Essays on Inequality and Intergenerational Mobility

Christopher Rauh

International Doctorate of Economic Analysis

Departament d'Economia i d'Historia Econòmica

Universitat Autònoma de Barcelona

Supervisor: Prof. Nezih Guner

PhD Thesis in Economics

Cerdanyola del Vallès, May 2014

I would like to dedicate this thesis to my parents, Patricia and Wolfgang,
and my sisters, Alison and Vanessa.

Acknowledgements

First of all, I would like to thank Nezih Guner for his guidance and supervision. Nezih chose to help me in a moment when I needed it most. He helped me to discover a new field and channel my ideas. His passion and excitement for economic research are contagious.
Another person that helped me in a crucial moment, and this even though I had never met him, was Steven Durlauf. His extensive feedback provided the necessary boost to my research. He also was the organizer of my greatest source of inspiration, the Summer School of Socioeconomic Inequality at the Becker Friedman Institute of the University of Chicago.
I would like to specifically thank Joan Esteban, who sacrificed time for me when he had it least. I am also grateful to Susanna Esteban, who besides conducting excellent research and teaching, manages to be the most passionate placement officer around.
There are many others in the faculty of the UAB and the IAE I should be mentioning. Hannes Müller was an excellent discussion partner from whom I learned a lot and also was a supportive figure during my job market experience. I am looking forward to our future collaboration. Caterina Calsamiglia and Hugo Rodrìguez Mendizábal deserve special mention as they were always willing to provide advice and guidance. Also, I would like to thank Angels López and Mercè Vicente, the glue that keeps the department together.
I acknowledge financial support from the Barcelona Graduate School of Economics through the Severo Ochoa Program for Centers of Excellence (SEV-2011-0075).
Besides all the important academic feedback I received from Ezgi Kaya and Arnau Valladares-Esteban, I would also like to thank them for the discussions and gossip that made the daily office routine an enjoyable experience.

Talking about office routines, I have to express my gratitude to Pau Agulló, Mikel Arizaleta, Rodrigo Peláez, Guglielmo Bartolozzi, and the entire Kernel Analytics team with whom I worked and from whom I learned during the two years working at Neo Metrics.
A special thanks goes to my friend and coauthor Rajesh Ramachandran. He did not only manage to explain MWG to me (of course all prices in Rupees) back in our first year of coursework in Bielefeld, but also personally was a great companion during this long journey. QEM was a challenging but fabulous experience, thank you to everybody involved. Too bad Jack Willis decided to stay in Paris, whom I would like to thank for his patience and willingness to carefully read preliminary drafts and provide feedback. The job market experience had not been the same without all the people I met and on who's sofa I crashed on my travels. In particular, I would like to thank Jan Stuhler for repeatedly providing me a roof over my head.
Finally, I would like to thank those that helped me balance work and leisure during the last years. Thank you Ania for smoothing the stresses of a Phd with your care and love. I am grateful to my friends and roommates Carolina, Claudia, Francois, Gary, Gretel, Giorgio, Julia, Marco, Rajesh, Silvia, and Yehenew. Last but not least, I would like to thank Lok Trier and my "Ersatz" family during the last years, my basketball club SAFA Claror, too numerous to mention every single one of them even though they would deserve it. Thank your for having adopted me.

Introduction

In this thesis I study how inequalities emerge and persist. Inequality has risen within the developed world over the past decades. However, I do not only focus on the inequality of outcomes but particularly on the inequality of opportunity. A large part of income inequalities is the consequence of unequal educational outcomes in meritocratic societies. Therefore, it is important to understand at which educational stages the gaps are widened and why access to education might differ for different segments of the population. Many ethnicities and social groups are characterized by persistently low socio-economic indicators compared to the rest of the population. Therefore, it is also essential to ask whether education alone can account for inequalities or whether discrimination might be an important explanatory factor as well.

In the first chapter, I investigate why high earnings inequality goes hand in hand with low intergenerational earnings mobility across developed countries. I study this relationship in a dynastic overlapping generations model, in which a parent can invest in the early education of his child and decides whether to send the child to college. An increase in the college premium translates into increased incentives to invest in early education because of assumed dynamic complementarities between early and tertiary education. It also increases the earnings gap between college and non-college attendants, which results in larger differences in parents' ability to finance education. Public education could mitigate the relationship between inequality and intergenerational mobility. However, public expenditure on education is negatively correlated with inequality. I replicate this cross-country relationship by endogenizing education policies via probabilistic voting, while accounting for biases in voter turnout towards the educated. The model is calibrated to the US as the benchmark economy, which exhibits high inequality and low mobility. Experiments comparing the US to other OECD countries demonstrate that tertiary education characteristics can account for two-thirds of the differences in in-

equality. Patterns of voter turnout across countries explain nearly one-quarter of the differences in inequality and mobility. A counterfactual exercise for the US suggests that compulsory voting could foster intergenerational mobility, whereas the effect on pre-tax inequality is comparably low.

In the second chapter, I study why intergenerational earnings mobility is lower in countries where females are less likely to work and public child-care expenditures are lower. In contrast, when the quality of preschool/child care is high and mothers are more likely work, then intergenerational mobility is high. If less females are working, then more mothers are at home with their children providing a direct channel of transmission of abilities. If both parents actually do work, they will have to either invest in private preschool, which generally is costly and of higher quality, or, if the household cannot afford to do so, arrange for informal child care, which is likely to be less costly and of lower quality. In order to account for cross-country differences in earnings persistence through differences in public preschool expenditures, a dynastic overlapping generations model is calibrated to the US. I find that differences in public preschool expenditures on average explain 17% of the differences in earnings persistence between the US and 15 OECD countries and can explain 22% of the variation.

In the third chapter, joint with Rajesh Ramachandran, we present a model where the collective memory of past discrimination can cause coordination failures leading to present discrimination. The presented type of discrimination can arise for activities characterized by interlinkages, which require the input of more than one individual for the production process to be carried out. The model shows how discrimination can persist forever under perfectly observable ability, when taste for discrimination has died out, individual ability is ex-ante and ex-post identical, and in the absence of discriminatory social norms. Agents without a preference for discrimination rationally discriminate in equilibrium driven by the belief about discriminatory actions by others. The model predicts lower participation rates and higher cost of establishing interlinkages for the discriminated group in equilibrium. Empirically, we analyze differences between blacks and whites in terms of self-employment in the United States, a market characterized by the need for interlinkages. We show, consistent with the theoretical prediction, that the discriminated group has lower participation rates and payoffs in the data. We find beliefs about discrimination to be a significant and economically meaningful factor in explaining lower self-employment rates of blacks in the US.

Contents

List of Figures

List of Tables

Chapter 1

The Political Economy of Early and College Education - Can Voting Bend the Great Gatsby Curve?

1.1 Introduction

Inequality is negatively linked to intergenerational mobility across countries.[1] Recent increases in income inequality across OECD countries highlight the importance of understanding the underlying causal mechanisms of the relationship between inequality and mobility. Why is a country such as the US, which is characterized by an unequal distribution of earnings, at the same time amongst the most rigid societies in the developed world? What causes these unequal outcomes and how are they transmitted to future generations? Education is a strong determinant of earnings and its public provision can be of redistributive nature. According to traditional voting models the provision of public education, in particular non-tertiary education, should be increasing in inequality. However, the share of GDP dedicated to non-tertiary as well as tertiary public education is negatively associated with inequality across countries. This could occur if the rich prefer private education and public policies are biased in their favor. Indeed

[1] The curve resulting from plotting this relationship is commonly referred to as the "The Great Gatsby Curve" and is exhibited in Figure 1.1. The name is adopted from the novel "The Great Gatsby", in which the author F. Scott Fitzgerald challenges the "American Dream" by telling the story of Jay Gatsby, who rises to the high society via shady business deals.

private education expenditures are found to be increasing in inequality across countries. In this paper two potential channels, i.e. tertiary education characteristics and biased voter turnout, are investigated. I quantify how much of the differences in inequality and mobility between the US benchmark economy and other OECD countries, namely Canada, Denmark, Finland, Germany, Norway, Sweden, and the UK, these channels can explain.

The dynamic stochastic general equilibrium model adopts a structure resembling that of Restuccia and Urrutia (2004). Parents invest in early private education of their offspring, which is a substitute to public education, in order to enhance the imperfectly inherited ability. Subsequently, they decide whether or not to send their child to college, where the probability of dropping out of college is decreasing in ability. In each period households vote on two separate proportional tax rates. Thus, early public education and the college subsidy, which is decreasing in parental earnings, are determined endogenously. The model is characterized by the dynamic complementarity between early and late investment (e.g., Cunha, Heckman, Lochner, and Materov 2006, Cunha and Heckman 2008, Caucutt and Lochner 2012), while incorporating the discrete nature of the college investment as in Galor and Zeira (1993).

Tertiary education characteristics in terms of the college premium, tuition, and completion and enrollment rates vary across countries. In particular, the college premium is highly correlated with inequality and intergenerational earnings persistence. By comparing steady states, I find that differences in tertiary education can explain up to 65% of the gaps in the Gini coefficient and 21% of the gaps in intergenerational earnings persistence. High returns to education create an incentive to invest in your child, while richer parents have more funds available to invest than the less well off. The combination of these two effects contributes to the positive cross-country relationship between inequality and intergenerational earnings persistence. High tuition costs magnified by a considerable probability of dropping out make college a costly lottery, relatively more affordable for wealthier households. If the probability of dropping out is decreasing in skills and dynamic complementarities exist between early and college education, then parents have greater incentives to invest in their children's education. Given that parents earn less when they are younger and cannot borrow against the future of their children, large incentives for private investment combined with imperfect credit markets increase the gap between the education received by rich and poor children at the early stage, and

consequently at the college level as well.

Rich parents are likely to prefer private early education, given that a large share of public education expenditures would have to be financed out of their pockets. Relatively high voter turnout among the educated, as in the US, might bias policies in their favor. In contrast, relatively high voter turnout among the less educated could increase public expenditures on early education due to its redistributive nature. In the model economy public education expenditures are endogenous and households vote via probabilistic voting. This allows me to exploit the skewness of voter turnout by age and level of education across countries to explain variations in education expenditures and the effects on inequality and mobility. The weights of individuals in the voting process are assigned according to voter turnout by age group and level of education using the voting supplement of the Current Population Survey (CPS) of 2006 for the US, and the European Social Survey 2010 (ESS) and the Canadian Election Study of 2008 for the experiments. I find that observed patterns of public and private education expenditures, inequality, and intergenerational mobility can be reconciled by voter turnout. On average 23% of differences in intergenerational mobility and 21% of differences in the Gini index compared to the US can be explained by voter turnout. As a robustness check, I repeat the analysis while weighing voters by the fraction of party members per age group and education level. The data is obtained from the World Values Survey 1981-2007 (WVS) and the results exhibit similar patterns. This indicates that the political participation of a society, whether through voting or through party membership, shapes public policy, and thereby influences inequality and intergenerational mobility.

Given that the patterns of voter turnout perform well at explaining cross-country differences, two possible voting policies in the US are considered. First, I enforce compulsory voting. Chong and Olivera (2008) provide empirical evidence that this has an equalizing effect on the income distribution across countries, while Fowler (2013) shows that the introduction of compulsory voting in Australia increased public pension expenditures. Second, I allow parents to vote for their children. This policy has recently received attention in the public debate and was first discussed by Demeny (1986) in order to counter decreasing fertility rates. I find that intergenerational earnings persistence is reduced by 10% under compulsory voting, whereas the extension of electoral franchise to children as of birth by letting their parents vote for them nearly halves earnings persistence. However, the equalizing effect of these policies on pre-tax earnings is

comparably low.

By including the political economy of education expenditures, the model demonstrates how policies and outcomes across countries might react differently to an exogenous change in the college premium. Simulations with an exogenous increase in the college premium in the US indicate that earnings persistence could decrease substantially. More children from the bottom of the earnings distribution enroll in college, hence increasing public support for financial aid. Increased enrollment raises aggregate returns to early education, which raises the share of earnings taxed to finance early education, further augmenting intergenerational mobility.

The paper can be summarized in terms of the five principle findings, of which two explain cross-country differences and three deal specifically with the US. First, the characteristics of tertiary education in terms of the college premium, tuition costs, and completion probability, can explain a large share of differences to the US in inequality and to a lesser extent intergenerational mobility. Second, the skewness of voter turnout in terms of age and education can account for a considerable part of the gaps in intergenerational earnings persistence. Third, a further rise in the college premium in the US could increase intergenerational mobility despite pronounced earnings inequality, which is in contrast to the outcome when the political economy is ignored and households are not allowed to vote. Fourth, policies of compulsory voting or extending the electoral franchise to children as of birth could foster intergenerational mobility. Fifth, in line with recent research (e.g., Belley and Lochner 2007, Bohacek and Capicka 2012, Winter 2013) nearly one in four individuals is found to be financially constrained in the college decision.

Becker and Tomes (1979, 1986) and Loury (1981) pioneered the microfoundation of the three fundamental institutions in the determination of a child's future: the family, the labor market, and the state. Their approach is extended and formalized by Solon (2004). Galor and Zeira (1993) present on the macroeconomic level how credit market imperfections and indivisibilities in human capital investments can transmit distributions of wealth. Glomm and Ravikumar (1992), Durlauf (1996), Bénabou (1996), and Fernandez and Rogerson (1998) demonstrate how locally provided public education affects growth, intergenerational income transmission, and inequality when households are sorted geographically by income. Abstracting from different stages of education and heterogenous innate ability among individuals, Fernandez and Rogerson (1995) show

that richer individuals can be those capturing higher education subsidies, thereby making the transmission of unequal distributions more likely.

The paper is most closely related to Restuccia and Urrutia (2004), Herrington (2014), and Holter (2014), who use dynastic overlapping generations models to quantitatively study intergenerational mobility. Assuming exogenous taxation and education policies, Restuccia and Urrutia (2004) find that to reduce intergenerational persistence an increase in early education expenditures would be more efficient than increasing college subsidies. Herrington (2014) accounts for a share of the gap in inequality and intergenerational persistence between Norway and the US through variations in the distribution of public spending and revenue sources. Holter (2014) exploits cross-country differences in the progressivity of income taxation to explain part of the gap in intergenerational mobility between the US and Denmark. However, in his partial equilibrium setting, as well as in Restuccia and Urrutia (2004) and Herrington (2014), the shape and level of the taxation function are taken to be exogenously given. This inevitably leads to the question of the political economy of education expenditures at the different stages of human capital formation. Ichino, Karabarbounis, and Moretti (2010) use a microeconomic approach to the political economy of the underlying structure of intergenerational income mobility considering only a single stage of education. Chetty, Hendren, Kline, and Saez (2013) corroborate the importance of the effect of tax expenditures on mobility. Using administrative tax records, they find that schools with higher expenditures per student have higher rates of upward mobility and that tax policies remain correlated with mobility even after controlling for other important factors, such as local social capital and family structure.

There is a growing political science literature investigating individual preferences for expenditures on public education. Most empirical studies are limited to general preferences on education expenditures expressed in surveys, not allowing for a distinction between early and college education. One finding is that these preferences are not only driven by income, but are increasing in the level of education (Busemeyer 2012, Horn 2012). This occurs in my model as well, however the causality is not driven by the education of an individual. Education serves as a proxy for ability, which is positively correlated with the ability of the child. Now since returns to education are assumed to be increasing in ability, holding earnings fixed, a more educated individual on average has a greater demand for public education. Busemeyer (2012) also finds that those still

enjoying the benefits of education, as well as people with children are more likely to support public education. These findings align well with the voting mechanism specified in the model, where individuals' preferences on public education are driven by their own payoff and altruism for their children, and not by altruistic motives for the society as a whole. Ansell (2010) models preferences for tertiary education expenditures in the setting of a class conflict where coalitions are formed. When enrollment rates are low, access is restricted to the elite, and consequently the poor and the middle class do not support public expenditures. With increasing enrollment more individuals from the lower end of the income distribution enter tertiary education, thereby expanding support and consequently public subsidization. Similar dynamics emerge from the probabilistic voting process embedded in my model.[2] Wolf (2009) does not focus on individual preferences, but explains the cross-country ratios of public to private expenditures through union density, federalism, the share of protestant population, and the given composition of the cabinet, thereby arguing in terms of value-related and ideology-based factors.

The remainder of the paper is organized as follows: In Section 2.2 the stylized facts of earnings inequality, intergenerational earnings persistence, and education expenditures are presented. The model is explained in Section 3.2 with the equilibrium definition following in Section 2.4. The parameterization is described in Section 1.5, whereas the benchmark results and policy experiments for the US are presented in Section 2.5.3. In Section 2.6 experiments explain differences between the US benchmark economy and other OECD countries. The robustness of these results is analyzed in Section 1.8. Finally, Section 1.9 concludes and outlines questions for future research.

1.2 Inequality and intergenerational earnings persistence

Intergenerational mobility is generally measured by regressing the log of son's earnings on the log of father's earnings to quantify earnings elasticities across generations, which captures the percentage change in a son's adult earning that is associated with a one percentage point increase in paternal earnings. Reliable estimates are difficult to obtain due to the requirement of data on earnings of both generations and are complicated by

[2] Ansell (2010) goes further by showing how leftist parties support for expenditures on higher education shift with the enrollment rates. Since in my model there are no left or right parties the details of partisan behavior escape my analysis of the political economy.

lifecycle and macroeconomic fluctuations. The empirical estimation was pioneered by Solon (1992) and Zimmerman (1992) finding an earnings persistence of 0.4 in the US.[3]

Wage inequality, as captured by the Gini coefficient of household earnings before

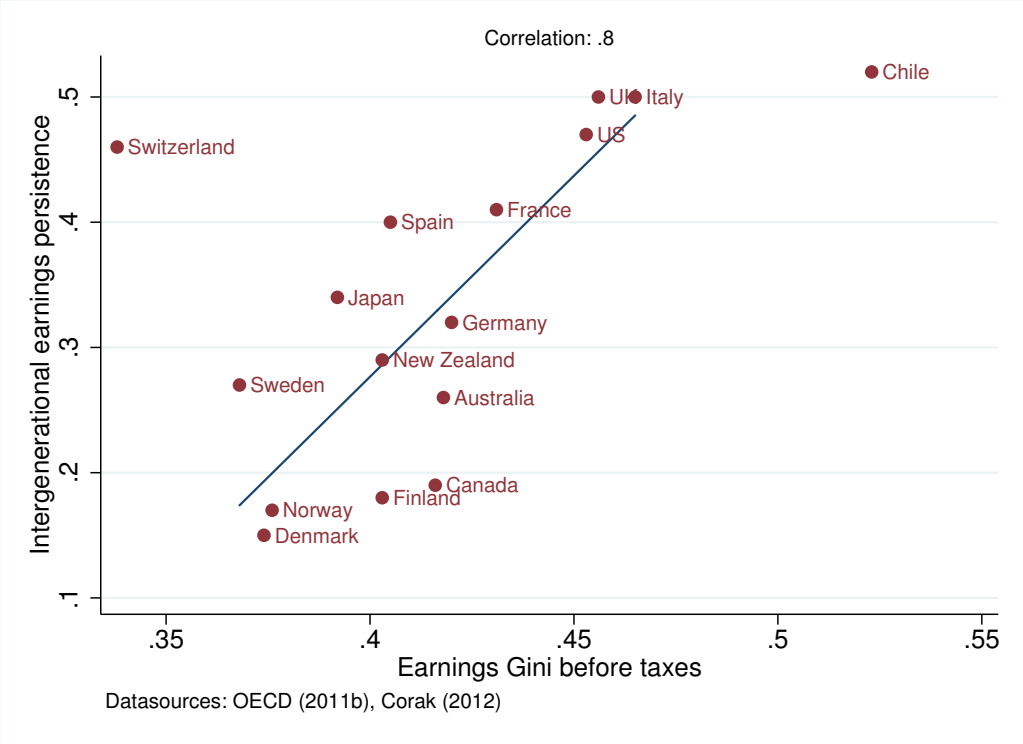


Figure 1.1: Gini before taxes versus intergenerational earnings persistence

taxes and transfers of the population aged 18-65 by the OECD (2011b), is highly correlated with the intergenerational persistence of earnings summarized by Corak (2012), as

[3]Concerning estimates of intergenerational mobility see Blanden, Gregg, and Macmillan (2007) for the UK, Björklund and Jäntti (1997) for a comparison of the US and Sweden, and Solon (2002) or Corak (2012) for a survey of cross-country estimates. Black and Devereux (2011) provide an excellent overview of recent findings. A range of empirical papers has investigated the relationship between inter- and intragenerational inequality. Cooper (1998) finds empirically that in poor neighborhoods redistributive expenditures on human capital have a significant effect in reducing the persistence of economic status across generations, thereby hinting towards a causal relationship. Andrews and Leigh (2009) similarly find that sons that grew up in more unequal states during the 1970s experienced less social mobility by the late 1990s. Aaronson and Mazumder (2008) present how the "snapshot" and the "moving-picture" measure of inequality move parallel between 1940 and 2000 in the US. Duncan, Khalil, and Ziol-Guest (2013), controlling for family structure and characteristics, find that between two-thirds and three-quarters of the increase in schooling gaps between low and high income children of cohorts born in the 50's and 80's in the US can be accounted for by increases in income inequality. In contrast, Bloome (2013) finds little evidence for an effect of inequality during childhood on intergenerational mobility across US states.

exhibited in Figure 1.1.[4] The positive slope of the "Great Gatsby Curve" indicates that a higher level of inequality is positively associated with a greater transmission of economic status across time in the past. Even if inequality can be justified as an outcome of differential efforts, this strong positive relationship raises the question of whether inequality also negatively affects the equality of opportunities.

In the top panels of Figure 1.2 one can see the high correlation between the Gini

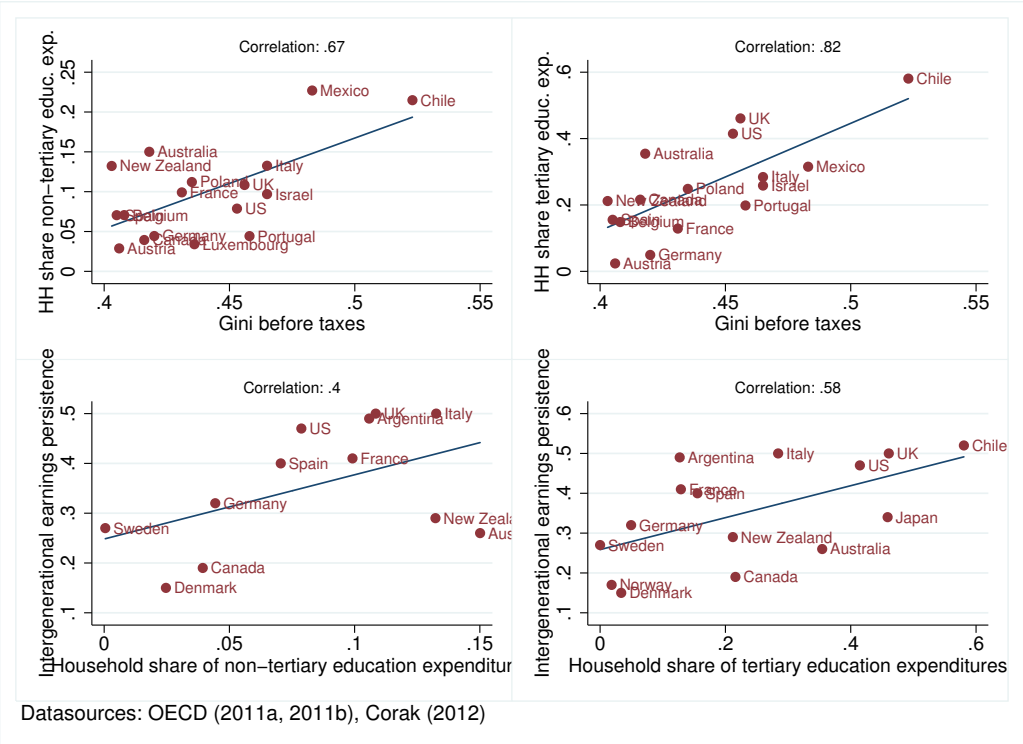


Figure 1.2: Non-tertiary (left) and tertiary (right) education expenditures financed by households versus pre-tax Gini (top) and intergenerational earnings persistence (bottom)

coefficient of earnings before taxes and the share of total education expenditures financed privately by households of non-tertiary and tertiary education, respectively.[5] As

[4]The fitted line and the correlation coefficient are computed excluding the outliers Chile and Switzerland.

[5]I use the expression "non-tertiary" or "early" to summarize primary and secondary education. Given the evidence of the early formation of cognitive and non-cognitive skills (e.g., Cunha et al. 2006, Heckman 2010) the consideration of pre-primary education is surely an important factor in the examined dynamics and outcomes from which I abstract. Peer and neighborhood effects, as well as health, are other channels worth mentioning, through which parental income could affect learning aptitude of children, but

exhibited in the bottom panels of Figure 1.2, the shares of total education expenditures financed privately by households are highly positively correlated with the intergenerational transmission of earnings.

Assuming public early education to be of redistributive nature, according to traditional voting models one would expect public education expenditures to be increasing in inequality. As can be seen in Figure 1.3 the opposite is the case in the data, whereas private education expenditures are indeed found to be positively related to inequality. Under the assumption that the rich prefer private education, this could be driven by a bias in voter turnout, as exhibited for a small sample of countries in Figure 1.10 in the Appendix. The share of GDP dedicated to public education is decreasing in the relative turnout of college graduates compared to those with no college education.

