home. Thus, a small brain gain was the effect for the source countries.

Chapter six investigated the female side of migration during the first global era. Although well above 40 percent of the migration flows already consisted of women; their experience has received little attention, until now. Therefore, a data set that includes female migrants from 42 source countries to five destination countries was constructed. Women, too, turned out to have been negatively selected on average during the first global era. Furthermore, the relationship status played a prominent role in the migration decision. It could be shown that human capital selectivity has different determinants for married than for single women. For married women, relative inequality of source and host countries is a significant determinant. Single women's selectivity, however, is mainly determined by immediate cost of migration proxies like distance and common language.

What are the major aspects than can be learned from this thesis? There are several: First of all, it was shown that human capital formation is a highly path dependent and persistent process, that impacts on the standard of living. Already in the 19th century, Portugal, for example, had difficulties keeping up the pace of human capital formation that prevailed in other European regions. In a recent article, Teixeira (2005) provided new estimates of

[214] http://www.oecd.org/insights/humancapital, last accessed April, 15th, 2011.

Portuguese human capital for the period 1960-2001. She found that, although considerable progress was made, average years of education of the working population is still considerably below even the compulsory schooling level (Teixeira 2005, p. 112). Therefore, policies which are aimed at augmenting the human capital of the population must take into account the sluggishness of reaction of human capital trends. Therefore they need to be designed sustainably, taking a long-term perspective.

Secondly, while most researchers today are convinced, migrants were positively selected in terms of their skills, a glance at historical data shows that this was not always the case. This suggests that probably, today's positive selection hypothesis might actually be a result of highly selective immigration policies in major receiving nations. One important result of this thesis was that relative inequality turned out to be a very robust and prominent determinant of human capital selectivity of migrants during the era of mass migrations. For countries like Germany which has, for example, a more equal income distribution than Italy or the US (Gottschalk 2000), this implies that highly skilled people would be attracted to labor markets with a more unequal income distribution, causing a brain drain phenomenon. On the other hand, people from countries with higher inequality would be attracted by Germany as a possible migration destination. This could have at least, implications for immigration policies, if not for tax policy also.

One has to keep in mind that the human capital which is captured in this thesis is a very basic form of numeracy skills. Furthermore, even if migrants turn out to be slightly negatively selected in terms of numeracy, they still might be a positively selected sample in terms of risk friendliness, motivation, and perhaps entrepreneurial spirit.

Therefore thirdly, there exists a demand for further studies on human capital selectivity of migrants employing new and innovative human capital proxies. One step in this direction could be the application of quality augmented measures of human capital, as Woessmann (2003) proposed. Yet, when it comes to historical data, this might be hard to accomplish. Another possibility would be the use of health-related measures. Only recently, researchers have begun to point to the impact of health on economic growth (Bloom, Canning and Sevilla 2004, Weil 2007). Massey, Rosenzweig and Smith (2004) investigate health selectivity of international migrants and stress the increasing importance of the health status of immigrants in shaping the well-being of populations. Already Gary S. Becker includes the dimension of health into his concept of human capital, therefore, health related measures could also be employed in studies on migrant selectivity in the future.

7.1 References

Bloom, David E., Canning, David, and Sevilla, Jaypee (2004): "The Effect of Health on Economic Growth: A Production Function Approach" *World Development* 32, no. 1, pp. 1-13.

Gottschalk, Peter (2000): "Empirical evidence on income inequality in industrialized countries" *Handbook of income distribution* 1, Chapter 5, pp. 261-307.

Jasso, Guillermina, Massey, Douglas S., Rosenzweig, Mark R., and Smith, James P (2004): "Immigrant Health: Selectivity and Acculturation", in: Anderson, N.B., Bulatao, R.A. and Cohen, B. (eds) *Critical Perspectives on racial and ethnic differences in late life*, National Research Council, Washington.

Roy, A. (1951): "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3, pp. 135–46.

Teixeira, Aurora A.C. (2005): "Measuring aggregate human capital in Portugal: 1960-2001", Portuguese Journal of Social Science 4, no. 2, pp. 101-120.

Weil, David N. (2007): "Accounting for the Effect of Health on Economic Growth" *The Quarterly Journal of Economics* 122, no. 3, pp. 1265-1306.

Woessmann, Ludger (2003):"Specifying Human Capital", Journal of Economic Surveys Vol. 17, No. 3, pp. 239-270.