On the one hand, the provision of public education affects how much education can be guaranteed to each individual. On the other hand, high returns to education increase the incentives parents have to invest in education. Therefore, assuming credit market imperfections high returns to education can drive a wedge between the education that poor and rich children achieve. An increase in the college premium raises the returns to investment directly at the college level, and assuming dynamic complementarities between early and college education also raises the returns to investment in early education. As can be seen in Figure 1.4, the college premium, here defined as the relative earnings of college graduates compared to upper secondary education of the population aged 25-64 (OECD 2013), is positively associated with the Gini as well as with intergenerational earnings persistence. Therefore, now we have two possible explanations for cross-country differences in inequality and intergenerational mobility consistent with the data patterns. First, a technological explanation in terms of the college premium.[6] Second, a political explanation in terms of the bias in voter turnout. In the following, a model will be specified, which allows for the quantification of the contribution of these two channels to cross-country differences in inequality and mobility.

are not included in the model. Also parental connections, though not an influence on cognitive abilities, could possibly be of high importance to the intergenerational linkage of earnings. Corak and Piraino (2011) find that about 40% of young Canadian men have worked for an employer that employed their father at some point in time, while 6-9% have the same employer in adulthood. The percentage is found to increase with paternal earnings, especially at the top.

[6]The college premium has mainly been linked to technological developments. See references in Section 1.6.2.1.

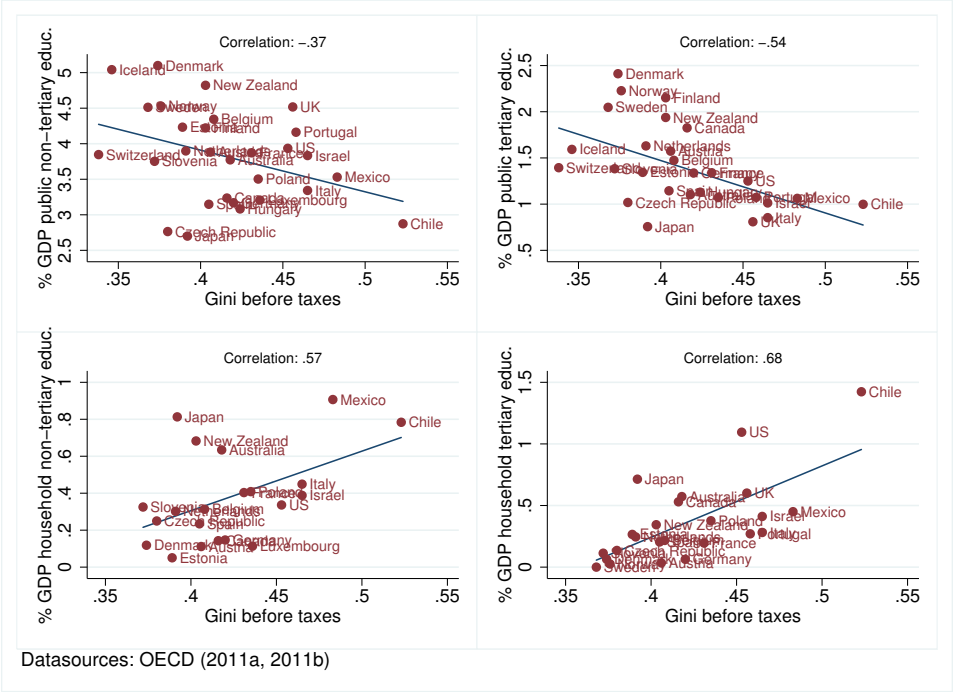


Figure 1.3: Gini before taxes versus public (top) and household (bottom) expenditure (% of GDP) dedicated to non-tertiary (left) and tertiary (right) education.

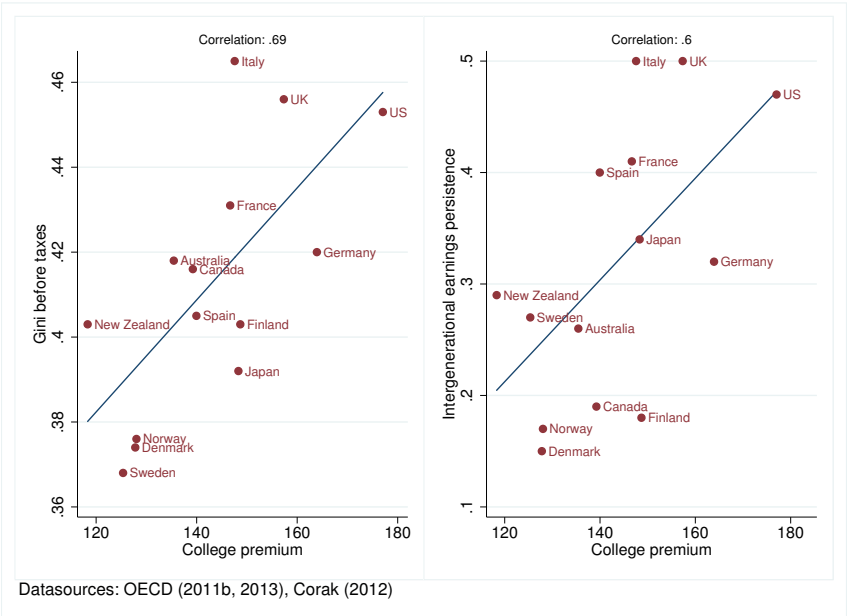


Figure 1.4: College premium versus Gini (left) and intergenerational earnings persistence (right)

1.3 The model

The overlapping generations model resembles the one presented by Restuccia and Urrutia (2004). A household exists for two periods and is composed of either a young parent and a young child, or an old parent and an old child, which I will refer to as young and old households, respectively. In a subsequent period old parents die, whereas the old children become young parents and form households of their own. Consequently, everybody lives for four periods, but only makes decisions during the last two periods of the lifecycle. Population growth is zero and all parents have the same number of children.[7] Parents take education decisions for their children. Households are heterogenous in their levels of innate ability of the offspring and human capital of the parents.

Lifetime utility is composed of consumption as a young household c_y, consumption as an old household c_o, and "warm glow" for the level of human capital of their child h_c. The assumption that parents derive utility from ensuring that their children are equipped with an adequate level of human capital is common in the literature (Glomm and Ravikumar 1992, Galor and Zeira 1993, Bénabou 2000). The utility function for consumption $u(c)$ is increasing and concave, while altruistic utility gained from human capital of the child $v(h_c)$ is nondecreasing. Labor is provided inelastically, where human capital is remunerated by competitive firms at a wage rate w.

1.3.1 Ability and human capital

Innate ability is assumed to be correlated with parental innate ability, which is not altered over the lifecycle and can be interpreted as the genetic component. Acquired ability $\hat{a}$ is a function of innate ability and public and private education investments when young, which later will affect the probability of college completion and wages. When the child is old, acquired ability $\hat{a}$ transforms into human capital h_c. Once he becomes a parent, human capital evolves exogenously, while capturing the lifecycle earnings profile. When becoming a young parent human capital experiences a shock and converts to h_y, whereas for the old parent human capital is denoted by h_o.

[7] While De la Croix and Doepke (2004) demonstrate theoretically how differential fertility rates might play a considerable role in the distribution of human capital investments and preferences for public education among families, Björklund et al. (2004) find no differences in intergenerational mobility by family size in Norwegian data.

Innate ability a, when passed from one generation to the next, follows a first-order discrete Markov process with mean normalized to one, and transition matrix Ψ, while restricting the vector of states for a and the elements in the transition matrix such that the process is a discrete approximation of a continuous AR(1) process:

$$ln(a') = \rho ln(a) + \varepsilon \; where \; \varepsilon \sim N(0, \sigma_a^2). \tag{1.1}$$

Innate ability a is transformed into acquired ability $\hat{a}$ via public education g and private early education e according to the function $f(a,e,g)$ when the offspring is young. The function $f(a,e,g)$ is assumed to be positive, strictly increasing, and concave in all its arguments. Public and private education are assumed to be perfect substitutes. Following the literature on early skill formation, acquired ability $\hat{a}$ is increasing in innate ability a. Also the return to investment in education is increasing in innate ability, such that $\frac{\partial^2 f}{\partial a \partial e}$, $\frac{\partial^2 f}{\partial a \partial g} > 0$, as skill has been found to beget skill (Cunha et al. 2006, Cunha and Heckman 2008).

The functional form chosen is close to the Ben-Porath (1967) specification and assumed to be given by

$$\hat{a} = f(a,e,g) = \chi a^{1-\gamma}(g+e)^{\gamma}. \tag{1.2}$$

The efficiency of private and public investment increases in parameter γ, which is also responsible for the relative importance of innate ability versus investment. Parameter χ is the efficiency parameter regulating the level effect of human capital creation.

Everybody enjoys public education, while choosing the level of private education to supplement it. This sort of structure is not limited to the growing number of charter schools, as investments in early private education can be imagined as anything from piano lessons to out of school tutoring. Data for the US reveals that the gap in "enrichment expenditures" between poor and rich parents has become greater over time (Duncan and Murnane 2011, Kornrich and Furstenberg 2013). Also, in the United States a large share of non-tertiary education is financed locally through property taxes or other local sources. Households are assumed to be able to choose the level of investment by moving into a neighborhood which provides the preferred level of investment and is priced accordingly due to housing prices differentials (Epple and Romano 1996, Fernandez and Rogerson 1997). Hoxby (1998) and Herrington (2014) show a positive relation between per-pupil spending and income.

Old parents decide on whether or not to send their offspring to college $s \in \{0,1\}$. If they do not go to college the offspring works for the entire period, whereas if they do go they spend share $\underline{n}$ of the period in college if they drop out, and share $\bar{n}$ in case they graduate, while working the rest of the period. The earnings of the offspring are shared with the parents at the household level. If the child goes to college, he completes college with probability $\pi(\hat{a})$, consequently dropping out with probability $(1-\pi(\hat{a}))$. The probability π of completing college is increasing in acquired ability (Light and Strayer 2000, Chatterjee and Ionescu 2010) and is assumed to take the functional form

$$\pi(\hat{a}) = min\{1, \psi_0(1+\hat{a})^{\psi_1}\}, \tag{1.3}$$

where $\psi_0 > 0$ and $\psi_1 > 0$. The parameter ψ_0 is responsible for the level effect of acquired ability on the probability of college completion, while the convexity of the function is increasing in ψ_1. College education comes at a tuition cost T per period. The government subsidizes a share decreasing in parental earnings. The share q of tuition cost T covered by the government is linearly decreasing in parental earnings and is given by

$$q(h_o) = \max\{0, 1-\kappa w h_o\}. \tag{1.4}$$

The share covered by the college subsidy is bound from below by 0, such that no household has to pay an additional fee, and from above by 1, assuring that no household receives a subsidy beyond compensation of the actual cost of attending college. If the offspring completes college, its acquired ability is multiplied by $\bar{\theta}$, while if it drops out only by $\underline{\theta}$ (where $\bar{\theta} > \underline{\theta}$). Therefore, acquired ability is mapped into human capital of the old child h_c via the functional form

$$h_c = \begin{cases} \hat{a} & \text{if does not attend college} \\ \bar{\theta}\hat{a} & \text{if completes college} \\ \underline{\theta}\hat{a} & \text{if drops out of college,} \end{cases} \tag{1.5}$$

where the college completion probability is given by (2.5). This functional form satisfies the dynamic complementarity between early and late investments in human capital discussed by Cunha et al. (2006) and Cunha and Heckman (2008). Despite recent discussions about the importance of the increased uptake of student loans (Belley and

Lochner 2007, Lochner and Monge-Naranjo 2012, Abbott, Gallipoli, Meghir, and Violante 2013), I do not allow households to borrow and abstract from physical capital accumulation.[8]

In the transition from old offspring to young parents two things occur. Firstly, individuals can experience a shock ζ_y commonly referred to as market luck. The shock accounts for the fact that earnings dispersion within a cohort increases over the lifecycle (Huggett, Ventura, and Yaron 2006) and for the finding by Huggett, Ventura, and Yaron (2011) that only 61% of the variance in lifetime earnings can be attributed to preworking conditions. The shock is multiplicative and takes either value $\zeta_y \in \{-\zeta, 0, \zeta\}$ with equal probability $1/3$. This shock creates no aggregate uncertainty and is permanent. Secondly, their human capital increases exogenously by the lifecycle component $\eta_0 > 1$, such that

$$h_y = \eta_0(1+\zeta_y)h_c. \tag{1.6}$$

Young parents are equipped with human capital h_y and choose how much to invest in private early education e. Old parents have human capital h_o, which is given by

$$h_o = \eta_1 h_y, \tag{1.7}$$

where $\eta_1 > 1$ represents the increase in the earnings profile through experience later in the lifecycle. Earnings are determined by human capital and the competitive wage w, the rate of return to human capital in the market.

1.3.2 Households

Households take tax rates for financing of early (τ_p) and college (τ_q) education, as well as public expenditures on early education (g) and the college education subsidy (q) in t and $t+1$ as given when maximizing discounted lifetime utility. The discount factor is $\beta < 1$. The utility of consumption is assumed to take the functional form

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma} \tag{1.8}$$

[8] According to a Sallie Mae-Ipsos Report in 2013 student loans pay for only 18% of costs of college attendants.

while the utility derived from leaving the child with human capital h_c is

$$v(h_c) = \phi h_c^{\xi}, \tag{1.9}$$

where $\phi > 0$ and $\xi > 0$.

The state variables of the old household are $(h_o, \hat{a})$. Let λ be equal to one if the child completes college and zero if he drops out. Given the wage rate w then the problem of the old parent household can be written as

$$V_o(h_o, \hat{a}) = \max_{s \in \{0,1\}} \{V_o^{s0}(h_o, \hat{a}), E_\lambda[V_o^{s1}(h_o, \hat{a}, \lambda)]\}, \tag{1.10}$$

where $V_o^{s0}(h_o, \hat{a})$ is the value function of not sending the child to college, whereas $V_o^{s1}(h_o, \hat{a}, \lambda)$ is the value of sending the child to college, and hence depends on the college completion probability given by (2.5). The value of not sending a child to college is given by

$$V_o^{s0}(h_o, \hat{a}) = \max_{c_o \geq 0} \{\frac{c_o^{1-\sigma}}{1-\sigma} + \phi h_c^{\xi}\} \tag{1.11}$$

subject to

$$c_o = (1 - \tau_p - \tau_q)(wh_o + wh_c) \tag{1.12}$$

and

$$h_c = \hat{a}. \tag{1.13}$$

The expected value of sending a child to college conditional on college completion λ is given by

$$E_\lambda V_o^{s1}(h_o, \hat{a}, \lambda) = \pi(\hat{a}) V_o^{s1}(h_o, \hat{a}, 1) + (1 - \pi(\hat{a})) V_o^{s1}(h_o, \hat{a}, 0), \tag{1.14}$$

where

$$V_o^{s1}(h_o, \hat{a}, \lambda) = \max_{c_o \geq 0} \{\frac{c_o^{1-\sigma}}{1-\sigma} + \phi h_c^{\xi}\} \tag{1.15}$$

subject to

$$c_o + (1 - q(h_o))(\lambda \bar{n} + (1-\lambda)\underline{n})T = (1 - \tau_p - \tau_q)(wh_o + (1 - (\lambda \bar{n} + (1-\lambda)\underline{n}))wh_c) \tag{1.16}$$

$$h_c = (\lambda \bar{\theta} + (1-\lambda)\underline{\theta})\hat{a}, \tag{1.17}$$

where the probability $\pi(\hat{a})$ of completing college is given by (2.5) and the college subsidy $q(h_o)$ by (1.4). Let b be the level of education obtained by a parent. This can either be high school, college dropout, or college graduate. Let the distribution of old parent households be summarized by $\mu(x_o)$, where $x_o = (h_o, \hat{a}, a, b)$, and let $s(x_o)$ and $c_o(x_o)$ be the solution to (1.10), and therefore the policy functions of the college decision and consumption of the old household, respectively.

The state variables of the young household are (h_y, a). The problem of the young household is given by

$$V_y(h_y, a) = \max_{e, c_y \geq 0} \{ \frac{c_y^{1-\sigma}}{1-\sigma} + \beta V_o(h_o', \hat{a}') \} \tag{1.18}$$

subject to

$$c_y + e = (1 - \tau_p - \tau_q) w h_y \tag{1.19}$$

$$\hat{a}' = \chi a^{1-\gamma}(g + e)^{\gamma} \tag{1.20}$$

$$h_o' = \eta_1 h_y. \tag{1.21}$$

Let $x_y = (h_y, a, b)$ and let $c_y(x_y)$ be the policy functions of consumption associated with the young, and $e(x_y)$ the policy function determining early private education. The distribution of young parent households is described by $\mu_y(x_y)$. The total mass of households is assumed to be constant and of mass unity each, young parents and old parents households. Therefore, the total mass of individuals is four (since each household contains one parent and one child).

1.3.3 Firms

A representative firm produces a single good and has a linear constant return to scale production technology producing aggregate good Y using aggregate human capital H as an input

$$Y = F(H) = AH. \tag{1.22}$$

Therefore, the firms problem can be written as

$$\max_{H}\{F(H) - wH\}. \tag{1.23}$$

1.3.4 Government

The government levies two proportional tax rates, while running a balanced budget for early and late education separately. First, tax τ_p which is used to finance public early education g. Second, tax τ_q which covers part of tuition cost T for those going to college. Therefore, government expenditures g on early education are given by

$$g = \tau_p Y. \tag{1.24}$$

The college subsidy depends not only on tax τ_q and total production in the economy, but also on the composition of college students. More specifically, it depends on how many students attend college, how long they attend, and what the earnings of their parents are. The government budget for tertiary education can be summarized by

$$T \int s(x_o)q(h_o)n(\hat{a})\, d\mu_o(x_o) = \tau_q Y, \tag{1.25}$$

where $q(h_o)$ is given by (1.4), $n(\hat{a}) = \pi(\hat{a})\bar{n} + (1 - \pi(\hat{a}))\underline{n}$ is the time spent in college, and $\pi(\hat{a})$ is given by (2.5).

An increase in tax rate τ_q, ceteris paribus, leads to a reduction in κ, which for a household sending its offspring to college increases subsidy rate q, given their earnings are not too high and satisfy $wh_o < \frac{1}{\kappa}$.

1.3.5 Political economy

The two proportional tax rates are determined jointly through probabilistic voting, which allows for a weighted average of preferences across households instead of only relying on the median voter.[9] Due to the two-dimensional problem preferences are unlikely to be single-peaked, not allowing for the identification of the median voter.

In probabilistic voting parties commit to policies before elections take place. The policy platform is chosen by opportunistic candidates, which only care about being elected. It is assumed that parties differ along an ideological dimension observable to the voter. Candidates know the ideological preference distribution of the voters, wherefore chosen policies are directed towards those voters that are less driven by the ideological component. Candidates have an average popularity common to all voters, which is a random variable and could be subject to a shock, such as a scandal the day before elections take place. Since the policy platform is chosen when the outcome is uncertain, parties maximize the expected share of votes, and thereby the probability of winning the election. There exists a unique political equilibrium in which both parties propose the same policy by maximizing a weighted social welfare function, where weights are determined by how responsive voters are to policies, which might vary due to the ideological component (see Lindbeck and Weibull 1987, Persson and Tabellini 2000). I use voter turnout by age and level of education as weights in the voting process, as candidates might be best of catering to the segment of population actually voting.

The decision of the young parent household does not only depend on tax rates in t, but also on tax rates in $t+1$ (denoted by a prime), which will be decided upon in the following period. Therefore, the current policy choices (τ_p, τ_q) depend on anticipated future policy choices $(\tilde{\tau}'_p, \tilde{\tau}'_q)$ for two reasons. Firstly, the tax rates are important for them, as given returns to wages, the tax rates will determine disposable income. Secondly, τ'_q will determine q', which for parents wanting to send their children to college could be of importance. Current policy choices and the current distributions of households are mapped via G into the distributions of households in the subsequent period,

[9] Voters only decide on proportional tax rates, not on the actual shape of the tax function. Holter (2014) and Herrington (2014) find that the progressivity of the tax schedule plays an important role. Additionally, Herrington finds that public expenditures on early education are not always uniformly distributed across households as in my model, which could be an outcome of the political economy, as well. Due to computational complexity I abstract from these characteristics.

such that

$$(\mu_y'(x_y'), \mu_o'(x_o')) = G(\tau_p, \tau_q, \mu_y(x_y), \mu_o(x_o)). \tag{1.26}$$

It is assumed that when voting on the preferred policy in t, they take the anticipated value of $t+1$ to be the same as in t and ignore the impact their choice will have on the future policy choice. Since agents are atomistic this is a plausible assumption as they do not influence the outcome by themselves. In the steady state equilibrium agents will have rational expectations. Agents take correctly into account, how current policy choices affect aggregate decisions and, hence, consider how g and the function $q(h_o)$ react.

The welfare function W is composed of the weighted (remaining) lifetime utility of the young and old households and is maximized over the set of tax rates S. Assigning ω_y as the weight of the young and ω_o as the weight of the old household in the voting process, which depend on the level of education b, the problem is

$$Z(\mu_y(x_y), \mu_o(x_o)) = \underset{\tau_p, \tau_q \in S}{arg\,max} W \tag{1.27}$$

$$W = \int \omega_y(b) V_y(x_y) d\mu_y(x_y) + \int \omega_o(b) V_o(x_o) d\mu_o(x_o). \tag{1.28}$$

where $S = \left\{(\tau_p, \tau_q) \in [0,1]^2 | 0 \leq \tau_p + \tau_q \leq 1\right\}$. Since W is strictly concave in $\tau = (\tau_p, \tau_q) \in S$ the solution is unique.

1.3.6 Timing

1. Ability a of the offspring and earnings shock ζ_y are realized.

2. Voting on the contemporaneous tax rates takes place. Agents have expectations of future policies and do not vote strategically. When voting on preferred policy in t voters assume their choice will be in effect in $t+1$ and ignore the impact their choice will have on the future policy choice. Voters anticipate how their decision will affect economic decisions.

3. The investment in public early education and the college subsidy determine investment decisions of young and old parents.

1.4 Equilibrium

Let total consumption of young and old parent households be $C_y = \int c_y(x_y)\, d\mu_y(x_y)$ and $C_o = \int c_o(x_o)\, d\mu_o(x_o)$.

DEFINITION 1. *Given prices, policies, and tax rates, V_y solves the functional equations* (1.18) *satisfying* (1.19), (1.20), *and* (1.21), *while V_o solves* (1.10) *satisfying* (1.11), (1.12), (1.14), (1.15), *and* (1.16), *with $c_y(x_y)$, $c_o(x_o)$, $s(x_o)$, and $e(x_y)$ as associated policy functions.*

1. *Price w satisfies*

$$w = A \tag{1.29}$$

2. *Goods market clearing:*

$$Y = AH \tag{1.30}$$

$$Y = C_y + C_o + E + F + g \tag{1.31}$$

$$E = \int e(x_y)\, d\mu_y(x_y) \tag{1.32}$$

$$F = T \int s(x_o)(\pi(\hat{a})\bar{n} + (1-\pi(\hat{a}))\underline{n})\, d\mu_o(x_o) \tag{1.33}$$

3. *Labor market clearing:*

$$H = \int h_y d\mu_y(x_y) + \int h_o d\mu_o(x_o) + \int ((1-s(x_o)) + s(x_o)(\pi(\hat{a})(1-\bar{n})\bar{\theta} + \\ + (1-\pi(\hat{a}))(1-\underline{n})\underline{\theta})\hat{a}\, d\mu_o(x_o) \tag{1.34}$$

4. *The government balances budgets* (1.24) *and* (1.25), *which determines g and q.*

5. *The laws of motion Φ which map from state $x_y = (h_y, a, b)$ of the young to state $x_o' = (h_o', \hat{a}', a', b')$ of the old in the following period, such that $\mu_o'(x_o') = \Phi(\mu_y(x_y))$, are given by* (1.2) *and* (1.7). *The laws of motion Ω which map from state $x_o =$*

$(h_o, \hat{a}, a, b)$ *of the old to state* $x'_y = (h'_y, a')$ *of the young in the following period,* $\mu'_y(x'_y) = \Omega(\mu_o(x_o))$*, are given by* (1.1)*,* (1.5)*, and* (1.6)*.*

6. *The tax rates* (τ_p, τ_q) *to finance public education expenditures on early education* g *and the individual college subsidy* q*, respectively, are given by* (1.27)*.*

A stationary equilibrium is a competitive equilibrium in which policy functions, as well as subsidies and tax rates, are constant. It is a fixed point in the mapping (1.26)*, such that the expected tax rates* $\tilde{\tau}_p$ *and* $\tilde{\tau}_p$ *are equal to the solution for* τ_p *and* τ_p*, respectively, which are the solution to* (1.27)*. Also the distributions* $\mu_y(x_y) = \mu'_y(x'_y)$ *and* $\mu_o(x_o) = \mu'_o(x'_o)$ *are stationary, hence the distributions can be summarized by* $\mu(x)$*.*

1.5 Model parameterization

In order to analyze the US economy and conduct cross-country experiments the model is calibrated to the US benchmark economy by matching facts on inequality, mobility, and public and private education expenditures. The model is governed by 21 parameters summarized in Table 2.1. Eight parameters are chosen from a priori information or are standard in the literature, while the remaining are determined in the calibration by minimizing the squared distance between model output and 13 data moments. Of the targeted data moments two relate to public expenditures on education, two to private expenditures on education, three to college statistics, and six relate to the distribution of earnings.

1.5.1 Independently chosen parameters

One period in the model is equivalent to 16 years. The discount rate is standard at 0.96 per year, which results in β being set to 0.52. I choose a standard value in the consumption literature of 1.5 for the intertemporal preference parameter σ . The firm productivity parameter A, and hence the return per unit of human capital, is normalized to unity. College completion requires four years of attendance, which given a period length of 16 years translates into $\bar{n} = 0.25$. Supported by evidence of Stinebrickner and Stinebrickner (2007) dropping out occurs after two years of college attendance ($\underline{n} = 0.125$). The

increases in earnings through the lifecycle are determined by the earnings ratio of working males in the given age groups. For η_0 I take the ratio between those aged 33 and 48 to those between 22 and 32, whereas for η_1 I take the ratio of males aged between 49 and 64 to those between 33 and 48. Using the IPUMS data of 2011, the increases of earnings over the lifecycle are determined to be $\eta_0 = 1.8$ and $\eta_1 = 1.1$.

1.5.1.1 Voting

Participation in the elections of the president and of congress is highly correlated with the level of education in the US. Since it is more likely for the higher educated to vote, politicians might be better off catering to their interests. In frameworks based on the median voter theorem, this has been incorporated in models by Bénabou (2000) and Ichino et al. (2010) to account for the fact that the decisive voter might not actually be the median voter. Recent research points out the effects of skewed voter turnout and a bias in responsiveness towards policy preferences of the affluent (e.g., Gilens 2012, Schlozman, Verba, and Brady 2012, Bonica, McCarty, Poole, and Rosenthal 2013). Additionally, the older an individual, the more likely he is to cast his vote. To account for these potential biases, I use the 2006 voting supplement of the CPS to compute the share of eligible individuals by education and age group that casted their vote in the 2006 election of congress. The patterns of voter turnout by age and education are fairly constant across the three elections of congress and the three presidential elections of the years available in the CPS data (1996-2006), as can be seen in Table 1.7 in the Appendix. The political science literature has established a cross-country relationship between inequality and voter turnout (e.g., Ljiphart 1997), but to my knowledge there is no empirical evidence of this relationship holding or being causal within a country. Given that additionally no theoretical model is able to explain patterns of voter turnout consistently, we can take participation as exogenous.[10]

The voting weight for the old parent household ω_o is composed of the sum of the weight assigned to the parents as well as to the offspring. This accounts for the fact that they are on the verge of completing the age of eligibility to vote at the beginning of the period. Since the old offspring has not taken all educational decisions yet, individuals

[10]For more on this discussion see Section 1.7.1.

in this age group are only weighted by their age-specific weight, while disregarding education to have an exogenous voting weight. Otherwise, tax rates could alter decisions, and thereby the voting weights of the old offspring. The weights assigned to each age and education level are displayed in Table 1.1.

Voting preferences are determined at the household level because all decisions in the

Table 1.1: Voting weights

Age	High school or less	Some college	College
18 - 32	.27	.27	.27
33 - 48	.32	.48	.64
49 - 64	.47	.63	.77

Datasource: CPS November voting supplement 2006

model are determined at the household level. Additionally, Niemi and Jennings (1991) find that parents play a major role in determining the initial political direction in the early adulthood of their offspring. I abstract from voters' influence varying by income. Given the strength of lobbies and the importance of election campaign financing in the US, this could play an important role.

1.5.2 Calibrated parameters

In the following the targets of the remaining parameters are specified. The parameters are related to costs and expenditures of education, educational decisions and outcomes, and the distribution of earnings.

1.5.2.1 Education costs and expenditures

In 2009 the share of GDP dedicated to public early education was 3.9% (OECD 2011a), which is targeted by γ, the productivity parameter of early education investment. The parameter determining the curvature of the "warm glow" function, ξ, is anchored by the share of household wages parents spend on early private education, estimated to be a total of 2,198 US$ per child in 2007 according to Kornrich and Furstenberg (2013).[11]

[11] Their analysis includes expenditures on education, child care, and other miscellaneous goods and services, such as games and instruments, for children.

Using the IPUMS data for 2007, I calculate the average household earnings per child, where no parent is older than 48 and at least one parent is older than 32, and determine it to be 41,068 US$. Therefore, I estimate that an average parent spends 5.4% of his earnings on private early education of a child.

In 2010-2011 average costs of one year of undergraduate full-time studies at a 4-year institution are 22,092 US$ according to the U.S. National Center for Education Statistics (NCES).[12] Given a GDP per capita of 49,800 US$ in 2011 on average the costs of one year of college rack up to 44% of GDP per capita, which I target via the tuition cost T. The total expenditures of households on tertiary education sum up to 1.1% of total GDP according to the OECD (2011a), which is targeted by the "warm glow" parameter ϕ, the importance parents attach to the human capital outcome of their children. The share of students receiving federal grants is retrieved from the 2011 Digest of Education Statistics.[13] Of all full-time undergraduate students 64% received financial aid through grants in 2007-2008, which pins down κ, the parameter determining the slope of the college subsidy with respect to parental earnings.

1.5.2.2 Education decisions and outcomes

In 2009 of those completing high school 71% enrolled into college according to the National Science Foundation, while the high school completion rate is estimated at 88% (Heckman and Lafontaine 2010).[14] This results in 62% enrolling into college and 38% of non-college workers, which I approach through ψ_0, a parameter of the probability of college completion. The other parameter of the college completion function, ψ_1, is pinned down by the fraction of individuals completing college, 58% of first-time, full-time students who enrolled at a 4-year institution in fall 2004 according to the NCES.[15] I use the IPUMS data of the year 2011 to calculate the college completion as well as the dropout premium. The ratio of the average earnings of men aged 33 to 48 with at least four years of college to those with no college education is 2.53, while the ratio of those with less than four years of college to those with no college education is 1.32.

[12] For more information see the website http://nces.ed.gov/fastfacts/display.asp?id=76.
[13] See http://nces.ed.gov/programs/digest/d11/tables/dt11_353.asp.
[14] See http://www.nsf.gov/statistics/seind12/c1/c1s4.htm.
[15] See http://nces.ed.gov/fastfacts/display.asp?id=40.

Table 1.2: Benchmark model parameters

Description	Independently chosen Parameter	Value	Reference
Coefficient of relative risk aversion	σ	1.5	Standard
Discount factor	β	0.96	Standard
Time for college completion	$\bar{n}$	4 years	Standard
Time for college dropout	$\underline{n}$	2 years	Stinebrickner and Stinebrickner 2007
Lifecycle wage premium	η_0	1.8	IPUMS 2011
Lifecycle wage premium	η_1	1.1	IPUMS 2011
Voting weights of young HH	ω_y	See Table 1.1	CPS voting supplement 2006
Voting weights of old HH	ω_o	See Table 1.1	CPS voting supplement 2006

Description	Calibrated Parameter	Value	Target	Data
Elasticity wrt early education	γ	0.34	Public early education/GDP	0.039
Tuition cost per period	T	0.44	Annual tuition costs/GDP per cap.	0.44
Slope college subsidy wrt earnings	κ	0.46	Share college students with grant	0.64
"Warm glow"	ϕ	1.375	Private college expenditure/GDP	0.011
Curvature of "warm glow"	ξ	0.78	Private early educe./Mean earnings	0.054
College completion wrt ability	ψ_0	0.29	Fraction attending college	0.62
College completion wrt ability	ψ_1	1.02	College completion rate	0.58
College completion premium	$\bar{\theta}$	1.31	Average college premium	2.53
College dropout premium	$\underline{\theta}$	0.87	Average dropout premium	1.32
Persistence ability transmission	ρ	0.25	Intergenerational earnings persist.	0.40
Level effect of HC prod. func.	χ	1.075	Gini before tax	0.39
STD of noise in ability trans.	σ_a	0.50	Variance of log hourly wages	0.47
Magnitude of market luck shock	ζ	0.375	Share earnings variance post-educ.	0.39

These are targeted in the model by the college completion and dropout premium $\bar{\theta}$ and $\underline{\theta}$, respectively.

1.5.2.3 Earnings

Intergenerational mobility in the US, as captured by the coefficient of the intergenerational earnings persistence, is 0.4 (e.g., Solon 1992, Zimmerman 1992, Solon 1999) and is targeted by the parameter for the intergenerational transmission of innate ability ρ, the coefficient of the autoregressive process. The variance of the random shock in the transmission of innate ability σ_a is linked to the variance of log hourly wages of males. In the US in 2005 the variance of log hourly wages of males was 0.47 (Heathcote, Perri, and Violante 2010). The Gini coefficient of hourly male wages in the US, my measure for earnings inequality, is 0.39 in 2005 (Heathcote et al. 2010) and is targeted by χ, which is the parameter in charge of the level effect of human capital creation at the early stage.[16] The relevant moment for the post-education earnings shock ζ is the share of the variance in earnings of 0.39 which is not explained by initial conditions, such as education, when entering employment (Huggett et al. 2011).

1.6 Benchmark economy

The model performs well at replicating the fraction of individuals going to college, the dropout rate, the share tuition costs of one year of college relative to GDP per capita, and the share of GDP households dedicate to tertiary education, as can be seen in Table 2.2. The average college and dropout premium are overestimated, whereas the share of college students receiving federal grants and intergenerational earnings persistence are slightly underestimated. While the correlation in innate ability ρ is 0.25, the intergenerational elasticity in earnings is 0.39. A large part of intergenerational mobility is

[16] The focus is on the pre-tax and pre-transfer Gini in form of hourly wages as I am trying to capture inequalities coming about through education and human capital accumulation and their respective returns in the market.

Table 1.3: Calibration of the US economy

Target	Data	Model
Public early education exp./GDP	.039	.044
Annual tuition costs/GDP per capita	.44	.42
Share of college students with grants	.64	.57
Private college expenditure/GDP	.011	.010
Private early education exp./Mean earnings	.054	.061
Fraction attending college	.62	.62
College completion rate	.58	.59
Average college premium	2.53	2.91
Average dropout premium	1.32	1.61
Intergenerational earnings elasticity	.40	.39
Gini before tax	.39	.38
Variance log wages	.47	.44
Share of earnings variance post-schooling	.39	.34

determined by the underlying structure of the model.[17]

In the calibration public early education investment accounts for 4.4% of GDP, which is slightly higher than the 3.9% in the data. This can be attributed to the high elasticity of returns to early education expenditures ($\mu = 0.34$). The elasticity of returns to education expenditures has been estimated to be lower, taking values around 0.2 (e.g. Card and Krueger 1992, 1996). However, these estimates might be underestimating the returns, as they exclude private investments (remember in the model investments are $e+g$) and tend to have small time horizons. Increasing a teacher's salary might increase motivation marginally, but in the long-run this measure would also improve the talent pool of teachers.

The model estimates that 57% of the students receive grants, whereas in the data 64% of full-time undergraduates do so. Even though not targeted by the calibration, I find that those receiving grants on average can cover 36% of tuition, which is fairly close to the 32% (7,100 US$) in the data of the NCES.[18] In the model public expenditures account for 21% of total tertiary education expenditures, which is far from the

[17] The finding that the correlation of the permanent component of wages of brothers is higher than for their physical attributes such as height and weight (Mazumder 2008), which are arguably transmitted more directly via genes, casts doubts on accepting "nature" as the sole explanation of intergenerational persistence in earnings.

[18] According to a Sallie Mae-Ipsos Report in 2013 grants and scholarships pay for 30% of costs of college attendants.

41% in the data. This discrepancy originates in the assumption that colleges do not require input in the model, implemented due to the emphasis on financial aid given tuition costs. Since colleges in the model do not require inputs to produce, their productivity is not dependent on funding. According to the OECD (2012) public expenditures on tertiary education account for 1.3% of GDP in the US, while 18.5% of these funds are spent on scholarships and grants. Therefore, the US invests 0.24% of GDP in grants and scholarships, which is close to the 0.25% endogenously determined through probabilistic voting in the model.

The college premium parameter $\bar{\theta}$ is 1.31, but since it is multiplied with acquired ability the actual average college premium is 2.91, 15% higher than what we observe in the data. The college dropout parameter $\underline{\theta}$ is 0.87. The fact that the dropout premium is smaller than unity can be interpreted as dropping out being a negative signal in the labor market and/or the time spent in college without graduating is lost in terms of learning on the job, leading to a relatively lower work experience, which is penalized in the labor market.

The "warm glow" parents feel for the human capital accumulation of their offspring ($\phi = 1.375$) is close to the value of 1.37 calibrated by Blankenau and Youderian (2012). This is supported by comments by Keane and Wolpin (2001), Johnson (2012), and Lochner and Monge-Naranjo (2012) pointing out the high parental valuation of children's education given that the young would not attend college without attendance-contingent transfers by their parents even if credits were abundant.

1.6.1 Performance beyond targets

The model does not only perform well at replicating targeted moments, but also comes close to a range of other moments related to inequality and intergenerational mobility. The ratio of mean to median earnings is 1.29 compared to 1.34 in the data using only employed males between 18 and 64 of the IPUMS 2011. For male earnings in 2005 Heathcote et al. (2010) determine the P50/P10 percentile ratio to be 2.31 compared to 2.54 in the calibration, and the P90/P50 percentile ratio to be 2.33, close to 2.23 in the benchmark economy. Jäntti et al. (2006) estimate the mobility trace index at 0.86 in the

US.[19] In the benchmark economy the mobility trace index is 0.89, marginally higher than the value obtained by Jäntti et al. (2006).

Additionally, I compare a 3x3 matrix of intergenerational transition probabilities be-

Table 1.4: Earnings transition matrix

	Data			Model		
FatherSon	Bottom	Middle	Top	Bottom	Middle	Top
Bottom tercile	.48	.34	.19	.47	.36	.18
Middle tercile	.30	.37	.34	.38	.37	.29
Top tercile	.21	.29	.50	.15	.31	.54

Note: The left panel is calculated using permanent earnings of fathers and sons of the PSID between 2000-2008.

tween earnings terciles of the model to those I obtain using fathers and sons of the PSID between 2000 and 2008. Correcting for lifecycle effects by taking the residuals of a regression controlling for age and age square and year fixed effects, and then averaging observations over various years, I find, as displayed in Table 1.4, that persistence is especially high at the bottom and the top, where about half of all sons end up in the same earnings tercile as their father. Comparing the data to the transition matrix obtained for the benchmark case, one can see that the model replicates intergenerational dynamics accurately. While for fathers in the bottom earnings tercile it matches the data closely, it benignly overestimates persistence for sons who's fathers are in the top tercile, where 54% end up in the same cell as their father, compared to only 50% in the data. For sons of the middle quintile the model overestimates downward mobility, which is due to the punishing effect of dropping out of college.

1.6.1.1 Credit constraints

In Figure 1.5 the correlation between parental earnings and college attendance and completion becomes apparent. The model accurately replicates the pattern of college attendance conditioned on the parental earnings quintile determined by Bailey and Dynarski (2011) using the National Longitudinal Survey of Youth (NLSY). In the data more chil-

[19] The trace index is calculated using a 5x5 matrix P of transition probabilities for sons to move to a given earnings quintile conditioned on the earnings quintile of their father. Given the matrix the index is calculated as: $M_T = \frac{5-tr(P)}{5-1}$ where tr is the trace of the matrix P. The index is bound below by zero and is increasing in mobility.

dren of the bottom quartile drop out of college, which could be linked to two causes. First, despite the finding of Stinebrickner and Stinebrickner (2007) that dropout is usually explained by ability, as implemented in the model, financial constraints tightening during college attendance might force some students to drop out. Second, in the model abilities and completion probabilities are perfectly known, whereas in reality some students might realize that they are not prepared for college once they are actually attending college.

In order to determine the share of individuals that are credit constrained in their college

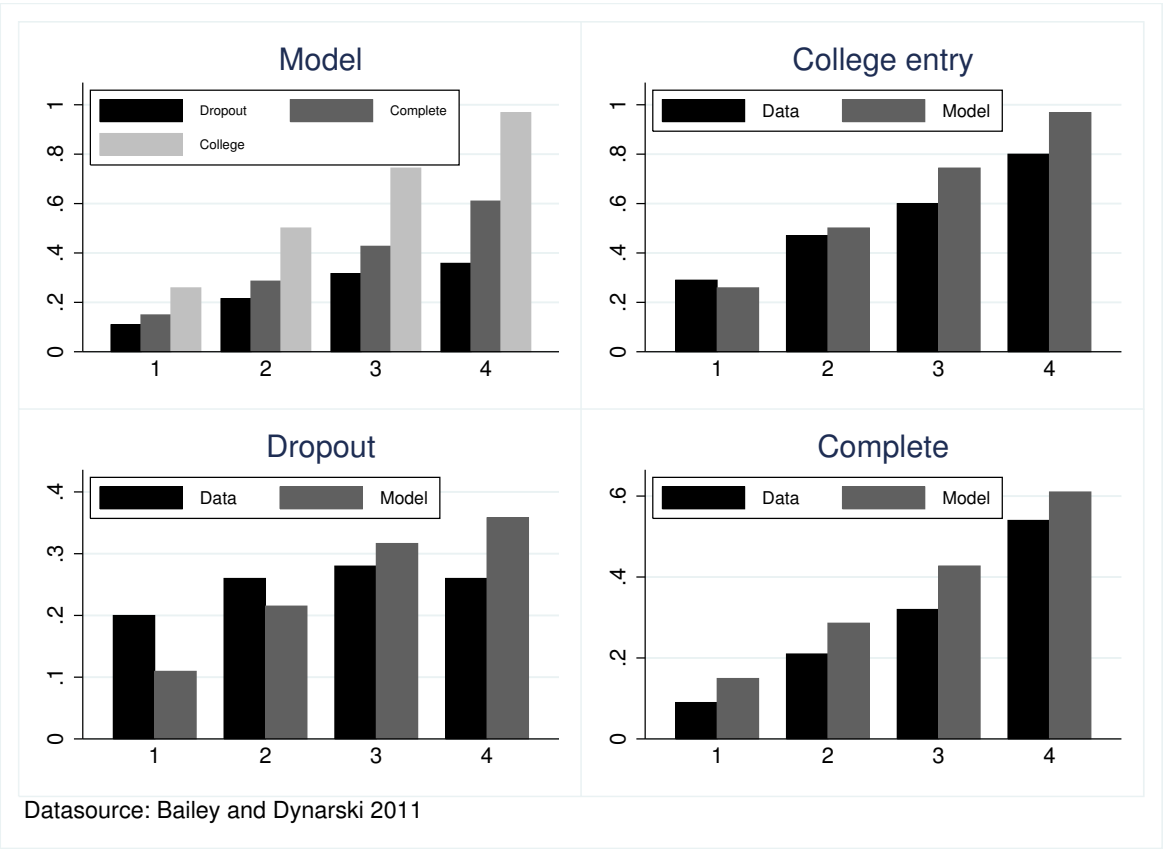


Figure 1.5: Data vs model of college attendance and completion rates (y-axis) by parental earnings quartile (x-axis)

decision I use the methodology of Carneiro and Heckman (2002). They define as credit constrained the sum of the gaps between the percentage of the highest income quartile enrolled for each ability tercile and the percentage enrolled in the remaining income quartiles of the given ability tercile. When undertaking the same exercise in terms of

acquired ability the model exhibits a large share of 23.2% of credit constrained individuals, compared to only 5.2% identified by Carneiro and Heckman (2002) using a sample of white males of the 1979 cohort of the NLSY. Belley and Lochner (2007), using the NLSY97, find that family income has become a much more important component of the college enrollment decision in the recent past. Given that tuition costs have more than doubled in real terms since the 80's this is not a surprising development. Recent estimates using the same approach have increased to 16.6% (Bohacek and Kapicka 2012), 24% (Winter 2013), and an upper bound of half of all children (Brown, Scholz, Seshadri 2012).[20] In the benchmark economy the same exercise in terms of innate ability instead of acquired ability, reveals that 27.5% of individuals are credit constrained in their college decision when one does not condition on early investments. In Figure 1.11 in the Appendix one can see how early investment increases with innate ability and parental earnings.

1.6.2 The role of political economy

With few exceptions (e.g., Ichino et al. 2010), the relationship between inequality and mobility is analyzed in isolation from the political economy. In order to develop a better understanding of the underlying mechanism, I illustrate voter preferences by age and earnings. In the left panel of Figure 1.6 the aggregate utilities as a function of τ_p of young and old households are displayed. For old households there is no gain from investment in early education, hence their utility is strictly decreasing in τ_p. For young households public early investment is beneficial due to its effect on their children, however at a decreasing rate. At some point the substitution effect of own earnings outweighs the benefits from the additional education investment and aggregate utility of the young household begins to decline. The aggregate utility of old households as a function of τ_q is presented in the right panel of Figure 1.6. Aggregate utility of the old is increasing until the function has a kink where tuition costs of all college attendants can be fully covered by the revenues, after which it is strictly decreasing.

[20] The share of potentially credit constrained children estimated by Brown et al. (2012) is so high due to the assumption that not all parents are willing to contribute the expected amount according to the expected family contribution schedule, which is used as a guide for the calculation of individual financial aid.

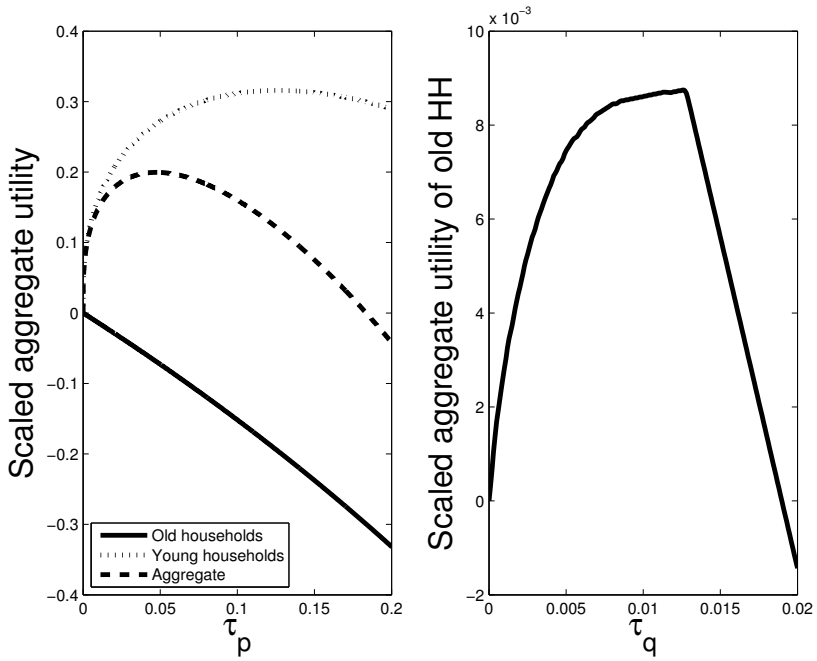
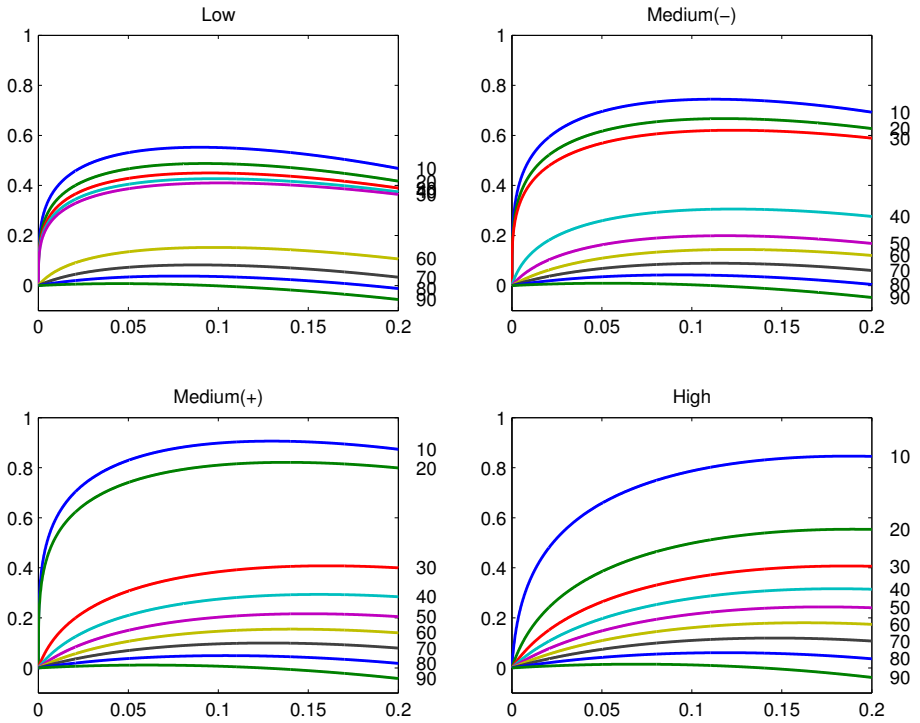


Figure 1.6: Aggregate utility of households as a function of τ_p (left) and τ_q (right).

In Figure 1.7 the utility of the 10^{th} to the 90^{th} earnings percentile by varying innate ability levels of the offspring is shown as a function of τ_p. The calculations are made given steady state private investments in early education, which is important because public and private education are substitutes. Comparing the trajectories for a low (top left panel) and a lower medium (top right) ability child, one can tell that parents of low ability children with earnings in the bottom tercile are less responsive to increases in τ_p than parents of children with slightly higher ability. This is due to low returns to investments in low ability children. For wealthier parents this is not necessarily the case because they have already allocated private funds to the child's education, wherefore marginal returns to investment are lower. In the bottom right panel, where children are of high ability, one can see that the poorest deciles are still in favor of a higher τ_p. This means that even when children are of high ability poor parents are still underinvesting in their children despite the high returns.



The number on the right hand y−axis is the earnings percentile within young parents, whereas the left hand y−axis is utility scaled to zero. Low, medium(−), medium(+), and high refer to the innate ability of the child.

Figure 1.7: Utility of young households by earnings percentile (y-axis) as a function of τ_p (x-axis) for varying levels of innate ability of the offspring

1.6.2.1 What if the college premium changes in the US?

Many argue that the digital revolution caused the rise in the college premium through a skill-biased demand shift combined with globalization (e.g., Katz and Murphy 1992, Krusell, Ohanian, Rios-Rull, and Violante 2000, Acemoglu 2002, Hornstein, Krusell, and Violante 2005). Given that both, technological development and globalization remain dynamic, further shifts in the college premium could take place. In order to demonstrate the effects of voting, exogenous movements of the college premium parameter $\bar{\theta}$ from 1.21 to 1.41 are simulated. The effects on intergenerational earnings persistence are illustrated in the left panel, and on earnings inequality in the right panel of Figure 1.8, where the dashed line represents the trajectory when taxes are fixed as in the benchmark model and the solid line displays what happens when taxes are voted on.

When the college premium declines, college enrollment decreases and the compo-

sition of college attendants changes. Fewer less well off households send their kids to college as the prize of the college lottery diminishes. Therefore, public support for college subsidies wanes. When the college premium is low, richer households do not have high incentives for investing in education, thus intergenerational earnings persistence decreases. Also rich households are not as rich, hence the distance and the incentive effect lead to a reduction in intergenerational earnings persistence. However, these reducing forces are countered by a lower chosen tax rate for early education. As a consequence the intergenerational earnings persistence is only 4% lower than in the benchmark case.

An exogenous increase in the college premium acts as a boost to the economy. The

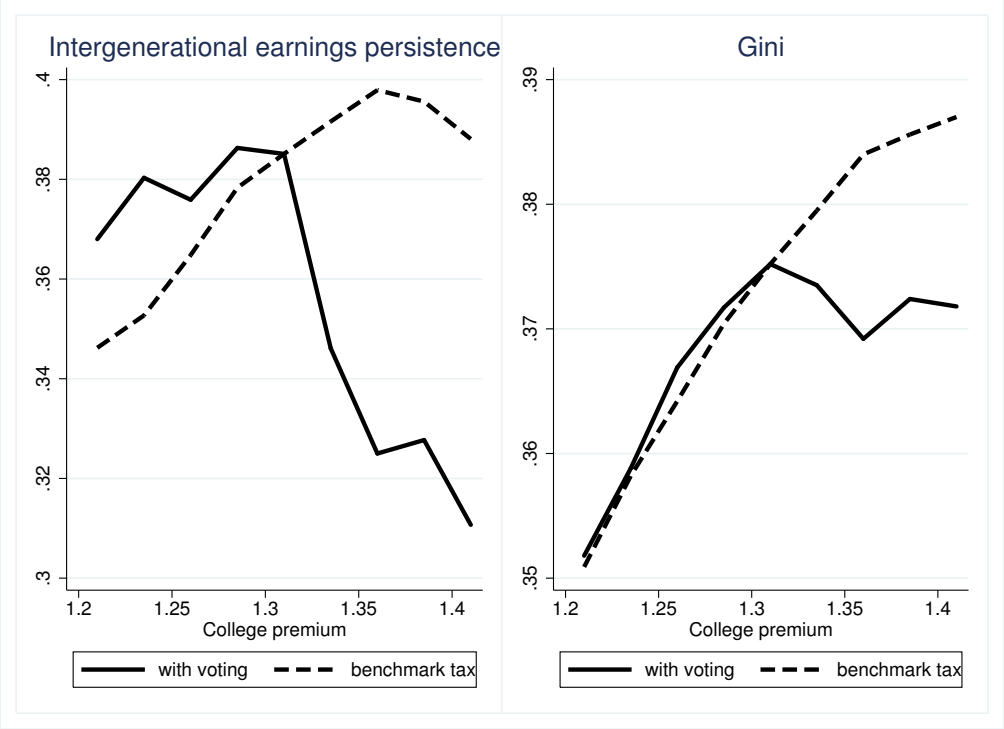


Figure 1.8: The effects of voting on earnings persistence and inequality

chosen tax rate τ_p dedicated to early public education increases. Thereby, the improved early education combined with the higher payoffs of college completion augment the share of college attendants. Given that amongst those entering, many are from less well off households, people vote for higher college subsidies, hence increasing the fraction of college students receiving grants and thereby alleviating the barriers to access. The

effect of enrollment on increased public financing has been emphasized by the political science literature (e.g., Ansell 2010, where it is shown that in a range of countries left-leaning parties only started supporting public expenditures on tertiary education after more children from poorer families began attending college). The intergenerational earnings persistence drops down to 0.31 when $\bar{\theta}$ takes its maximum value of 1.41, while the Gini declines only marginally to 0.37.

This simulation exhibits, though, that the general assumption of a positive relation between returns to tertiary education on the one side, and inequality and intergenerational earnings persistence on the other, does not hold when public expenditures are determined through voting. During an interval where incentives for investment increase and poorer households suffer from credit constraints, intergenerational earnings persistence and inequality increase. Then they peak before returns become so high that it is worth taking the risk of enrollment for less well off individuals, after which the intergenerational earnings persistence and the Gini decline.

I compare these outcomes to simulations where the tax rates are fixed as in the benchmark model. Here we do not observe the large decline in intergenerational earnings persistence when the college premium increases. However, at the highest simulated college premium intergenerational persistence decreases marginally, due to increased incentives for less wealthy households to dedicate a large share of their earnings to college tuition.

In order to decompose this effect I simulate the exogenous variations in the college premium while allowing households to vote only on one single stage of public education financing in isolation, as displayed in Figure 1.21 in the Appendix. For a lower college premium voting has a smaller impact, whereas considerable differences emerge towards the maximum value of $\bar{\theta}$. When only allowing for voting on college subsidies intergenerational earnings persistence reduces to 0.38, whereas it decreases to 0.36 when only permitting voting on early education. Therefore, 4% of the drop the political economy in earnings persistence can be attributed to the endogenous changes in public financing of late education alone, while 32% can be accomplished by endogenous changes in early education. This decomposition highlights the importance of sound coordination of reforms in early and late education, as the increases in public subsidies at both levels combined allow for a multiplicator effect, which decreases intergenerational earnings persistence by three times the amount compared to when public subsidies at only one

level are adjusted endogenously.

Pre-tax earnings inequality is affected less by the inclusion of voting on taxes. For the largest value of the college premium the Gini decreases by 4% compared to no voting, with only a small difference to when individuals only vote on one of the levels of education financing.

1.6.2.2 Voting policy experiments in the US

Voter turnout in the US is lower and more skewed towards the educated than in most OECD countries. Therefore, one possible policy to counter this would be compulsory voting, as already exists in a number of countries (e.g., Argentina, Australia, Brazil, Peru). I simulate the policy experiment by imposing mandatory voting for all. If everybody were to cast their vote, public education funding of early education would increase from 4.4% to 5.2% of GDP and financial aid expenditures on college would augment from 0.25% to 0.28% of GDP, providing financial aid to 59% of the college students, who would account for 66% of the population. As a consequence the Gini reduces marginally to 0.37, whereas earnings persistence drops to 0.36.[21]

Another policy, which recently has received attention in the public debate, is the extension of electoral franchise to children as of birth.[22] The practical implementation discussed would include parents voting for their children. I simulate this by doubling the benchmark voting weight of young parent households.[23] As a result inequality reduces to a Gini of 0.36, whereas intergenerational mobility surges due to the increases in public early education expenditures, thus reducing earnings persistence to 0.21.

[21] This is not to claim that these changes would be immediate, as surely politicians would take time to identify preferences of the electorate, which would take several elections.

[22] The extension of franchise to children is commonly referred to as "Demeny voting", named after Paul Demeny, the author of a paper suggesting half a vote for children to counter low-fertility (Demeny 1986).

[23] This is debatable given that increased voting weight raises incentives to vote, which thereby could feedback into turnout of young parent households as well. Given that voter turnout is not modelled in the framework, I refrain from speculating about the magnitude of the possible increase in turnout among young parents.

1.7 Cross-country differences

As visible in the "Great Gatsby Curve" inequality and intergenerational mobility are negatively associated across countries. In order to identify potential drivers of this relationship I conduct experiments exploiting cross-country differences in voter turnout and tertiary education characteristics jointly and in isolation.

1.7.1 Voter turnout

To decompose the drivers of cross-country differences in inequality and intergenerational mobility, I run the simulations with the US benchmark specification while only adapting one of the explanatory factors separately, either the voting weights or the characteristics of tertiary education. Voter turnout shows strong variations across age groups and levels of education across countries. I am agnostic here on whether this is due to culture or political institutions.[24] I use this variation to adjust the voting weights in the probabilistic voting process and explain differences in individual earnings inequality and intergenerational earnings persistence. The voter turnout by age group and education is determined using the ESS 2010 for European countries and the Canadian Election Study 2010 for Canada, as displayed in the Appendix in Table 1.10, and assign these weights to the respective age groups and education levels in the probabilistic voting mechanism of the benchmark model.

Voter turnout alone on average explains 21% of the difference in the Gini and 23% of the difference in intergenerational earnings persistence to the US. The results are summarized in Table 1.5 and in Figures 1.14, 1.15, and 1.17, in the Appendix. The model performs well in terms of replicating data moments of the Gini, intergenerational earnings persistence, and public expenditures on non-tertiary and tertiary education. In

[24] There is strand of literature attempting to explain cross-country differences in voter turnout and specifically low turnout in the US (e.g., Wolfinger and Rosenstone 1980, Powell 1986, Jackman 1987, Blais 2000, Perea 2002). General explanations for cross-country differences range from legal differences, such as compulsory voting or voting facilitation through postal or advanced voting, organizational factors and electoral systems, such as number of parties, party-group alignment, and proportional representation, and population size. Specifically concerning the US there seems to be consensus that the complexity of registration is one reason for the low turnout. Blais (2000) summarizes that there is a lack of solid explanations and there are few robust findings, which explain cross-country differences, with compulsory voting being an exception. So far neither theoretically nor empirically has anybody found a consistent answer to the voting paradox.

the following the results of each country are specified.

The Swedish society is characterized by a relatively low intergenerational earnings persistence of 0.27 (Corak 2012) and a pre-tax Gini of individual earnings of 0.32 (Domeij and Floden 2010). Swedes of all age groups are similarly likely to vote, and voter turnout is not biased towards the highly educated as in the US. This voting pattern causes an increase of τ_p from 4.4% in the benchmark case to 5.3%, while τ_q increases from 0.26% to 0.37%. These higher levels of public education funding decrease the intergenerational earnings persistence by 10% to 0.35 and the Gini to 0.37. Thereby, the voter turnout can explain 28% and 7% of the gaps between Sweden and the US in earnings persistence and inequality, respectively. These improvements in inter- and intragenerational equity are accompanied by a 4% increase in aggregate consumption. In Sweden inequality increased over the 90's, however to a lesser extent than in the US. Similarly, intergenerational mobility remains high despite increases in earnings inequality. The model indicates that the political economy, and more specifically voter participation, might be able to explain why the increase in the skill premium in the 90's had a less harsh effect on inequality and mobility in Sweden than in the US.

In the United Kingdom the Gini of individual earnings is 0.383 (OECD 2011b), while the intergenerational earnings persistence is estimated to be 0.31 (Jäntti et al. 2006). In the UK voter turnout is increasing in age and education, but the bias is relatively lower than in the US. The simulation reveals an earnings persistence of 0.35 accounting for 40% of the gap. The Gini only drops to 0.37, but this accounts for 66% of the difference in inequality between the UK and the US. Canada has a Gini of 0.37 (OECD 2011b) and a relatively low earnings persistence of 0.21 (Corak 2012). Voter turnout patters are capable of explaining 19% of the difference in earnings persistence to the US and 23% of the difference in earnings inequality. Denmark is amongst the most equal countries and has the lowest intergenerational earnings persistence of the OECD. The Gini of individual pre-tax earnings is estimated to be 0.294 (OECD 2011b), while the intergenerational earnings persistence is 0.15 (Corak 2012). In Denmark voter turnout is relatively high across all age groups and levels of education resulting in a Gini of 0.37 and intergenerational earnings persistence of 0.35, accounting for 6% and 19% of the gaps, respectively. Norway has a similarly low level of intergenerational earnings persistence of 0.17 (Corak 2012) and a Gini of 0.35 (OECD 2011b). In Norway voter turnout explains 10% of the difference in inequality and 13% of the difference in inter-

generational earnings persistence. Finland matches the pattern of the other Scandinavian countries with a pre-tax earnings Gini of 0.31 (OECD 2011b) and an intergenerational earnings persistence of 0.18 (Corak 2012). For the case of Finland voter turnout patterns can account for 3% and 12% of the respective gaps.

In Germany the pre-tax earnings Gini is estimated to be 0.37 (Fuchs-Schuendeln, Krueger, and Sommer 2010) and the intergenerational earnings persistence is 0.32 (Corak 2012).[25] Both values are in the middle of the distribution of OECD countries. Voting participation in Germany is biased towards the older and more educated, but not to such an extreme extent as in the US. As expected the publicly chosen education expenditures as well as the Gini and intergenerational earnings persistence are in the middle range, closing 29% and 33% of the gaps, respectively

Table 1.5: Counterfactuals (voter turnout)

	Intergenerational elasticity			Gini		
Country	Data	Model	$\Delta_{explained}$	Data	Model	$\Delta_{explained}$
US	.40	.385		.39	.375	
Sweden	.27	.348	.28	.32	.370	.07
UK	.31	.350	.40	.383	.371	.66
Canada	.21	.349	.19	.37	.371	.23
Denmark	.15	.34	.18	.294	.369	.06
Norway	.17	.355	.13	.35	.371	.10
Finland	.18	.358	.12	.313	.373	.03
Germany	.32	.358	.33	.37	.369	.29

Note: The share of a gap between a data moment in the US d_{US} and another country $d_{country}$ explained by moments m_j produced by the model, is calculated using the following definition: $\Delta = \frac{(m_{US} - m_{country})}{(d_{US} - d_{country})}$.

1.7.2 Tertiary education and voter turnout

Countries are characterized by varying types of tertiary education. Roughly labeling the countries included in the analysis into the categories defined by Ansell (2010), the Scandinavian countries and Canada could be considered as mass public, the US, and

[25] In Germany tuition costs vary strongly by state, which is why Germany is excluded from the simulation with tertiary education characteristics.

the UK as partially private, and neither as an elite tertiary education system.[26] Using data from the OECD (2012, 2013) I calculate the college premium and tuition costs relative to the US, which are displayed in Table 1.8 in the Appendix, and assign them as exogenous factors in the model.[27] Then I calibrate the model while targeting dropout rates of each country via ψ_0, a parameter of the college completion probability function, and enrollment via the dropout premium $\underline{\theta}$. Expenditures on early education and, for those countries with non-zero tuition costs, financial aid is determined endogenously via probabilistic voting given country-specific voter turnout, which is in Table 1.10 of the Appendix and discussed in further detail in section 1.7.1.[28] On average I find that the differences in tertiary education and voter turnout combined can explain 71% of the differences in inequality compared to the US and 29% of the gaps in intergenerational earnings persistence. The results are summarized in Table 1.6 and Figure 1.9. The high correlations of 0.67 and 0.83 between data and model moments in terms of the Gini and intergenerational earnings persistence, respectively, are illustrated in Figure 1.12 in the Appendix.

In Sweden college access is free of tuition costs and is rewarded with one-third of the premium of the US. I find that differences in the characteristics of tertiary education and voter turnout explain 38% of the difference in intergenerational earnings persistence, which drops to 0.34, and 62% of the gap in the Gini, which reduces to 0.33.[29] In the United Kingdom the college premium is relatively high and about three-quarters of the US college premium, whereas average tuition costs are nearly half as cheap. While 29 % of the gap in intergenerational earnings persistence is explained, the gap in inequality is fully accounted for. Canada's average tuition costs for tertiary education are about 40% of those of the US (Belley, Frenette, and Lochner 2011), while the college premium is about half as high as in the US (OECD 2013). Also the dropout rates are substantially lower at around 30% (Finnie and Qiu 2008). The model estimates inequality to be 0.35, which is even lower than it actually is in Canada, thereby covering the

[26] Elite tertiary education system are more prevalent in low and middle income countries with high inequality, such as Chile, Brazil, or Mexico.

[27] Since not all the data for Canada is available in the OECD report I obtain them from the literature.

[28] Since financial aid in the model is a tuition discount, there is no point in voting on financing of college education when tuition costs are zero.

[29] The share of a gap between a data moment in the US d_{US} and another country $d_{country}$ explained by moments m_j produced by the model, is calculated using the following definition: $\frac{(m_{US}-m_{country})}{(d_{US}-d_{country})}$.

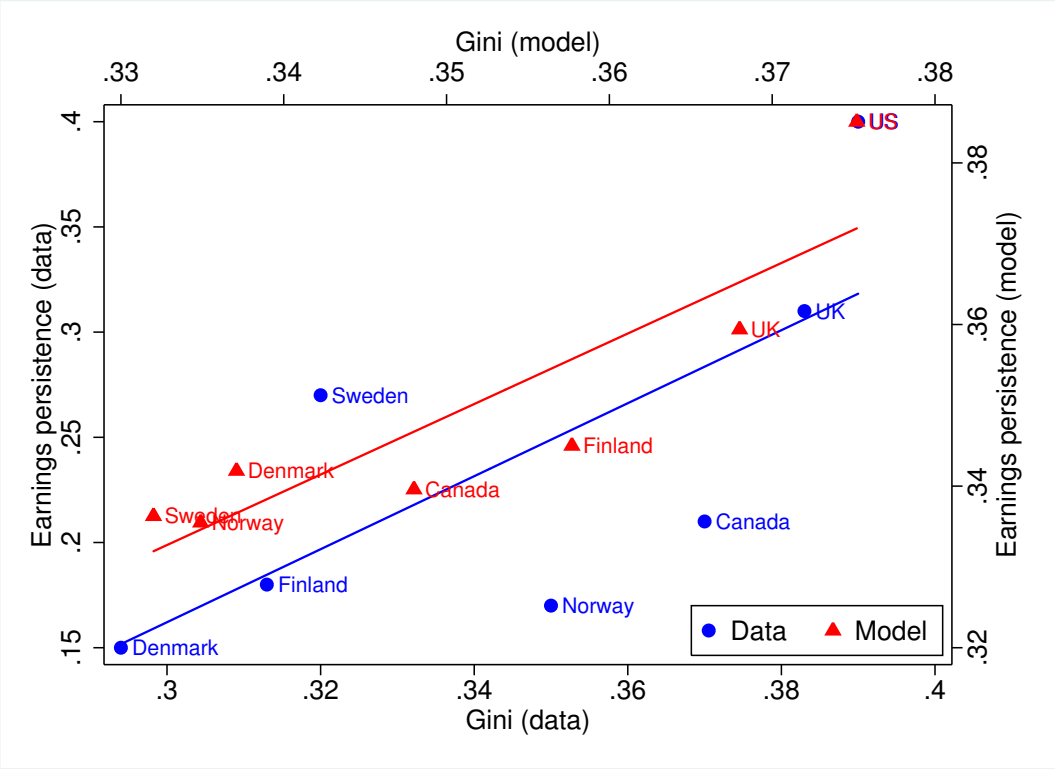


Figure 1.9: Comparing model moments with country-specific tertiary education and voter turnout by age and education (blue) and data moments (red) of the Gini versus intergenerational earnings persistence.

entire gap. Intergenerational earnings persistence drops to 0.34 explaining 24% of the difference.

In all Scandinavian countries tuition is free and, except for Finland, the college premium is about one-third of the US. In Denmark tertiary education combined with the patterns of voter turnout accounts for 17% of the difference in intergenerational earnings persistence and 40% of the gap in the Gini between Denmark and the US. Both, Gini and earnings persistence, diminish to 0.34 in the simulation. In Norway 22% of the gap in earnings persistence and the entire difference in earnings inequality are explained by the simulation. In Finland the college premium is nearly two-thirds of the US, and hence the model closes only 23% and 18% of the gaps in inequality and intergenerational earnings persistence, respectively.

Table 1.6: Counterfactuals (tertiary education and voter turnout)

	Intergenerational elasticity					Gini				
Country	Data	Ter.&Vote	$\Delta_{ter.\&vote}$	Tertiary	$\Delta_{ter.}$	Data	Ter.&Vote	$\Delta_{ter.\&vote}$	Tertiary	$\Delta_{ter.}$
US	.40	.385		385		.39	.375		.375	
Sweden	.27	.336	.38	.329	.43	.32	.332	.62	.33	.65
UK	.31	.359	.29	.369	.18	.383	.368	1.03	.370	.71
Canada	.21	.340	.24	.340	.24	.37	.348	1.36	.348	1.36
Denmark	.15	.342	.17	.347	.15	.294	.337	.40	.339	.38
Norway	.17	.336	.22	.350	.15	.35	.335	1.01	.334	.94
Finland	.18	.358	.18	.362	.11	.313	.345	.23	.360	.20

Note: The column "Ter.&Vote" refers to results where country-specific voter turnout and tertiary education charcteristics are varied, whereas the column "Tertiary" refers to results where only country-specific characteristics in tertiary education are varied. The share of a gap between a data moment in the US d_{US} and another country $d_{country}$ explained by moments m_j produced by the model, is calculated using the following definition: $\Delta = \frac{(m_{US} - m_{country})}{(d_{US} - d_{country})}$. $\Delta_{ter.\&vote}$ and $\Delta_{ter.}$ refer to the share of the difference explained by voting as well as tertiary education, and tertiary education, respectively.

1.7.3 Tertiary education

In order to isolate the importance of tertiary education I simulate the economy with the voting weights of the US, but the country-specific characteristics of tertiary education. The results are summarized in Table 1.6 and Figure 1.16 in the Appendix. For all countries the Gini is lower than when only using country-specific voter turnout, and therefore the gap in earnings inequality is closed to a larger extent. On average 65% of the differences in the Gini and 21% of the differences in intergenerational mobility are accounted for by characteristics of tertiary education. Concerning the intergenerational earnings persistence, varying levels of education expenditures at the two education stages caused by differences in voter turnout seem to be of a greater importance than the characteristics of tertiary education in some cases. In Canada, Norway, and Sweden tertiary education contributes more to the lower intergenerational elasticity than voter turnout, whereas in Denmark, Finland, and the UK the opposite is the case. In Denmark, Norway, and Sweden, where access to college is free and the premium is relatively low, the model exhibits the lowest levels of earnings inequality, as is the case in the data.

1.8 Robustness check

The educated are more likely to participate in a range of political activities such as signing petitions, attending meetings, writing to congress, or contributing to campaigns

(Schlozman et al. 2012). As a robustness check to the voter turnout experiment I instead use variations in party membership. Party members spread political propaganda, mobilize voters, and form and shape agendas.[30] Ichino et al. (2010) find that across countries intergenerational mobility is positively associated with party membership of the poor relative to the rich. I use the first, third, and fifth wave of the World Values Survey to determine the membership rate (active as well as inactive) by age group and level of education to assign these shares as weights, as displayed in Table 1.10 in the Appendix, in the probabilistic voting process. Including all countries, on average weighting voters by party membership propensity explains 3% of the difference to the Gini and 14% of the difference to intergenerational earnings persistence to the US, whereas if we exclude the UK these shares increase to 12% and 22%, respectively. The results are summarized in the Appendix in Table 1.9 and in Figures 1.18, 1.19, and 1.20.

In the US the share of individuals that are active or inactive members of a political party is highest among the countries examined. However, once again participation is biased towards the older and more educated. This causes intergenerational earnings persistence to increase slightly compared to the benchmark model due to public early education decreasing to 4.1% of GDP, which is even closer to the 3.9% in the data. The share of subsidies to college education remain unchanged, while earnings inequality experiences a benign increase.

In Sweden there is hardly a bias in party membership by age, and contrary to other countries, particularly the highly educated are less likely to be party members. This results in high subsidies to early and late education decreasing earnings persistence to 0.35 and the Gini to 0.37, thereby closing 34% and 10% of the gaps to the US, respectively. In Germany party membership is unusual. The relative membership propensity explains 30% of the difference in intergenerational earnings persistence and 17% in earnings inequality.

In the UK the old and educated are relatively more likely to be party members than in the US. This leads to an increase in intergenerational earnings persistence and inequality, thereby increasing the differences to the US. In Canada 11% of the gap to the US is closed, as well as 14% of the difference earnings inequality. In Norway party membership is not habitual, but more common among the older, yet there is no recog-

[30] For evidence on the importance I refer to Huckfeldt and Sprague (1992) and Green and Gerber (2008).

nizable bias towards the more educated. The patterns of party membership in Norway account for 9% of the gap in inequality and 11% of the gap in intergenerational earnings persistence.

1.9 Conclusions

I calibrate a model characterized by dynamic complementarity between early and college education to the US economy. Households vote on tax rates dedicated to the funding of each education level. The model performs well at replicating the US economy across several dimensions including inequality, intergenerational mobility, the share of GDP dedicated to early education, and financial aid to college students. In addition, the model matches details not targeted, such as the intergenerational earnings transition matrix and college attendance by earnings quartile. In the benchmark economy almost one in four individuals is financially constrained in the college decision and mobility is found to be low at the top and the bottom of the earnings distribution.

The negative relation between inequality and intergenerational mobility observed in the data is replicated by the model when simulating with country-specific characteristics of tertiary education. I find that differences in tertiary education can account for 65% of the gaps in earnings inequality and 21% of the gaps in intergenerational earnings persistence between the US and Canada, Denmark, Finland, Norway, Sweden, and the UK. When controlling for country-specific voter turnout by age and level of education in the probabilistic voting process, the negative association between inequality and public expenditures on education observed in cross-country data is reconciled. Political participation in form of voter turnout explains nearly one-quarter of the differences in inequality and intergenerational earnings persistence. As a robustness check, I assign voting weights according to party membership by age and education obtaining similar results.

Concerning voting policies in the US, I find that compulsory voting would reduce earnings persistence by one-tenth, whereas extending electorate franchise to children as of birth and letting their parents vote for them would nearly halve the share of earnings transmitted across generations. However, the effects of these policies on inequality are found to be comparably low.

Simulations with an exogenous increase in the college premium in the US exhibit a

non-linearity in the otherwise negative relation between inequality and mobility, a characteristic not identified in models that ignore the political economy. The incentives for poorer households to enroll their children in college increase, driving broader public support for college subsidies, as stressed by the political science literature. This raises aggregate returns to early education through the dynamic complementarity between the two educational stages, thereby increasing public funding to early education, hence increasing mobility further.

The model neglects pre-primary education and abstracts from savings and borrowing decisions, which play a role in college financing, thereby providing fruitful areas for future research.

1.10 Appendix

Table 1.7: Voting patterns US 1996-2006

	Congress			President		
Age	High school	Some college	College	High school	Some college	College
		1998			**1996**	
18-32	.16	.28	.41	.28	.47	.66
33-48	.32	.49	.63	.45	.65	.79
48-64	.47	.66	.73	.59	.76	.86
		2002			**2000**	
18-32	.15	.26	.44	.29	.47	.67
33-48	.33	.50	.64	.47	.65	.79
48-64	.46	.61	.76	.58	.75	.84
		2006			**2004**	
18-32	.17	.29	.46	.35	.58	.72
33-48	.32	.48	.64	.50	.69	.82
48-64	.47	.63	.77	.60	.78	.85

Datasource: CPS November voting supplement 1996-2006.

Table 1.8: Tertiary education characteristics

	Exogenous		Calibrated					
			Enrollment			Completion		
Country	Premium $\bar{\theta}$	Tuition T	Data	Model	$\underline{\theta}$	Data	Model	ψ_0
US	1.31	.44	.62	.62	.87	.58	.59	.29
Sweden	1.10	0	.76	.74	1.03	.49	.49	.26
UK	1.23	.25	.62	.62	.78	.81	.81	.40
Canada	1.16	.26	.60	.62	.95	.70	.70	.35
Denmark	1.11	0	.65	.66	.98	.82	.82	.44
Norway	1.11	0	.76	.74	1.01	.63	.64	.34
Finland	1.20	0	.68	.66	.88	.72	.72	.36

Datasources: College premium OECD 2013 Table A6.1, Tuition OECD 2012 Table B5.1 (for Canada: Belley et al. 2011), Enrollment OECD 2012 Table C3.3, Completion rates OECD 2011a Table A4.1 (For Canada: Finnie and Qiu 2008).
Note: The relative college premium $c\tilde{p}_j$ of country j is calculated by comparing the college premium cp_j to cp_{US} in the following manner: $c\tilde{p}_j = \frac{cp_j - 1}{cp_{US} - 1}$. Then $\bar{\theta}_j = 1 + c\tilde{p}_j(\bar{\theta}_{US} - 1)$.

Table 1.9: Counterfactuals (party membership)

	Intergenerational elasticity			Gini		
Country	Data	Model	$\Delta_{explained}$	Data	Model	$\Delta_{explained}$
US	.40	.393		.39	.378	
Sweden	.27	.349	.34	.32	.370	.10
UK	.31	.408	-.16	.383	.379	-.31
Canada	.21	.373	.11	.37	.375	.14
Norway	.17	.369	.11	.35	.374	.09
Germany	.32	.370	.30	.37	.374	.17

Note: The share of a gap between a data moment in the US d_{US} and another country $d_{country}$ explained by moments m_j produced by the model, is calculated using the following definition: $\Delta_{explained} = \frac{(m_{US} - m_{country})}{(d_{US} - d_{country})}$.

Table 1.10: Voting weights based on voter turnout and party membership

		Voter turnout			Party membership		
	Age*	High school	Some college	College	High school	Some college	College
US							
	18 - 32	.27	.27	.27	.38	.38	.38
	33 - 48	.32	.48	.64	.30	.48	.61
	48 - 64	.47	.63	.77	.44	.58	.75
Sweden							
	18 - 32	.89	.89	.89	.07	.07	.07
	33 - 48	.93	.91	.94	.11	.13	.11
	49 - 64	.95	.99	.97	.17	.13	.16
UK							
	18 - 32	.51	.51	.51	.09	.09	.09
	33 - 48	.65	.65	.81	.08	.15	.10
	49 - 64	.77	.83	.87	.13	.16	.28
Canada							
	18 - 32	.78	.78	.78	.14	.14	.14
	33 - 48	.80	.95	.90	.10	.25	.19
	49 - 64	.88	.94	.96	.15	.19	.30
Denmark							
	18 - 32	.81	.81	.81			
	33 - 48	.89	.90	.94			
	49 - 64	.93	.99	.95			
Norway							
	18 - 32	.78	.78	.78	.08	.08	.08
	33 - 48	.85	.91	.90	.14	.20	.12
	49 - 64	.86	.92	.93	.22	.28	.26
Finland							
	18 - 32	.69	.69	.69			
	33 - 48	.66	.76	.88			
	49 - 64	.75	.90	.92			
Germany							
	18 - 32	.73	.73	.73	.04	.04	.04
	33 - 48	.68	.83	.92	.06	.04	.10
	49 - 64	.78	.87	.93	.07	.05	.15

*Age at time of elections

Datasource: CPS November voting supplement 2006, European Social Survey 2010, World Values Survey 1981-2007, Canadian Election Study 2010. Missing values are left blank.

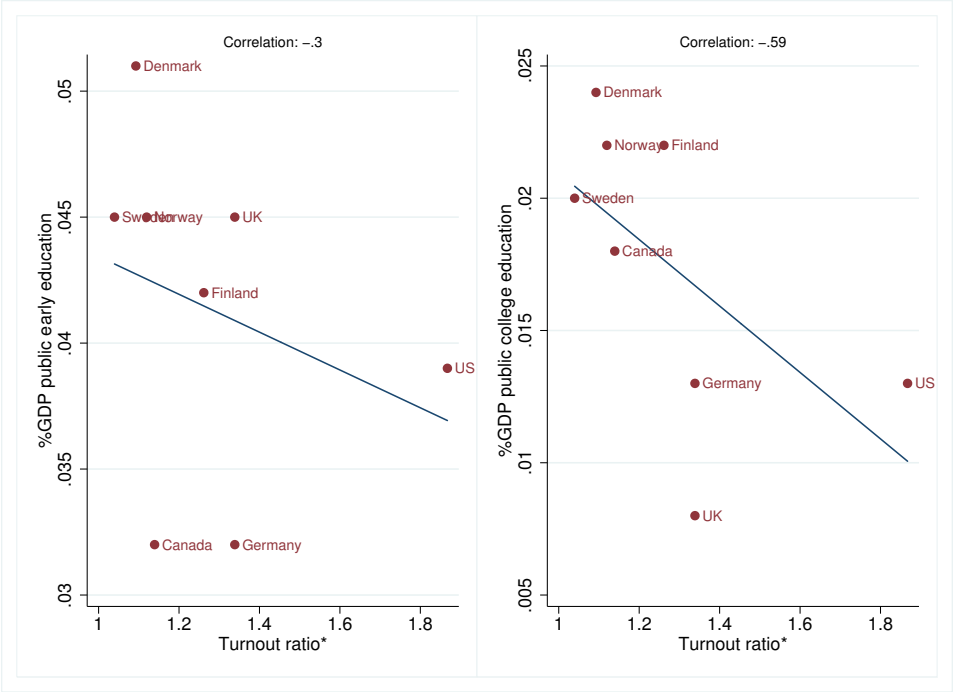


Figure 1.10: Turnout ratio of college graduates to those that did not attend college aged 18-65 versus public expenditures on early (left) and college (right) education

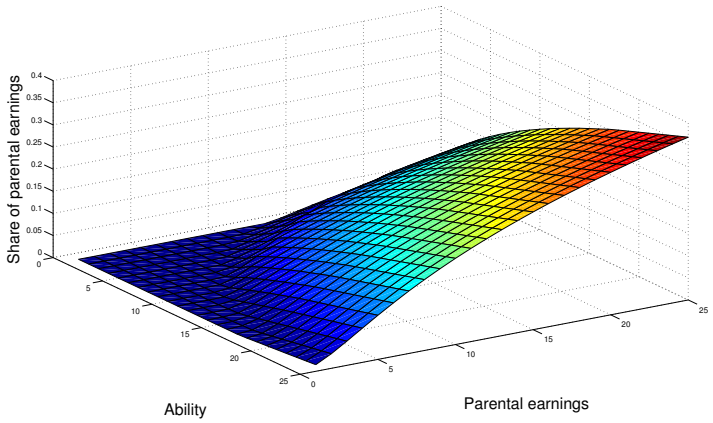


Figure 1.11: The share of parental earnings allocated to private early education as function of innate ability and parental earnings

Figure 1.12: Comparing model moments with country-specific tertiary education and voter turnout by age and education and data moments of the Gini (left) and intergenerational earnings persistence (right)

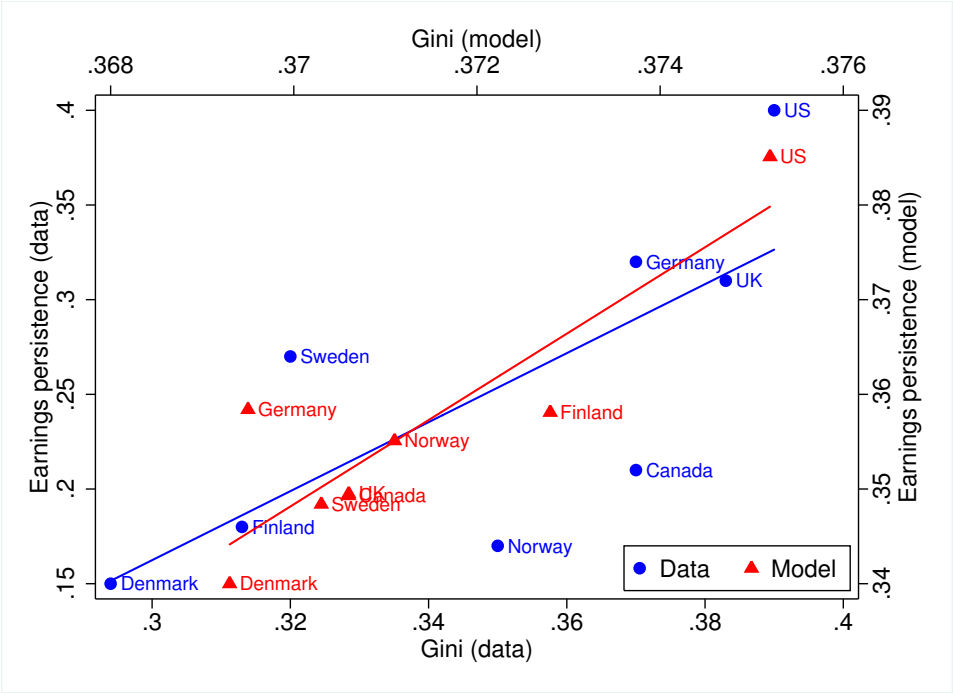


Figure 1.13: Comparing model moments with voter turnout by age and education (red) with data moments (blue) of the Gini versus intergenerational earnings persistence

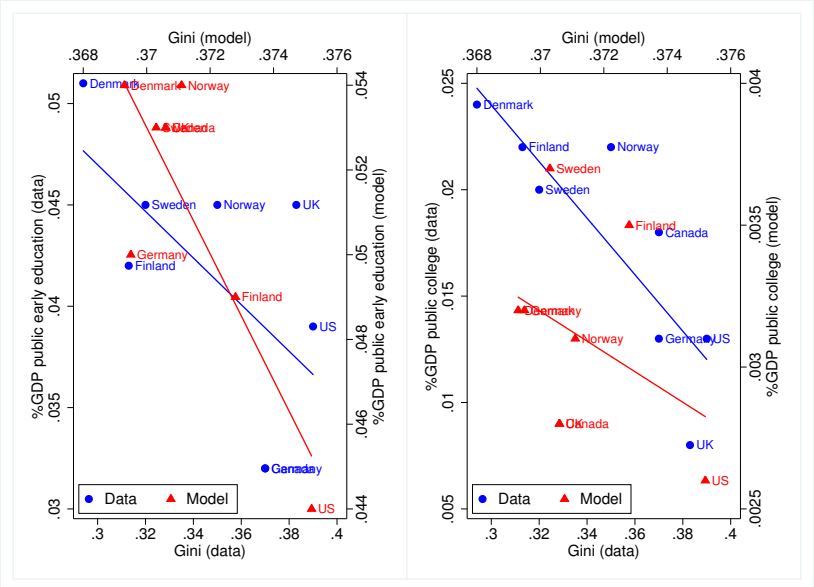


Figure 1.14: Comparing model moments with country-specific voter turnout by age and education (blue) and data moments (red) of the Gini versus public expenditures (% of GDP) on non-tertiary (left) and tertiary (right) education

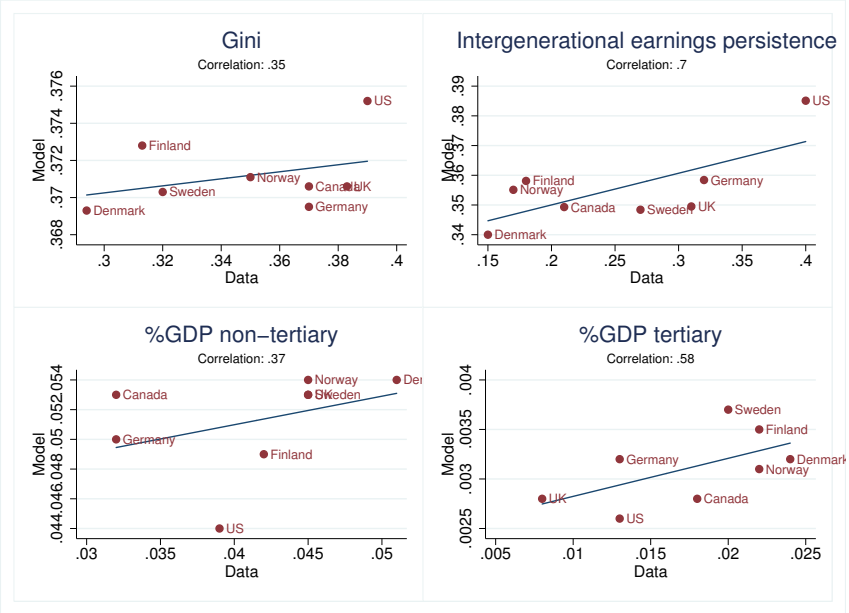


Figure 1.15: Comparing model moments with country-specific voter turnout by age and education and data moments of Gini (top left), intergenerational earnings persistence (top right), and public expenditures (% of GDP) on non-tertiary (bottom left) and tertiary (bottom right) education

Figure 1.16: Comparing model moments with country-specific tertiary education and data moments of the Gini (left) and intergenerational earnings persistence (right)

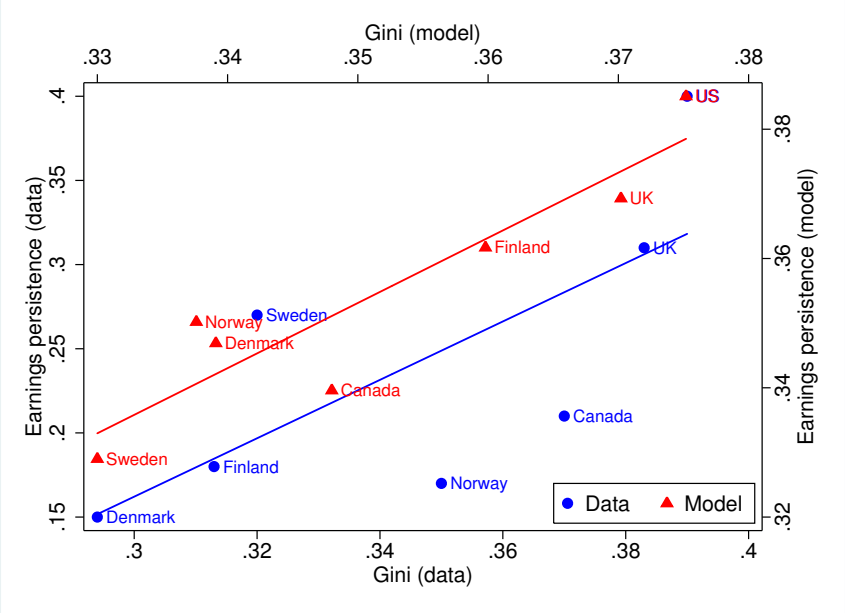


Figure 1.17: Comparing model moments with country-specific tertiary education (red) and data moments (blue) of the Gini versus intergenerational earnings persistence

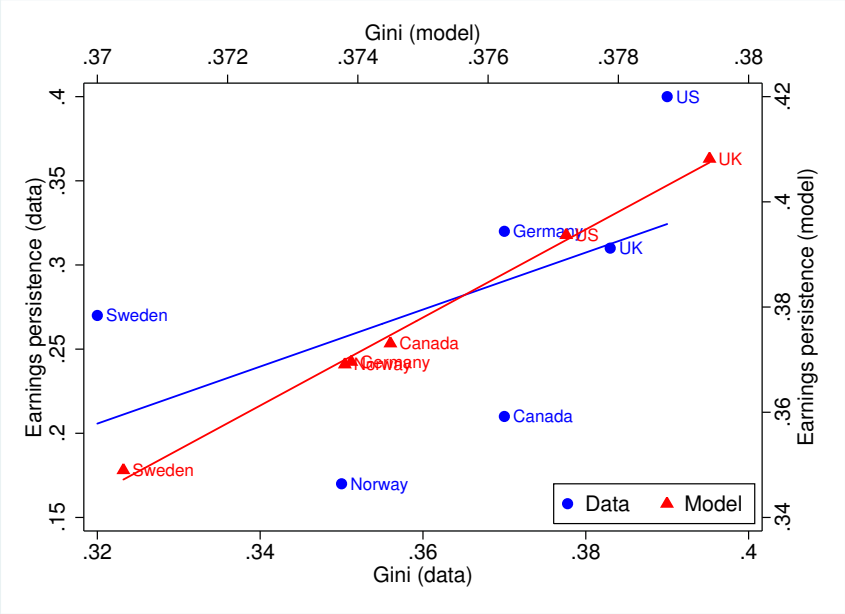


Figure 1.18: Comparing model moments with country-specific party membership (red) and data moments (blue) of the Gini versus intergenerational earnings persistence.

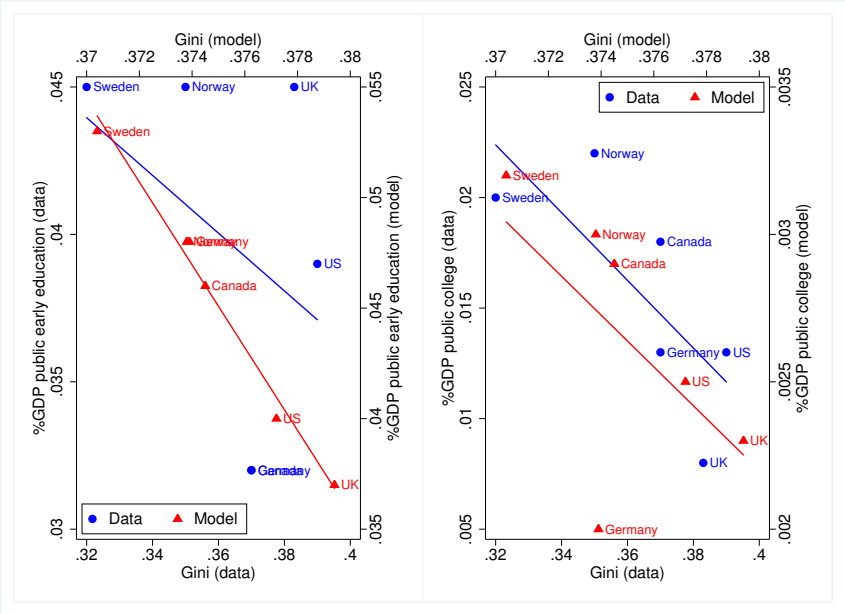


Figure 1.19: Comparing model moments with country-specific party membership by age and education moments (red) with data moments (blue) of the Gini versus public expenditures (% of GDP) on non-tertiary (left) and tertiary (right) education

Figure 1.20: Comparing model moments with country-specific party membership by age and education with data moments of the Gini (top left), intergenerational earnings persistence (top right), and public expenditures (% of GDP) on non-tertiary (bottom left) and tertiary (bottom right) education

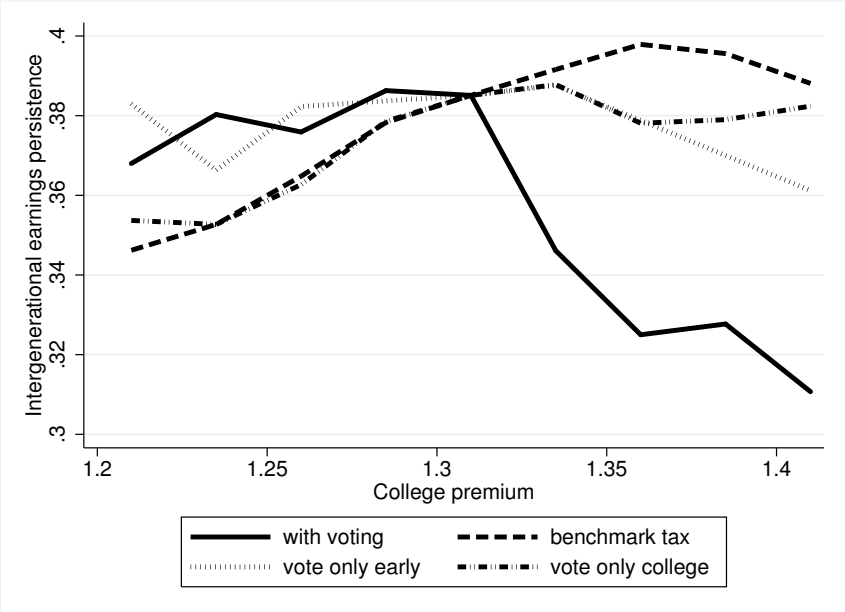


Figure 1.21: The effects of voting on early and/or late education on the intergenerational earnings persistence

References

[1] Aaronson, D., and Mazumder, B. (2008). Intergenerational economic mobility in the United States, 1940 to 2000. *Journal of Human Resources*, 43(1), 139-72.

[2] Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2013). Education Policy and Intergenerational Transfers in Equilibrium (No. w18782). National Bureau of Economic Research.

[3] Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, 40(1), 7-72.

[4] Andrews, D., and Leigh, A. (2009). More inequality, less social mobility. *Applied Economics Letters*, 16(15), 1489-92.

[5] Ansell, B. W. (2010). From the ballot to the blackboard: The redistributive political economy of education. Cambridge University Press.

[6] Bailey, M. J., and Dynarski, S. M. (2011). Gains and gaps: Changing inequality in US college entry and completion (No. w17633). National Bureau of Economic Research.

[7] Becker, G. S., and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, 1153-89.

[8] Becker, G. S., and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3 Pt. 2), 1.

[9] Belley P, and Lochner L. 2007. The changing role of family income and ability in determining educational achievement. *Journal of Human Capital*, 1:37-89

[10] Belley, P., Frenette, M., and Lochner, L. (2011). Post-secondary attendance by parental income in the US and Canada: What role for financial aid policy? (No. w17218). National Bureau of Economic Research.

[11] Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4), 352-65.

[12] Bénabou, R. (1996). Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance. *American Economic Review*, 584-609.

[13] Bénabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review*, 96-129.

[14] Black, S. E., and Devereux, P. J. (2011). Recent developments in intergenerational mobility. Handbook of labor economics, 4, 1487-541.

[15] Blais, A. (2000). To vote or not to vote?: The merits and limits of rational choice theory. University of Pittsburgh Press.

[16] Blanden, J., Gregg, P., and Macmillan, L. (2007). Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. *Economic Journal*, 117(519), C43-C60.

[17] Blankenau, W., and Youderian, X. (2012). Early childhood education expenditures and the intergenerational persistence of income. Working Paper.

[18] Bloome, D. (2013). Income Inequality and Intergenerational Income Mobility in the United States. mimeo.

[19] Björklund, A., and Jäntti, M. (1997). Intergenerational income mobility in Sweden compared to the United States. *American Economic Review*, 87(5), 1009-18.

[20] Björklund, A., Eriksson, T., Jäntti, M., Rauum, O., and Österbacka, E. (2004). Family Structure and labour market success: the influence of siblings and birth order on the earnings of young adults in Norway, Finland and Sweden. In: Miles Corak (ed.) Generational income mobility in North America and Europe, 207-25, Cambridge University Press.

[21] Bohacek, R., and Kapicka, M. (2012). A Quantitative Analysis of Educational Reforms in a Dynastic Framework, Manuscript, University of California Santa Barbara.

[22] Bonica, A., McCarty, N., Poole, K. T., and Rosenthal, H. (2013). Why Hasn't Democracy Slowed Rising Inequality?. *Journal of Economic Perspectives*, 27(3), 103-24.

[23] Brown, M., Scholz, J. K., and Seshadri, A. (2012). A new test of borrowing constraints for education. *Review of Economic Studies*, 79(2), 511-38.

[24] Busemeyer, M. R. (2012). Inequality and the political economy of education: An analysis of individual preferences in OECD countries. Journal of European Social Policy, 22(3), 219-40.

[25] Card, D., and Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1), 1-40.

[26] Card, D., and Krueger, A. B. (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina. *Journal of Economic Perspectives*, 10(4), 31-50.

[27] Caucutt, E. M., and Lochner, L. (2012). Early and late human capital investments, borrowing constraints, and the family (No. w18493). National Bureau of Economic Research.

[28] Chatterjee, S., and Ionescu, F. A. (2010). Insuring college failure risk. Working Paper No. 10-1 Federal Reserve Bank of Philadelphia.

[29] Chetty, R., Hendren, N., Kline, P., and Saez, E. (2013). The Economic Impacts of Tax Expenditures: Evidence from Spatial Variation across the U.S.. Working paper.

[30] Chong, A., and Olivera, M. (2008). Does compulsory voting help equalize incomes?. *Economics Politics*, 20(3), 391-415.

[31] Cooper, S. J. (1998). Redistribution and the persistence of income inequality. Research Programs, John F. Kennedy School of Government, Harvard University.

[32] Corak, M. (2006). Do poor children become poor adults? Lessons from a cross-country comparison of generational earnings mobility (Vol. 13, 143-88). Emerald Group Publishing Limited.

[33] Corak, M. (2012). Inequality from generation to generation: the United States in comparison. The Economics of Inequality, Poverty, and Discrimination in the 21st Century, ABC-CLIO, forthcoming.

[34] Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives* , 27(3), 79-102.

[35] Corak, M., and Piraino, P. (2011). The intergenerational transmission of employers. *Journal of Labor Economics*, 29(1), 37-68.

[36] Cunha, F., and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-82.

[37] Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. Handbook of the Economics of Education, 1, 697-812.

[38] De la Croix, D., and Doepke, M. (2004). Public versus private education when differential fertility matters. *Journal of Development Economics*, 73(2), 607-29.

[39] Demeny, P. (1986). Pronatalist policies in low-fertility countries: Patterns, performance, and prospects. *Population and Development Review*, 12, 335-58.

[40] Domeij, D., and Floden, M. (2010). Inequality trends in Sweden 1978-2004. *Review of Economic Dynamics*, 13(1), 179-208.

[41] Duncan, G. J., Kalil, A., and Ziol-Guest, K. M. (2013). Increasing Inequality in Parent Incomes and Children's Completed Schooling: Correlation or Causation? Working paper.

[42] Duncan, G. J., and Murnane, R. J. (Eds.). (2011). Whither opportunity?: Rising inequality, schools, and children's life chances. Russell Sage Foundation.

[43] Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth*, 1(1), 75-93.

[44] Epple, D., and Romano, R. E. (1996). Ends against the middle: Determining public service provision when there are private alternatives. *Journal of Public Economics*, 62(3), 297-325.

[45] ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3.0. Norwegian Social Science Data Services, Norway - Data Archive and distributor of ESS data.

[46] Fernandez, R., and Rogerson, R. (1995). On the political economy of education subsidies. *Review of Economic Studies*, 62(2), 249-62.

[47] Fernandez, R., and Rogerson, R. (1997). Keeping people out: income distribution, zoning, and the quality of public education. *International Economic Review*, 23-42.

[48] Fernandez, R., and Rogerson, R. (1998). Public education and income distribution: A dynamic quantitative evaluation of education-finance reform. *American Economic Review*, 813-33.

[49] Finnie, R., and Qiu, T. (2008). The Patterns of Persistence in Post-Secondary Education in Canada: Evidence from the YITS-B Dataset. Educational Policy Institute.

[50] Fournier, P., Cutler, F., Soroka, S., and Stolle, D. (2011). The 2011 Canadian Election Study. [dataset]

[51] Fowler, A. (2013). Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia. *Quarterly Journal of Political Science*, 8(2), 159-82.

[52] Fuchs-Schuendeln, N., Krueger, D., and Sommer, M. (2010). Inequality trends for Germany in the last two decades: A tale of two countries. *Review of Economic Dynamics*, 13(1), 103-32.

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

[54] Gilens, M. (2012). Affluence and influence: Economic inequality and political power in America. Princeton University Press.

[55] Glomm, G., and Ravikumar, B. (1992). Public versus private investment in human capital: endogenous growth and income inequality. *Journal of Political Economy*, 818-34.

[56] Green, D. P., and Gerber, A. S. (2008). Get out the vote: How to increase voter turnout. Brookings Institution Press.

[57] Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006. *Review of Economic Dynamics*, 13(1), 15-51.

[58] Heckman, J. J. (2010). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48(2), 356-98.

[59] Heckman, J. J., and LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *Review of Economics and Statistics*, 92(2), 244-62.

[60] Herrington, C. M. (2014). Public Education Financing Systems, Earnings Inequality, and Intergenerational Mobility. Mimeo.

[61] Hornstein, A., Krusell, P., and Violante, G. L. (2005). The effects of technical change on labor market inequalities. Handbook of economic growth, 1, 1275-370.

[62] Holter, H. (2014). Accounting for cross-country differences in intergenerational earnings persistence: the impact of taxation and public education expenditure. Available at SSRN 1978261.

[63] Hoxby, C. M. (1998). How much does school spending depend on family income? The historical origins of the current school finance dilemma. *American Economic Review*, 88(2), 309-14.

[64] Huckfeldt, R., and Sprague, J. (1992). Political parties and electoral mobilization: Political structure, social structure, and the party canvass. *American Political Science Review*, 70-86.

[65] Huggett, M., Ventura, G., and Yaron, A. (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics*, 53(2), 265-90.

[66] Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of Lifetime Inequality. *American Economic Review*, 101(7), 2923-54.

[67] Ichino, A., Karabarbounis, L., and Moretti, E. (2010). The political economy of intergenerational income mobility. *Economic Inquiry*, 49(1), 47-69.

[68] Jackman, R. W. (1987). Political institutions and voter turnout in the industrial democracies. *American Political Science Review*, 405-23.

[69] Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Österbacka, E., and Eriksson, T. (2006). American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States. Working paper.

[70] Johnson, M. T. (2012). Borrowing Constraints, College Enrollment, and Delayed Entry (No. 2011-006). Human Capital and Economic Opportunity Working Group.

[71] Karabarbounis, L. (2011). One Dollar, One Vote. The Economic Journal, 121(553), 621-51.

[72] Katz, L. F., and Murphy, K. M. (1992). Changes in relative wages, 1963-1987: supply and demand factors. *Quarterly Journal of Economics*, 107(1), 35-78.

[73] Keane, M. P., and Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4), 1051-103.

[74] Kornrich, S., and Furstenberg, F. (2013). Investing in Children: Changes in Parental Spending on Children, 1972-2007. *Demography*, 50(1), 1-23.

[75] Krusell, P., Ohanian, L. E., Rios-Rull, J. V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5), 1029-53.

[76] Lijphart, A. (1997). Unequal participation: democracy's unresolved dilemma. *American Political Science Review*, 1-14.

[77] Light, A., and Strayer, W. (2000). Determinants of college completion: school quality or student ability? *Journal of Human Resources*, 299-332.

[78] Lindbeck, A., and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice*, 52(3), 273-97.

[79] Lochner, L., and Monge-Naranjo, A. (2012). Credit Constraints in Education. *Annual Review of Economics*, 4(1), 225-56.

[80] Loury, G. C. (1981). Intergenerational transfers and the distribution of earnings. *Econometrica*, 843-67.

[81] Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of Population Economics*, 21(3), 685-701.

[82] Niemi, R. G., and Jennings, M. K. (1991). Issues and inheritance in the formation of party identification. *American Journal of Political Science*, 970-88.

[83] Organisation for Economic Co-operation and Development (OECD) (2011a). Education at a Glance. Paris.

[84] Organisation for Economic Co-operation and Development (OECD) (2011b). Divided We Stand: Why Inequality Keeps Rising. Paris.

[85] Organisation for Economic Co-operation and Development (OECD) (2012). Education at a Glance. Paris.

[86] Organisation for Economic Co-operation and Development (OECD) (2013). Education at a Glance. Paris.

[87] Perea, E. A. (2002). Individual characteristics, institutional incentives and electoral abstention in Western Europe. *European Journal of Political Research*, 41(5), 643-73.

[88] Persson, T., and Tabellini, G. E. (2000). Political economics: explaining economic policy. The MIT press.

[89] Powell Jr, G. B. (1986). American voter turnout in comparative perspective. *American Political Science Review*, 17-43.

[90] Restuccia, D., and Urrutia, C. (2004). Intergenerational persistence of earnings: The role of early and college education. *American Economic Review*, 94(5), 1354-78.

[91] Schlozman, K. L., Verba, S., and Brady, H. E. (2012). The unheavenly chorus: Unequal political voice and the broken promise of American democracy. Princeton University Press.

[92] Solon, G. (1992). Intergenerational income mobility in the United States. *American Economic Review*, 393-408.

[93] Solon, G. (1999). Intergenerational mobility in the labor market. Handbook of labor economics, 3, 1761-800.

[94] Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *Journal of Economic Perspectives*, 16(3), 59-66.

[95] Solon, G. (2004). A model of intergenerational mobility variation over time and place. In: Miles Corak (ed.) Generational income mobility in North America and Europe, 38-47. Cambridge University Press.

[96] Stinebrickner, T. R., and Stinebrickner, R. (2007). The effect of credit constraints on the college drop-out decision a direct approach using a new panel study (No. w13340). National Bureau of Economic Research.

[97] Winter, C. (2013). Accounting for the changing role of family income in determining college entry. Mimeo.

[98] Wolf, F. (2009). The division of labour in education funding: A cross-national comparison of public and private education expenditure in 28 OECD countries. *The Acta Politica*, 44(1), 50-73.

[99] Wolfinger, R. E., and Rosenstone, S. J. (1980). Who votes. Yale University Press.

[100] World Values Survey 2005 Official data file v.20090901, 2009. World Values Survey Association (www.worldvaluessurvey.org).

Chapter 2

Early Childhood Education and Intergenerational Earnings Mobility - Is Mobility Higher when Mothers Work?

2.1 Introduction

Why does intergenerational earnings mobility vary to such a large extent across countries? Despite a large body of empirical literature insisting on the importance of preschool education in the development of cognitive and non-cognitive skills, which successfully explain labor earnings, little attention has been given to whether differences in public child-care and preschool expenditures can account for differences in intergenerational earnings mobility across countries. So far the literature has focused on differences in public expenditures at the level of primary, secondary, and/or tertiary education, which have been shown to be able to account for a considerable share of cross-country differences in intergenerational mobility (Herrington 2014; Holter 2014; Rauh 2014).

The early childhood environment does not only depend on preschool but also on the home environment. Children of more educated parents face a more stimulating home environment. So the more time a child spends at home with his parents, the more the offspring adopts parental skills. In Figure 2.1 one can see that in countries where fe-

males are more likely to work, the childrens' future earnings depend less on parental earnings.[1]

I build a model where differences in female labor force participation as well as

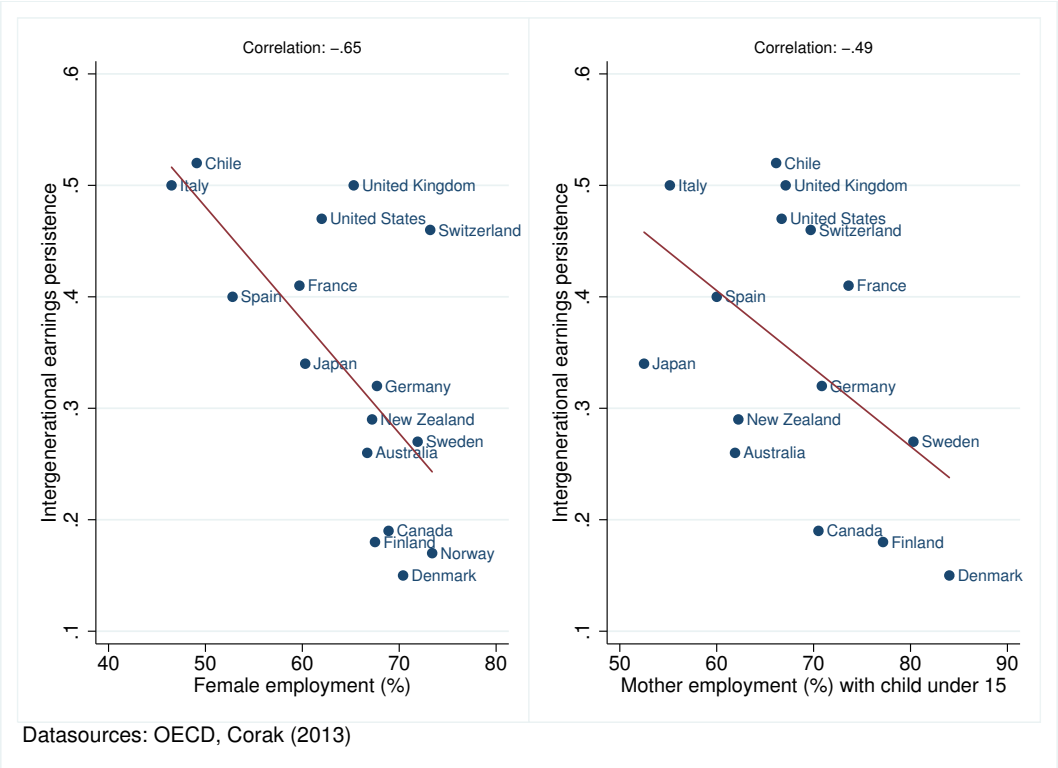


Figure 2.1: Female employment and intergenerational earnings persistence

differing levels of public preschool financing are exploited to explain cross-country differences in intergenerational mobility. When young mothers are less likely to participate in the labor force, parents might not have a large demand for public preschool expenditures, as maternal nurturing could be considered a substitute to public preschool. This increases the private investments of households where both parents are working. Hence, when public preschool expenditures are low, the quality of care a child receives is dependent on parental financial resources. When the mother is not working she spends more time with her child. During the time a mother spends with her child the maternal cognitive abilities affect the development of a child. Therefore, when more children

[1]This relationship holds when using the difference as well as the ratio between female and male employment rates.

face maternal nurturing instead of public provision of preschool education, cognitive abilities are transmitted with a higher correlation.

I calibrate the dynastic overlapping generations model to the US as a benchmark, matching key features in terms of wage dispersion and distribution, earnings mobility, and private education expenditures. As a counterfactual exercise, I simulate the economy by replacing the public preschool expenditures of the US with those of 15 OECD countries. On average this exercise explains 17% of the difference in intergenerational earnings persistence between the US and these countries and 22% of the variation in earnings persistence.

In the model human capital creation takes place in three stages. First, children are born with an innate ability, which is positively correlated to parental innate ability, but transmitted through a noisy process. Innate ability affects the learning capacity throughout education and the efficiency of how investments are transformed into human capital. The initial human capital endowment is enhanced through public and private education investments as well as maternal time investment. The efficiency of the maternal time investment depends on her human capital. Second, children attend primary and secondary education, commonly referred to as K-12 education. Third, once leaving home individuals can choose to attend college, where one might drop out, hence making it a risky investment. At the second and third educational stage, I only consider private and public expenditures, whereas at the earliest stage I include the maternal time input, of which the quality depends on maternal human capital. Recent structural estimates confirm that the parental time input seems to be of decreasing importance as the child ages (Abbott 2014). Later parents can take direct influence on children's development by investing privately and by endowing the children with a intervivos transfer when they leave home and make the college decision.

Across OECD countries one observes a hump-shaped relationship between public preschool expenditures and intergenerational earnings persistence, as can be seen in Figure 2.2. Countries that spend little on public preschool have low earnings persistence, those with expenditures in the medium range have relatively high earnings persistence, and those spending a lot have low earnings persistence. The simulations exploiting cross-country differences in public preschool provision replicate this hump-shaped relationship. When public preschool expenditures are very high and crowd out private investments, all children leave preschool with a considerable level of human capital,

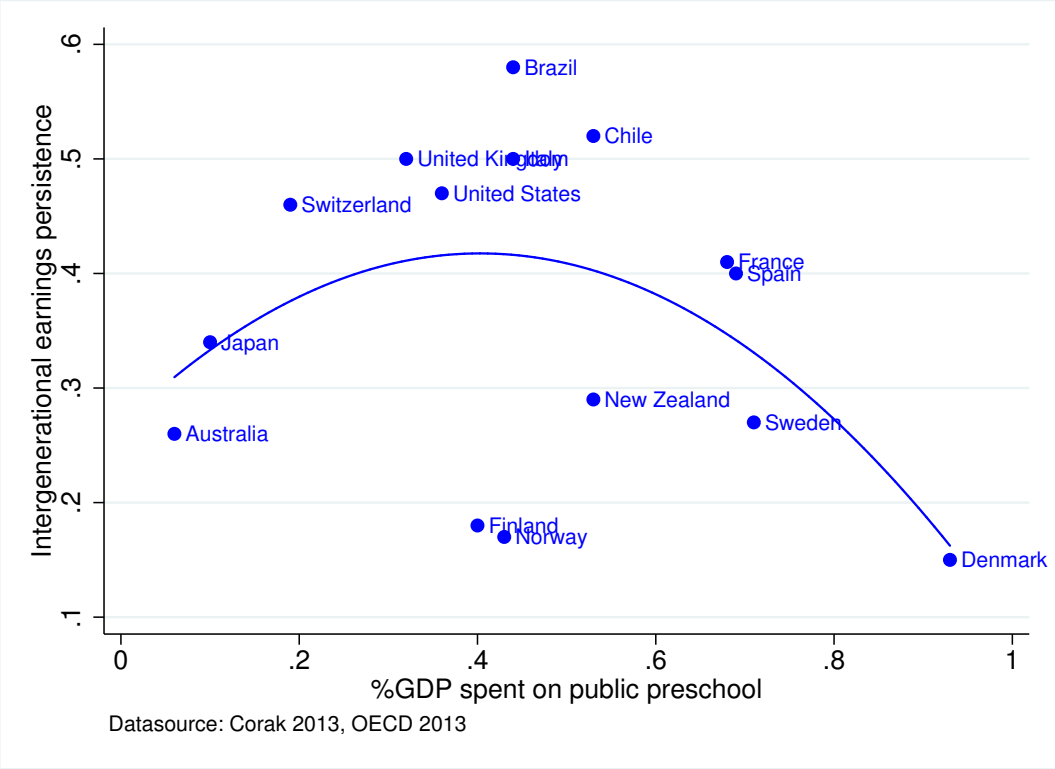


Figure 2.2: Female employment and intergenerational earnings persistence

such that future earnings depend less on parental investments at the subsequent educational stages. At a medium level, parental preschool investments play an important role. The differences in early human capital are amplified due to the dynamic complementarity of human capital at the subsequent stages. High levels of human capital increase the returns to investment, wherefore richer parents invest relatively even more, which leads to high levels of earnings persistence. However, very low public preschool provision leads to overall lower levels of human capital. Therefore, at later stages the differences of human capital acquisition of poor and rich children is dampened, because the returns to investment, and consequentially the incentives to invest are lower.

A lower female labor force participation in the model, generated by an increase in disutility associated with joint work of husband and wife, results in an increase in earnings persistence. The intuition is, when more mothers stay at home with their children this provides a direct channel of ability transmission. Conversely, when the disutility is reduced, mothers work more and earnings persistence decreases.

The body of literature providing evidence on the importance of child care and preschool on children's cognitive abilities is extensive. However, a number of studies detect negative outcomes of children receiving child care (Blau 1999; Bernal 2008; Bernal and Keane 2010; Herbst and Tekin 2010, 2012; Hawkinson et al. 2013). Most of these studies, though, disregard the quality of child care provided, which has been shown to be of crucial importance (Love el al. 2003). In some cases the negative results are admittedly driven by the use of informal child care, whereas for formal child care no negative effect can be made out.[2] In a survey by Brilli et al. (2013), they find that large-scale policies of the provision of non-parental child care have positive effects on children's cognitive outcomes, both in the short and in the medium run, and that the provision of early childhood education can have long lasting effects on adult outcomes, also enhancing non-cognitive skills. Reviewing the returns to the Perry Preschool Program, Heckman et al. (2010) report substantial benefits to the participating individuals and society as a whole.

Witte and Trowbridge (2004) summarize the early care and education offered in the US and compare it to those encountered in Europe. They find that public expenditures are lower in the US and are targeted towards disadvantaged families, whereas in Europe they are universal and staff credentials are uniformly higher. In the US there are three major programs, namely, Head Stat, Pre-Kindergarten, and a child care voucher program, which have different sources of funding, administration, and minimum standard regulations, thereby creating transaction costs for parents trying to inform themselves and coordinate stable arrangements for their children. Only 50% of eligible low income families wanting their children to participate actually receive services from those three programs. Additionally, these programs mostly provide only part-day and part-year care, whereas in Europe most offer full-day and full-year care.

Del Boca et al. (2012) find that mothers' work has a negative effect on children's academic results due to a reduction in time dedicated to their children. This impact,

[2] McCartney (1984), Melhuish et al. (1992), and NICHD (2000) find that one aspect that differentiates formal from informal child care is the amount of language stimulation. Teachers of formal child care are more likely to have training in child development and are more educated in general, both of which are associated with more verbal stimulation. Hart and Risley (1995) and Rowe (2008) document that children of poor, uneducated parents are exposed to a much poorer language environment in terms of amount of words heard. Suskind et al. (2013) make parents aware of the importance of the language environment and of ways how to improve it, finding that this increases the children's vocabulary substantially. According to NICHD (2000), formal child care also tends to provide more supportive, attentive, and interactive care.

however, is offset by the use of formal child care. The positive effects of formal child care are stronger for children from households with a low socio-economic status. In a structural estimation Del Boca, Flinn, and Wiswall (2013) identify conditional cash transfers as the most efficient way to improve average child quality. In contrast, Griffin (2011) finds that an expansion of Head Start to families that are now not eligible, would be more efficient than cash transfers, and would be beneficial because now children from high quality home environments are spending significant amounts of time in low quality child care. Also looking at Head Start, Ludwig and Miller (2007) and Deming (2009) identify positive effects on educational attainment. Havnes and Mogstad (2010) emphasize that looking only at mean impacts misses a lot. Studying a publicly provided child care in Norway they detect effects to be positive over most of the distribution, but particularly large below the median. Again analyzing the large-scale expansion of subsidized child care in Norway, Havnes and Mogstad (2011) provide evidence for strong and persistent positive effects on children's' educational attainment, labor market participation, and decreased welfare dependency.

The macroeconomic literature generally focuses on the responsiveness of female labor supply to child-care subsidies, for which most identify positive responses (Attanasio, Low, and Virginia Sanchez-Marcos 2008; Bick 2013; Domeij and Klein 2012; Guner, Kaygusuz, and Ventura 2012). Few macroeconomic models explicitly include early childhood education and child care to explain human capital outcomes. Youderian (2013) includes parental time investment in early education as an education impact. Restricting her analysis to the US, she finds that subsidized private education spending and adopting paid parental leave would increase human capital, while labor supply decreases when introducing paid parental leave.

The differences in female participation in the labor force could be driven by various forces, which are not necessarily competing arguments. One driver could be cultural forces, due to which the male is seen as the breadwinner of the family, wherefore women are more likely to give up/not take up a job in order to care for their children. In some countries there also seems to be a social stigma attached to mothers of young children that work, as people assume this might negatively affect the child's development. There exists a well established empirical literature which studies how culture affects decisions of females. El-Attar (2013) shows that the level of trust females have, influences whether they are willing to send their child to formal child care. Fernandez

and Fogli (2005) provide empirical evidence that cultural differences have a significant effect in explaining fertility and female labor force participation. Another force could be discriminatory actions in the hiring process, where females are avoided due to the possibility of them requiring a break induced by childbearing. Institutions could play a role as well. For instance, paternal work leave facilitates labor market participation of females and has the potential to reduce a discriminatory wedge in the hiring process.[3]

Del Boca (2002) and Baker, Gruber, and Milligan (2008) show empirically that the availability of child care increase the probability of mothers working, while Tekin (2005) finds that single mothers are highly responsive to child-care provision by increasing their employment and replacing parental and relative care with formal child care. Blau and Tekin (2007) also estimate a positive effect of child-care subsidy receipt on the employment of mothers. However, Havnes and Mogstad (2011) find no causal effect of child-care expenditures on maternal employment.

The rest of the paper is organised as follows. In Section 2.2 stylized facts are presented, In Section 3.2 the model is specified, which is paramterized in Section 2.5. Counterfactual experiments are conducted and discussed in Section 2.6. Finally, Section 3.5 concludes.

2.2 Stylized facts

In Figure 2.3 the strong positive correlation between female labor supply and public child-care expenditures is illustrated. If public facilities providing child care are unavailable then households have to invest privately in case both parents work. When preschool quality depends on private investments then children from poorer households face a poorer environment (informal child care), whereas richer households can afford high quality private facilities. In Figure 2.4 it becomes clear that in most countries with low public child-care and preschool expenditures, the demand for informal child care is high. Informal child care is likely to be provided by individuals that are not professionally trained to stimulate cognitive and non-cognitive skills of children.

On the one hand, in Figure 2.5 we see that in countries with a large sector of in-

[3] In Figure 2.9 in the Appendix one can see that paternal weeks of work leave after childbirth are found to be positively correlated with employment rates of mothers as well as intergenerational earnings mobility.

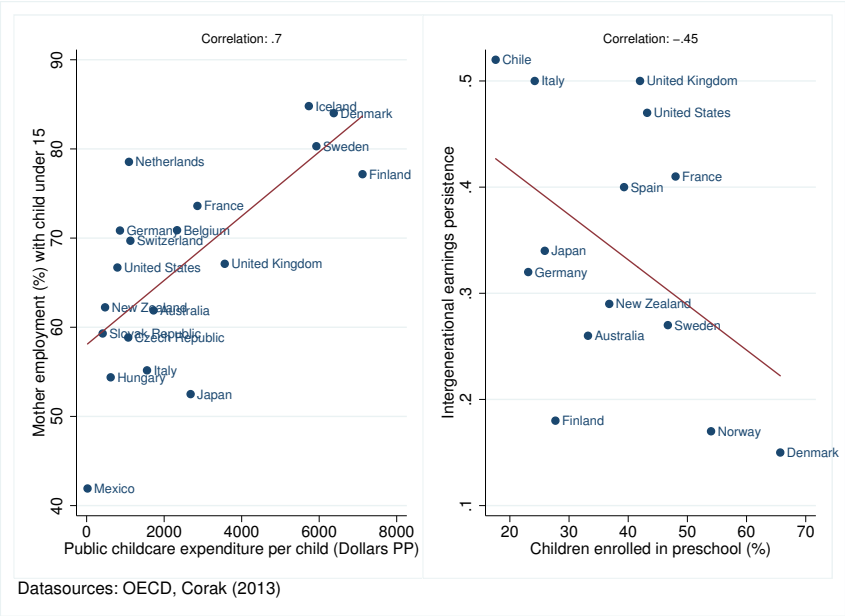


Figure 2.3: Child care expenditures, preschool, and intergenerational earnings persistence

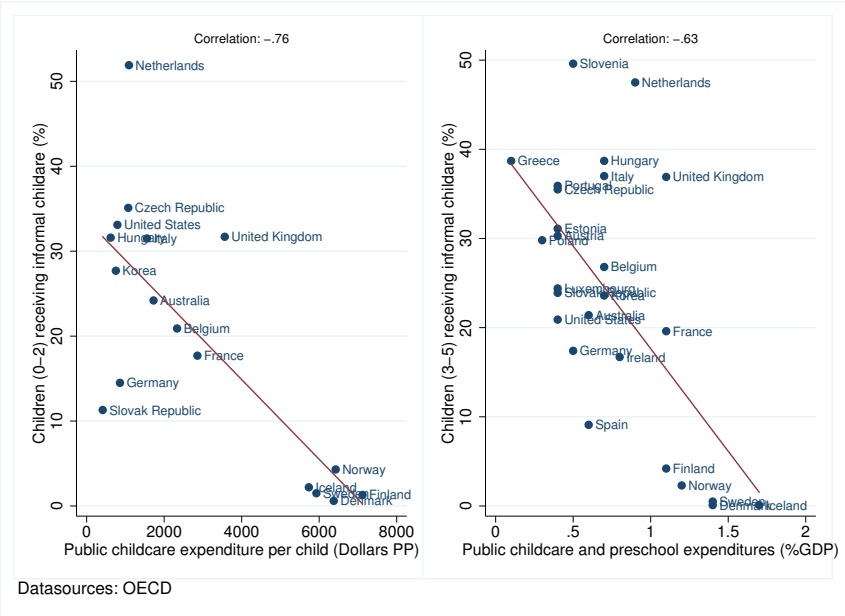


Figure 2.4: Preschool expenditures and informal child care

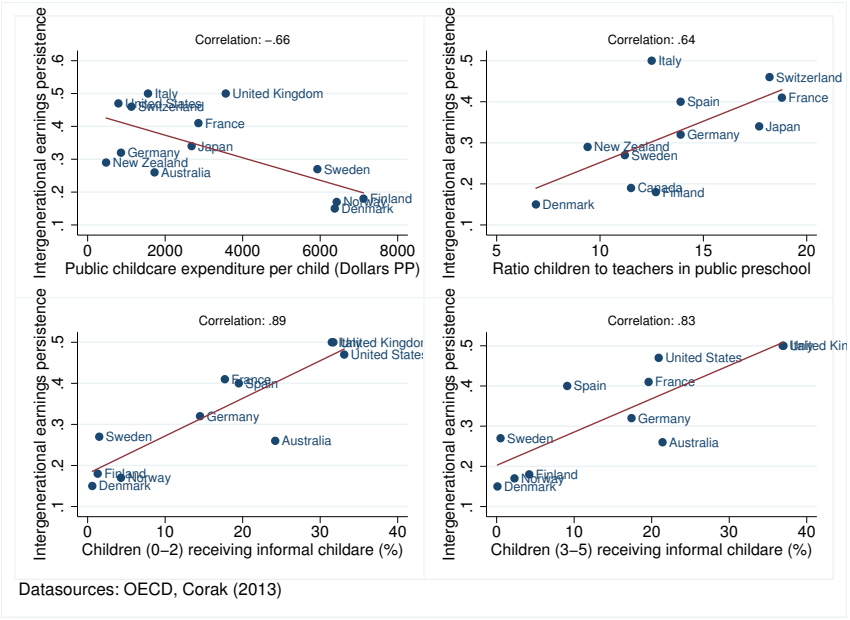


Figure 2.5: Preschool quality and intergenerational earnings persistence

formal child care the intergenerational earnings persistence tends to be higher. On the other hand, in Figure 2.3 it seems that in countries where more children enroll in public preschool, the children's future earnings exhibit a lower correlation with parental earnings. This seems to be especially the case in countries where child-care expenditures are higher and where the ratio of children to teachers in preschool is lower (Figure 2.5). This indicates that when the quality of preschool is high and mothers are more likely work, then intergenerational earnings persistence is low. However, when there is little supply of public preschool and child care, and females are unlikely to work, then earnings persistence is high. The mechanism could be twofold. First, if fewer females are working, then more mothers are at home with their children providing a direct channel of transmission of abilities. Second, if both parents actually do work, they will have to either invest in private preschool, which generally is costly and of higher quality, or, if the household cannot afford to do so, arrange for informal child care, which is likely to be less costly and of lower quality.

In Figure 2.6 (and Figure 2.10 in the Appendix) one can see that cultural forces might play an important role in determining whether a mother works or not. In Figure 2.6 the negative associations between the share of the population in a country believ-

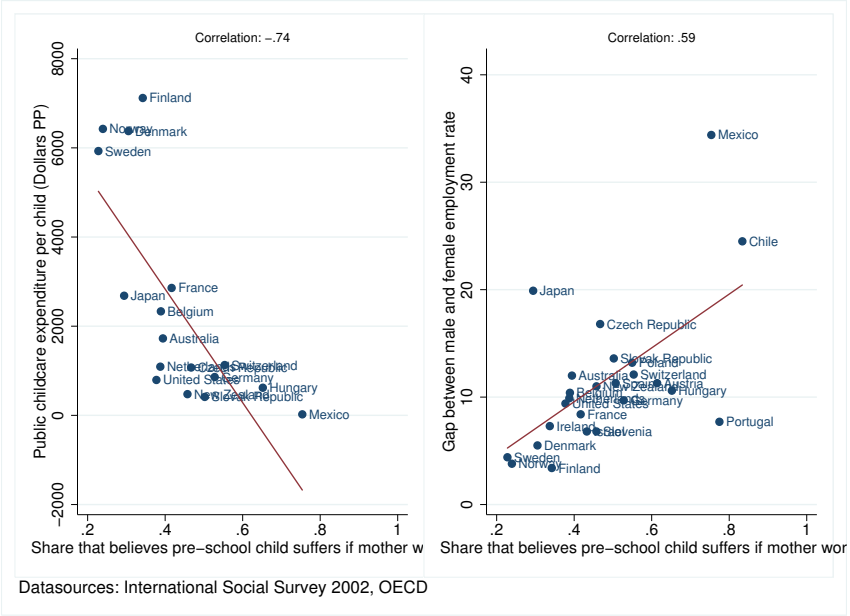


Figure 2.6: Attitudes to mothers working, preschool expenditures, and female employment

ing that a preschool-aged child suffers if a mother works and the employment rate of mothers, as well as public child-care expenditures are exhibited. In Figure 2.10 in the Appendix one observes similar correlations, just that instead the survey question asks whether the family suffers when a mother works.

2.3 Model

In the model time is discrete and an individual can live up to J periods. All individuals run through at least three educational periods composed of one period of child care/preschool, followed by two periods of primary and secondary schooling. After three periods with their parents, individuals leave home and might receive an intervivos transfer from their parents. In the same period they decide whether to attend college or not. The period in which an individual starts making own decisions when leaving home is initialized as $j = 1$, which occurs in the fourth period of an individual's life. The rest of life is dedicated to work in order to finance consumption, children's education, and saving for retirement, which occurs as of $j \geq 9$. Upon retirement individuals receive a

pension but might die with an exogenous probability, whereas death takes place with certainty at the latest at the end of period J.

Throughout working life men provide labor elastically, whereas mothers of preschool-aged children can decide to work full time, part time, or not at all in order to stay home with their children. In all other periods mothers also provide labor inelastically.

Labor income can either be consumed, invested in private education of the children, saved, or passed on to the children as an intervivos transfer when they leave home. Parents invest in their children because they are altruistic for their children and the discounted future utility flow of their offspring.

The model abstracts from population growth so that the size of the population is constant. Each household is of mass unity and composed of two adults (one male and and one female), and additionally during periods $j = 3,4,5$ two simultaneously born children of mass unity are present.

The main sources of heterogeneity in the model are that households differ in terms of their innate ability, accumulated human capital, and assets. The distribution of households' state variables x_j (ability, human capital, assets, etc.) is summarized by $\mu_j(x_j)$, where subscript j denotes the period of the life cycle.

2.3.1 Human capital creation

A child is born with innate ability a, which is imperfectly correlated to the parental innate ability. This can be interpreted as the genetic transmission of learning ability. Innate ability remains constant over the entire life and is transmitted according to

$$ln(a') = \rho ln(a) + \varepsilon \; where \; \varepsilon \sim N(0, \sigma_a^2), \tag{2.1}$$

where a is parental innate ability and a' denotes the offsprings ability. The parameter ρ is responsible for the correlation in ability between generations, whereas ε is the noise in the ability transmission.

2.3.1.1 Preschool

The preschool education for children has three components: (i) public provision g_e, (ii) private expenditures i_e, and (iii) through the time mothers spent with their children.

Spending small amounts privately can be interpreted as the before mentioned informal child care one observes in many countries that lack public provision. Informal child care is typically provided by a nanny or a grandparent. Large spendings on private preschool are to be understood as high quality facilities with well trained educators.

The labor supply decision of the mother is given by $n \in \{0, \frac{1}{2}, 1\}$. The time a mother spends with her children depends on whether she works or not. If she works full time ($n = 1$) she provides a fraction λ less of time compared to a mother that does not work. The quality of the time a mother spends with her child is influenced by her own human capital h and the efficiency parameter ν. These three time-quality inputs are assumed to be substitutes and their efficiency is governed by the productivity parameter γ_e. Therefore, the following period's human capital h_k of a child is given by

$$h'_k = h_0 + a[h_0(g_e + i_e + \nu(1 - n\lambda)h)]^{\gamma_e}, \tag{2.2}$$

where h_0 is the initial human capital all individuals are born with. Since more able children learn faster, returns to investment are increasing in innate ability a.

2.3.1.2 K-12 education

Primary and secondary education are commonly referred to as K-12 education, which in the model children attend in the period two and three of their life. Here the inputs are public and private education expenditures g_m and i_m, respectively, where subscript m stands for middle, as this educational stage is between preschool and college. The efficiency of these substitutes depends on γ_m, such that

$$h'_k = h_k + a[h_k(g_m + i_m)]^{\gamma_m}. \tag{2.3}$$

2.3.1.3 College

After three periods individuals leave the parental household and receive an intervivos transfer b, which is a one time gift by their parents. This gift can be saved in terms of capital k, spent on consumption c, or used to finance private college expenditures i_c. Borrowing is possible only for those whom enroll into college and they can only take up a loan in the period they study (more on this in Section 2.3.4). Their borrowing limit is ψ. Individuals spent time $\eta \in \{\bar{\eta}, \underline{\eta}\}$ in college and work the remaining share of the

period. Those that drop out only spend a fraction $\underline{\eta}$ of the period studying in college, while the duration of completion is $\bar{\eta}$ (where $\bar{\eta} > \underline{\eta}$). The curvature of the returns to investment in college education are governed by γ_c, such that

$$h' = h + a[h(g_c + i_c)]^{\gamma_c}. \tag{2.4}$$

Completion depends on the college completion probability $\pi(a,h)$, which is an increasing function in innate ability a and human capital h. The importance of ability and human capital for completion is determined by ω, and is given by

$$\pi(a,h) = 1 - e^{\omega ah}. \tag{2.5}$$

Henceforth, college graduates are considered as skilled $(s = 1)$ and those without a college degree as unskilled $(s = 0)$.

2.3.2 Labor supply

Each household has two adult members, a male and a female. Males inelastically provide one unit of labor each period. In all periods females' labor is supplied inelastically except for the labor supply of mothers of a preschool child. Mothers of a preschool-aged child face a disutility of ξ per unit of labor supplied, which can be interpreted as the cost of joint work for a household where both parents are working. It could also include the cost from a stigma attached to a mother of a preschool child working. The cost could be influenced by institutions or the cultural attitudes in society towards young mothers engaging in wage employment.[4] Depending on this cost, the family's capital, wage earnings, and public provision of preschool education the mother can decide to work part-time or full-time, or not work at all.

There are two reasons why female labor supply is inelastic in all periods except for during early childhood. First, the empirical literature emphasizes that the presence of preschool children has a substantially larger effect on parental employment than the presence of older children. Second, parental nurturing in early years has been found to have the largest impact on skill formation (Abbott 2014).

[4] Whether institutions are due to attitudes or whether attitudes are shaped by institutions is not investigated here.

2.3.3 Wages

The curvature of the life-cycle earnings profile is different for skilled and unskilled. Wages change over the life cycle, depend on human capital h and the potential periods of work experience j, and is given by

$$w(h, j, \upsilon, s) = h\phi_0 e^{\phi_1^s j + \phi_2^s j^2 + \phi_3^s j^3 + \upsilon}, \tag{2.6}$$

where ϕ_0 controls a level effect in common for both skilled and unskilled. The coefficients ϕ_1^s, ϕ_2^s, and ϕ_3^s are intended to capture changes in earnings over the life cycle through a cubic approximation.

Due to the gender wage gap, working females only earn the exogenous fraction θ of what males earn. Therefore, total earnings of a household in which both adults are working is $w + \theta w$.

In period $j = 1$, when individuals take up their first job, the earnings shock υ is assumed to be equal to zero. In all following periods, υ is meant to capture market luck and reflects the fact that wages are not fully explained by education. The distribution of υ is assumed to be given by $N(0, \sigma_\upsilon)$.

In period $j = 3$ in which the children are born, the young mother might decide not to work ($n = 0$), to work part-time ($n = 1/2$), or to work full-time ($n = 1$). Remember, the household will be missing less than half of its earnings when the young mother does not work. Therefore, the labor income in that period is given by

$$w(h, j = 3, n, \upsilon, s) = w(h, j = 3, \upsilon, s) + n\theta w(h, j = 3, \upsilon, s). \tag{2.7}$$

A working mother faces two opportunity costs if she decides to work. First, the household suffers utility cost $n\xi$. Second, she is able to spend less time with the children from which the children would profit cognitively. These two costs the household will weigh against the increased consumption and educational investments the additional wage can accomplish.

2.3.4 Capital

There is a risk free investment opportunity providing an exogenous rate of return r per period. Conditional on college attendance, individuals are permitted to borrow up to the borrowing limit ψ. This can be understood as college enrollment being the eligibility criteria for access to a loan program. College attendants are only allowed to take out a loan in period $j = 1$, when they are actually attending college. Subsequently, the borrowing limit declines linearly over the following periods so that all debt has to have been repaid before retirement. An individual that attended college is not allowed to increase the amount borrowed or take up a loan once the college period has ended.

2.3.5 Government

The government finances public education at the three stages and provides a pension to retired folks.

2.3.5.1 Public education

The government levies three proportional taxes on labor income to separately balance the budgets of public preschool, K-12 education, and college. Let Y be the total labor income and C define total consumption. For public preschool the government balances the following budget:

$$\tau_e Y = g_e. \tag{2.8}$$

For K-12 education the per capita expenditures are halved because it extends over two periods, such that

$$\frac{\tau_m Y}{2} = g_m. \tag{2.9}$$

The public college expenditures per student depend on how many students actually enroll in college and how much time η they spend in college. The budget constraint can be written as

$$\tau_c Y = g_c \int 1_{college=1}(\pi(a,h)\bar{\eta} + (1-\pi(a,h))\underline{\eta}) d\mu_1(a,h,k), \tag{2.10}$$

where $\mu_1(a,h,k)$ is the distribution of ability, human capital, assets of households of age $j = 1$ and $1_{college=1}$ is an indicator function taking the value one for those that attend

college and zero otherwise.

2.3.5.2 Social security

Social security benefits p are provided by the government as a function of averaged potential lifetime earnings, which depend on human capital h and whether the individual is skilled s.

A consumption tax τ_v and and additional earnings tax τ_p are used to finance public pension expenditures. Let μ_r summarize the entire distribution of retired folks. For social security the government has to satisfy the budget

$$\int p(s,h)d\mu_r(s,h) = \tau_p Y + \tau_v C. \tag{2.11}$$

Let τ be the sum of all proportional labor income taxes, such that $\tau = \tau_e + \tau_m + \tau_c + \tau_p$.

2.3.6 Utility

Utility from consumption c is given by a constant relative risk aversion utility function. When children are with their parents, which occurs in periods $j = 3,4,5$, per-capita consumption is given by $\frac{c}{\xi}$ with $\xi > 1$, such that

$$u(c) = \frac{(\frac{c}{\chi})^{1-\sigma}}{1-\sigma}. \tag{2.12}$$

Therefore, the present value of the lifetime utility can written as

$$U = \sum_{j=1}^{J} \beta^{j-1} u(c) + \beta^T \alpha U_c, \tag{2.13}$$

where α is the altruism felt for the child and U_c is the lifetime utility the child.

2.3.7 Recursive formulation

The life of a household can be separated into 3 stages. First, one makes the college decision, which in case of graduation does not only lead to a higher h, but also a different

life-cycle earnings profile due to being skilled ($s = 1$). Second, one works, has children, decides on how much to spend on private preschool and K-12 education, and when the child leaves home passes on a intervivos transfer. Finally, retirement takes place.

2.3.7.1 College decision

For $j = 1$ the state variables are a household's capital endowment, which is the intervivos transfer b received by the parents, the level of human capital as a child h_k, and innate ability a. Let $V_j(\cdot)$ denote the value function of period j, such that $V_1(b,h_k,a)$ is the value function of a household of age $j = 1$ when making the college decision, which is given by

$$V_1(b,h,a) = max\{W(college = 0,\cdot), W(college = 1,\cdot)\}, \tag{2.14}$$

where the value of not going to college $W(0,b,h,a)$ is given by

$$W(0,b,h_k,a) = \max_{c,i_c,k'} u(c) + \beta E[V_2(0,k',h,a,\upsilon')] \tag{2.15}$$

subject to

$$\begin{gathered} (1+\tau_v)c + k' = (1-\tau)(1+\theta)w + (1+r)b \\ h = h_k \\ w = h\phi_0 e^{\phi_1^0 j + \phi_2^0 j^2 + \phi_3^0 j^3} \\ k' \geq 0, c > 0. \end{gathered} \tag{2.16}$$

The value of going to college $W(1,b,h,a)$ is given by

$$W(1,b,h_k,a) = \max_{c,i_c,k'} u(c) + \beta\pi(a,h)E[V_2(1,k',h,a,\upsilon')] + \beta(1-\pi(a,h))E[V_2(0,k',h,a,\upsilon')] \tag{2.17}$$

subject to

$$
\begin{aligned}
(1+\tau_v)c + k' + i_c &= (1-\tau)(1-\eta)(1+\theta)w + (1+r)b \\
h &= \begin{cases} h_k + a[h_k(g_c+i_c)]^{\gamma_c} & \text{if } s=1 \\ h_k & \text{if } s=0 \end{cases} \\
\eta &= \begin{cases} \bar{\eta} & \text{if } s=1 \\ \underline{\eta} & \text{if } s=0 \end{cases} \\
w &= h\phi_0 e^{\phi_1^s t + \phi_2^s j^2 + \phi_3^s j^3} \\
k' &\geq -\psi, c > 0,
\end{aligned}
\tag{2.18}
$$

where the probability of college completion $\pi(a,h)$ is given by (2.5). In case of completion s is equal to one and zero otherwise.

2.3.7.2 Working and parental life

For $2 \leq j \leq 8$ the state variables are whether one is skilled s, the capital endowment k, human capital h, human capital of the children h_k, innate ability a, and the labor earnings shock υ. The value function of a household of age j is denoted by V_j and is given by

$$
\begin{aligned}
V_j(s,k,h,h_k,a,\upsilon) = \max_{c,i_e,i_m,b,k',n} u(c) - n\xi 1_{j=3} + \\
\beta E[V_{j'}(s,k',h,h_k',a',\upsilon')] + 1_{j=6}\alpha E[V_1(b,h_k'',a')],
\end{aligned}
\tag{2.19}
$$

subject to

$$
\begin{aligned}
(1+\tau_v)c + k' + i_e 1_{j=3} + i_m 1_{j=4,5} + b1_{j=6} &= k(1+r) + (1-\tau)(1+n\theta)w \\
h_k' &= \begin{cases} h_0 + a'[h_0(g_e+i_e+\nu(1-n\lambda)h)]^{\gamma_e} & \text{if } j=3, \\ h_k + a'[h_k(g_m+i_m)]^{\gamma_m} & \text{if } j=4,5 \end{cases} \\
w &= h\phi_0 e^{\phi_1^s j + \phi_2^s j^2 + \phi_3^s j^3 + \upsilon} \text{ where } \upsilon \sim N(0,\sigma_\upsilon^2) \\
ln(a') &= \rho ln(a) + \varepsilon \text{ where } \varepsilon \sim N(0,\sigma_a^2) \text{ if } j=3 \\
k' &\geq min\{0, max\{k, -\psi\frac{8-j}{7}\}\} \\
c &> 0, j' = j+1, n \in \{0, \tfrac{1}{2}, 1\},
\end{aligned}
\tag{2.20}
$$

where 1_j denotes an indicator function, which takes the value 1 in period j and zero otherwise. The indicator functions are required because mothers can choose not to work when their children are of preschool age, certain education investments only take place in the corresponding periods, and the altruistic utility is enjoyed when children leave home in period $j = 6$. The expression for assets in the following period, k', includes the two restrictions on assets. First, one cannot increase one's level of debt. Second, the borrowing limit declines linearly, thereby forcing individuals to repay their college loan before retirement.

2.3.7.3 Retired

For $j \geq 9$ the state variables are skill level s and human capital h, which together determine pension benefits, and assets k. Therefore the value functions $V_j(s,h,k)$ are given by

$$V_j(s,k,h) = \max_{c,k'} u(c) + \beta\Gamma(j)E[V_{j'}(s,k',h)], \tag{2.21}$$

subject to

$$\begin{aligned} (1+\tau_v)c + k' &= k(1+r) + p(s,h) \\ c > 0, k' &\geq 0, j' = j+1, \end{aligned} \tag{2.22}$$

where $\Gamma(j)$ is the survival probability of a household of age j and $p(s,h)$ determines pension benefits as a function of skill s and human capital h.

2.4 Equilibrium

The interest rate r is taken as exogenous. Labor income is given by the wage process defined in (2.6). One equilibrium object is the proportional tax rate τ_p, which is required to finance social security benefits. The remaining equilibrium objects are the stationary distribution of agents across states, the borrowing limit, and the public education expenditures dedicated to preschool/child care, K-12, and college.

Let x_j summarize the state variables of a household in life-cycle period j.

DEFINITION 2. *Given an interest rate r and tax rates $(\tau_e, \tau_m, \tau_c, \tau_v)$, an equilibrium is such that V_1 solves the functional equation (2.14), (2.15), and (2.17) satisfying the constraints in (2.16) and (2.18), $V_{2\leq j\leq 9}$ solves the functional equation (2.19) satisfying*

the constraints in (2.20)*,* $V_{9\leq j\leq 13}$ *solves the functional equations* (2.21) *satisfying the constraints in* (2.22)*, with* $i_c(x_1)$*,* $i_e(x_3)$*,* $i_m(x_4)$*,* $i_m(x_5)$*,* $c_j(x_j)$*,* $k_j(x_j)$*, and* $b(x_6)$ *as associated policy functions.*

1. *Goods market clearing*

$$Y + rK = C + I_e + I_m + I_c \tag{2.23}$$

$$Y = \int \int w(h, j, \upsilon, s) d\mu_j(x_j) dj \tag{2.24}$$

$$C = \int \int c_j(x_j) d\mu_j(x_j) dj \tag{2.25}$$

$$I_e = \int i_e(x_3) d\mu_3(x_3) \tag{2.26}$$

$$I_m = \int i_m(x_4) d\mu_4(x_4) + \int i_m(x_5) d\mu_5(x_5) \tag{2.27}$$

$$I_c = \int i_c(x_1) d\mu_1(x_1) \tag{2.28}$$

2. *The government balances budgets* (2.8)*,* (2.9)*,* (2.10)*, and* (2.11)*.*

3. *The laws of motion* Φ_c*, which map from state* x_1 *to* x_2*, are given by* (2.4) *and* (2.5)*. The law of motion* Φ_a*, which maps from state* x_2 *to* x_3*, is given by* (2.1)*. The law of motion* Φ_e*, which maps from state* x_3 *to* x_4*, is given by* (2.2)*. The law of motion* Φ_m*, which maps from state* x_4 *to* x_5 *and* x_5 *to* x_6*, is given by* (2.3)*. The law of motion* Φ_s*, which maps from state* x_9 *sequentially to all remaining states (*x_{13} *being the last), is given by* $\Gamma(j)$*. All remaining transitions between states, are the transitions of capital from* k *to* k'*, which are given the respective budget constraints, and the transitions between labor shocks from* υ *to* υ'*, which are independent of each other. The distribution of* υ *is given by* $N(0, \sigma_\upsilon)$*.*

A stationary equilibrium is a competitive equilibrium in which policy functions, as well as public education expenditures and tax rates are constant. The distributions μ_j *are stationary.*

2.5 Model paramerization

Given the complexity of the model deriving an analytical solution is not possible, wherefore the model is solved numerically. While 19 parameters are chosen a priori based on previous literature or on direct estimates from the data, the remaining 11 parameters are calibrated to match key statistics in terms of earnings inequality, earnings persistence, and private education expenditures in the US. The 11 parameters are determined by minimizing the sum of squared deviation in terms of percentage of 11 model moments to 11 targets. In the following, the sources of the a priori chosen parameters and targets are outlined.

2.5.1 Independently chosen parameters

The independently chosen parameters are related to the age structure of the model, preferences, earnings, human capital creation, college, tax rates, and the interest rate.

One period in the model is equivalent to 6 years and J is set to 13, so that individuals can live up to 96 years. The survival probabilities for retired households are obtained from Social Security data and are specified in Section 2.8.3 in the Appendix.

The discount rate is 0.98 per year, which results in $\beta = 0.8858$. I choose a standard value in the consumption literature of 1.5 for the intertemporal preference parameter σ. As standard in the literature the household consumption discount factor χ of 1.4 is obtained from the OECD modified equivalence scale.

All coefficients of the life-cycle wage schedule, except for ϕ_0, I obtain by estimation using the Panel Study of Income Dynamics as specified in more detail in Section 2.8.1 in the Appendix. The gender wage gap is taken from the OECD Employment Database according to which in 2007 women earn 80.2% of what men earn.

A mother of a young child that does work also tends to spend a considerable amount of active time with her child. According to Heiland and Price (2012) working mothers on average spend 60% as much time with their children as do those that do not work, such that $\lambda = 0.4$. The initial human capital h_0, which every individual is born with, is normalized to unity.

College completion requires four years of attendance, which given a period length of 6 years translates into $\bar{n} = \frac{2}{3}$. Supported by evidence of Stinebrickner and Stinebrick-

ner (2007) dropping out most frequently occurs after two years of college attendance ($\underline{n} = \frac{1}{3}$). As in Holter (2014) the college borrowing limit is set to 24,658US$ (Lochner and Monge-Naranjo 2011), which is the maximum amount available under the Stafford program to which most students are eligible.

Public education expenditures as a share of GDP are taken from the OECD (2013). The US only spends 0.4% of GDP on public child care and preschool. Given that a large share of public education expenditures is determined at the local level, I follow the approach by Restuccia and Urrutia (2004) and Holter (2014) and consider federal and state K-12 expenditures as public expenditures and local expenditures I allocate to the private expenditures. From the total public K-12 expenditures of 3.7% of GDP, 13% come from the federal level and 43% the state, which translates into 2.1% public education expenditures for primary and secondary education. At the college level the public expenditures amount to 1% of GDP. The average consumption tax is set to 9.6% (Vertex Inc. 2012).

The interest rate is set conservatively to the risk-free return of the average of 3-month T-bill rates minutes inflation based on data from the Federal Reserve Bank of St. Louis over the period 1947-2008, which Holter (2014) documents as 1.1% per annum, wherefore $r = 0.0678$.

2.5.2 Targets for calibration

2.5.2.1 Earnings

Intergenerational earnings persistence is amongst the highest in the developed world in the US. According recent estimates in a survey by Corak (2013), earnings persistence in the US is as high as 0.47, which is the target of ρ, the correlation coefficient of the innate ability transmission process from parents to children.[5] The standard deviation of the earnings shock σ_υ is pinned down by the Gini of males' (aged 25-60) hourly earnings before taxes and transfers, which in 2005 was 0.39 (Heathcote, Perri, and Violante 2010). The standard deviation of the noise in the innate ability transmission σ_a is linked

[5]Earnings persistence is generally measured as the intergenerational earnings elasticity obtained by regressing the log of the child's adulthood earnings on the log of parental lifetime earnings as in $ln(Y_{child}) = \alpha + \beta(Y_{parents}) + \varepsilon$. The coefficient of interest is β, which captures the percentage increase in the child's earnings for a 1% increase of parental earnings above the mean.

to the variance of log hourly wages of males aged 25-60, which according to Heathcote et al. (2010) in 2005 was 0.47.

2.5.2.2 Private education expenditures

The parameters governing the returns to investment in education are γ_e, γ_m, and γ_c. As a target for γ_e the amount of privately invested funds on preschool/child care is chosen. According to the Census Bureau report "Who's Minding the Kids? Child Care Arrangements", which uses the Survey of Income and Program Participation, households with a child under the age of 6 in 2008 on average spend 9.300US$ in child care/preschool per year. In the American Community Survey (ACS) of 2011 provided by the Integrated Public Use Microdata Series (Ruggles et al 2010) 16,736,779 households in the US have a child younger under the age of 6. Multiplying this with average expenditures of 9.300US$ this accounts for 1.08% of the GDP of 14369.4 Billion US$ in 2008.

In the US 0.3% of GDP were spent on private K-12 education in 2010, while total public expenditures on K-12 education in the US in 2010 accounted 3.7% of GDP (OECD 2013). Following Restuccia and Urrutia (2004) and Holter (2014), I count the expenditures financed at the local level as private expenditures as well, which account for 44% of total public expenditures. Therefore, γ_m, the parameter determining the efficiency of K-12 investments, is pinned down by private K-12 expenditures accounting for 1.9% of GDP.

College education in the US is characterized by very high and increasing private expenditures. In 2010 according to the OECD (2013) private expenditures on college education summed up to 1.8% of GDP, which serves as a target for γ_c, the parameter governing the curvature of returns to college education investments.

2.5.2.3 College and intervivos transfer

In the model a college degree does not only increase the level of human capital of an individual, but also alters the life-cycle earnings profiles. For college graduates the earnings profile is steeper at younger ages and decreases to a lesser extent when approaching retirement. Therefore, the remaining wage parameter ϕ_0 is targeted by the college enrollment rate. In 2009, of those completing high school, 71% enrolled into college

according to the National Science Foundation, while the high school completion rate is estimated at 88% (Heckman and Lafontaine 2010).[6] This results in 62% enrolling into college.

The parameter ω, determining the responsiveness the college completion function to the level of human capital and ability, is pinned down by the fraction of individuals completing college, 58% of first-time, full-time students who enrolled at a 4-year institution in fall 2004 according to the NCES.[7]

The intervivos transfer b plays an important role financing college education. Using the National Longitudinal Study of Youth, Abbott et al. (2013) find that between 1997 and 2003 on average 18-24 year olds received 30,566US$ as a gift from their parents. This serves as a target for α, the altruism parameter.

2.5.2.4 Mothers' labor force participation

Using the the ACS of 2011, I compute the share of mothers with a preschool-aged child at home that engage in full- or part-time employment. Full-time employment is 40% and is used to pin down ξ, the disutility of joint labor supply. Part-time employment is chosen by 17% of the mothers and serves as a target for ν, the efficiency parameter of the maternal time input in the children's human capital creation.

2.5.3 Benchmark results

The calibration's objective is to minimize the sum of squared deviations in percentage terms of the 11 moments for which the results can be seen in Table 2.2. The benchmark calibration successfully matches the intergenerational earnings persistence and earning inequality in terms of the Gini before taxes and transfers. The targets related to private education investment at the three educational stages come reasonably close to those encountered in the data. In the model, parents on average provide intervivos transfer which are slightly more generous than in the data. The percentage of individuals enrolling in college and the share of graduates is matched by the calibration. Mothers' employment is matched exactly. While one percentage point more than in the data work full time this is compensated by one percentage point less working part time.

[6] See http://www.nsf.gov/statistics/seind12/c1/c1s4.htm.
[7] See http://nces.ed.gov/fastfacts/display.asp?id=40.

Table 2.1: Benchmark model parameters

Description	Independently chosen: Parameter	Value	Reference
Coefficient of relative risk aversion	σ	1.5	Standard
Discount factor	β	0.98	Standard
Interest rate	r	0.011	Return 3-month T-bill (1968-2008)
Time for college completion	$\bar{\eta}$	4 years	Standard
Time for college dropout	η	2 years	Stinebrickner & Stinebrickner 2007
Wage coefficients non-college	$\phi_1^0, \phi_2^0, \phi_3^0$	.0001, .014, -.003	PSID (1968 - 2011)
Wage coefficients college	$\phi_1^1, \phi_2^1, \phi_3^1$	.0301, .051, -.004	PSID (1968 - 2011)
Gender wage gap	ν	0.8	OECD
Time working mothers w. child	λ	0.6	Heiland and Price (2012)
Consumption tax	τ_v	9.6%	Vertex Inc. (2012)
Public exp. on preschool	g_e	0.36% GDP	OECD 2013
Public exp. K-12	g_m	2.1% GDP	OECD 2013
Public exp. on college	g_c	1% GDP	OECD 2013
College borrowing limit	ψ	24,658$	Lochner & Monge-Naranjo 2011
HH consumption discount	χ	1.4	OECD modified equivalence scale
Death probability	Γ	Section 2.8.3	Social Security data
Pension	p	Section 2.8.2	Social Security data

Description	Calibrated: Parameter	Value	Target	Data
Persistence ability transmission	ρ	.34	Intergen. earnings persist.	0.47
STD of earnings shocks.	σ_υ	.023	Earnings Gini before tax	0.39
STD of noise in ability trans.	σ_a	.43	Variance of log hourly wages	0.47
Elasticity wrt early education	γ_e	.39	Private preschool exp./GDP	1%
Elasticity wrt early education	γ_m	.28	Private K-12 exp. /GDP	1.9%
Elasticity wrt early education	γ_c	.60	Private college exp.s/GDP	1.8%
Parental altruism	α	.30	Average intervivos transfer	30,566$
Wage level coefficient	ϕ_0	.17	College enrollment	0.62
College completion wrt a & h	ω	-.23	College completion rate	0.58
Utility cost mother working	ξ	1.07	% of mothers working full-time	0.40
Efficiency mother time with child	ν	.012	% of mothers working part-time	0.17

Table 2.2: Calibration of the US economy

Target	Data	Model
Intergenerational earnings persistence	.47	.47
Gini before tax	.39	.39
Variance log wages	.47	.52
Private preschool expenditure/GDP	.011	.009
Private primary&secondary expenditure/GDP	.019	.022
Private college expenditure/GDP	.018	.017
Average intervivos transfer	30,566$	32,005$
Fraction attending college	.62	.62
College completion rate	.58	.58
% of mothers working full-time	.40	.41
% of mothers working part-time	.17	.16

The parameter for the correlation in the intergenerational ability transmission, ρ, is 0.34. Given an intergenerational earnings persistence of 0.47, one can argue that earnings persistence is not fully accounted for by genetics but arises endogenously due to parental investments. The parameter governing the returns to preschool investments, γ_e, is 0.38, considerably higher than the efficiency parameter of K-12 investments ,γ_m, which is 0.28.[8] This is in accordance with the literature claiming that earlier investments are more efficient than later educational investments (e.g., Cunha and Heckman 2008). However, the efficiency parameter of college investments, γ_c, is very high at 0.6, which reflects the high college premium in the US.

2.5.3.1 Beyond targets

The model comes reasonably close to a range of data moments not explicitly targeted in the calibration. Concerning the earnings distribution, the model is disciplined to replicate the Gini. Percentile ratios of earnings of males aged 25-60 documented by Heathcote et al. (2010), such as the P50/P10 percentile ratio in the data of 2.31 versus 2.49 in the benchmark economy and the P90/P50 percentile ratio of 2.33 in the data versus 2.18 in the model, are closely aligned. Another relevant characteristic of the US earnings distribution is that the mean is greater than the median, with a ratio of 1.36

[8]Empirical estimates in the literature looking at changes in public expenditures, which might suffer from a downward bias due to neglecting substitution effects of private investments, range from below 0.2 (Card and Krueger 1992, 1996) to 0.4-0.6 (Haley 1976).

according to ACS of 2011 provided by Ruggles et al. (2010) for employed individuals between 18 and 66, which in the benchmark economy turns out to be 1.09.

2.6 Counterfactuals

In the following, two channels are exploited to explain cross-country differences in intergenerational earnings mobility. First, I simulate the model with varying levels of public preschool expenditures. Second, the disutility of joint work for households with a preschool-aged child is varied exogenously.

2.6.1 Cross-country differences in public preschool expenditures

The dispersion of expenditures on preschool across countries is large. In the following, I investigate whether differences in public preschool expenditures can explain cross-country differences intergenerational mobility. All parameters are left as in the US benchmark economy, except for τ_e, the tax rate financing public preschool, which is replaced by the share of GDP dedicated to public preschool in the corresponding country. The analysis is restricted to steady states only and the results are summarized in Table 2.5 in the Appendix.

Only accounting for differences in public preschool expenditures the model on average is able to explain 17% of the gap in intergenerational earnings persistence between the US and the 15 OECD countries of which the requisite data was available. Excluding the South American countries the counterfactual experiment on average accounts for as much as 30% of the differences in earnings persistence.

Mothers' labor supply is hardly affected by changes in public preschool expenditures. While this is a surprising finding it does not necessarily contrast the empirical evidence. While a range of studies do find a positive response of mothers' labor supply to public preschool/child care provision, Havnes and Mogstad (2011), for instance, find that an expansion of childcare in Norway did not lead to a causal change in the labor market attachment of mothers.

Importantly, the model does not only perform well at explaining differences in earnings persistence on average but also at replicating the patterns observed in the data. In Figure 2.7 one can see that public preschool expenditures and intergenerational earn-

ings persistence exhibit a hump-shaped relationship, just as the model generates. This happens due to the interaction of two complementarities embedded in the model. First, between innate ability and the returns to investment in preschool education. Second, the dynamic complementarity between human capital gained in the preschool period and returns to educational investments in the following periods. In general 22% of the variation in earnings persistence is explained, which becomes clear in Figure 2.8. The correlation between the data moments compared to the moments produced by the model is 0.47.

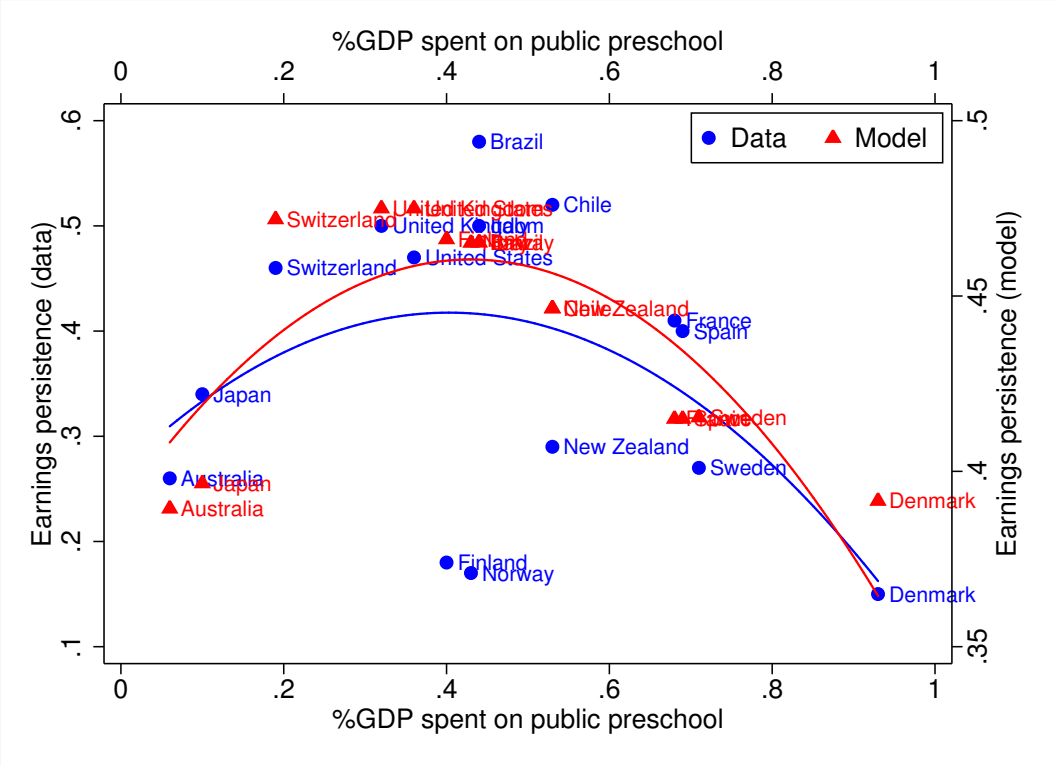


Figure 2.7: Public preschool expenditures and earnings persistence in data (red) and model (blue)

2.6.2 Mothers' labor supply and intergenerational ability

Given that changes in public preschool expenditures have a negligible effect on mothers' labor force participation, the previous counterfactual exercise does not shed light

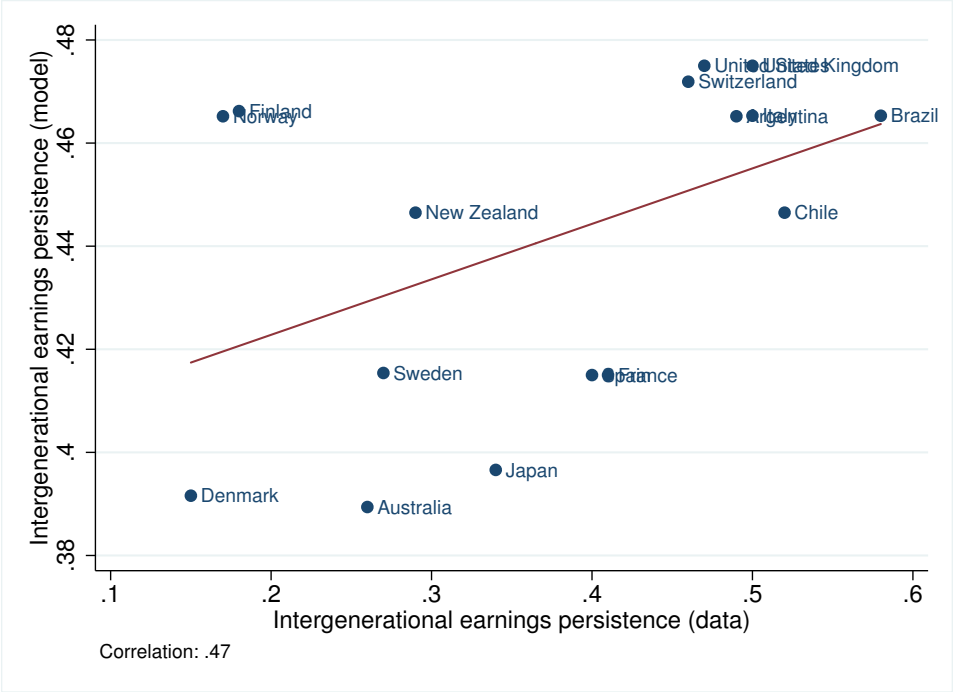


Figure 2.8: Intergenerational earnings persistence in data and model

on how intergenerational earnings persistence might be affected by mothers working. In the model, the disutility cost ξ of joint work when the child is of preschool age is the primary driver of mothers' work decision.[9]

In the benchmark case ξ takes the value 1.07. Full-time employment is chosen by 41% of mothers, while 16% decide to work part time. The remaining 57% do not work at all and stay at home with their children. In the following counterfactual exercise, female employment is facilitated by lowering, and hampered by increasing ξ exogenously. Table 2.3 shows the changes in mothers' employment as well as earnings persistence, which arise when the value of ξ varies from -10% to +10% of its benchmark value.

Unsurprisingly, total employment of mothers with preschool-aged children is negatively related with the disutility cost of joint work. When the utility cost ξ reduces by 10% compared to its benchmark value, the employment of mothers increases by 11 percentage points. Less women work part time, but this reduction is compensated by the increase in full-time employment. Similarly, an increase of 10% of ξ leads to a 14

[9]Cultural or institutional forces, such as parental leave regulations, could be responsible for cross-country variations in ξ. See Section 2.2 for further discussion.

Table 2.3: Counterfactual public preschool expenditures

	Mothers' employment			
Disutility ξ	Total	Full-time	Part-time	Earnings persistence
-10%	.69	.58	.11	.40
-5%	.62	.49	.13	.46
Benchmark	.57	.41	.16	.47
+5%	.50	.35	.15	.48
+10%	.43	.33	.10	.51

percentage point reduction of mothers employment. Here both full- and part-time employment decrease.

Interestingly, when more mothers stay at home with their children, we observe an increase earnings persistence. This is because the mothers human capital influences the child's cognitive development during the time they spend with them. So if now more mother are at home with their children, human capital is correlated to a larger extent across generations. Conversely, earnings persistence declines when more mothers are employed due to the analog mechanism.

2.7 Conclusions

I introduce a model which includes preschool education as a stage in the human capital creation process. Children of young age are not only influenced by preschool, but also by the time input by their mothers. The efficiency of the time input is increasing in the mother's human capital. Human capital accumulated early on exhibits dynamic complementarities with education investments during K-12 education and college. The model is calibrated to the US replicating key statistics in terms of intergenerational mobility, inequality, private education investments, and mothers' labor force participation.

By exploiting differences in public preschool expenditures across countries the model successfully explains a considerable share of the cross-country variation in intergenerational mobility and replicates the hump-shaped association between public preschool expenditures and earnings persistence. Mothers' labor supply is found to hardly respond to changes in public preschool provision.

When exogenously increasing (decreasing) the disutility of joint labor supply, mothers respond by exiting (joining) the labor force. When less (more) mothers participate in the labor force, intergenerational earnings persistence decreases (increases). This is due to mothers sacrificing time for work, which otherwise they would spend with their children.

Three interesting extensions are left to further research. First, the gender wage gap could be endogenized by penalizing females for accumulating less work experience due to the time not worked. Second, fertility is treated as constant in the present model but has been shown to be affected by the key model ingredients. Third, public preschool expenditures and the disutility of joint labor supply for households with a preschool aged child have been treated separately and as exogenous. A Political Economy approach in which the disutility of joint labor supply and public preschool provision are jointly determined could yield interesting dynamics.

2.8 Appendix

2.8.1 Wage coefficients

In order to compute the coefficients determining the life-cycle earnings profile, I use all observations of the Panel Study of Income Dynamics (PSID) from 1968 to 2011. The separate coefficients for those with and without a college degree are estimated exploiting the panel dimension of the data. In the panel regression, I control for individual and year fixed effects. In accordance with previous empirical literature (e.g., Kambourov and Manovskii 2009), college graduates' earnings profile initially increases more steeply than those that have no college degree, and drops off less drastically in the years leading up to retirement.

2.8.2 Social security

Using data provided by the Social Security Administration, I approximate the pension profile as a function of average lifetime earnings w_{life} and average earnings $\bar{w}$.[10] The share of lifetime earnings provided as retirement benefits is decreasing in average lifetime earnings w_{life}, which depend on the level of human capital and whether the individual is skilled. Including a cubic term of w_{life} the estimated profile of pensions p turns out to be:

$$p(w_{life}(s,h)) = (-0.1117(\frac{w_{life}}{\bar{w}})^2 + 0.6425\frac{w_{life}}{\bar{w}} + 0.1422)\bar{w}. \tag{2.29}$$

2.8.3 Survival probabilities

During retirement I account for the fact that people might die before age 96. The probability of survival I obtain from data the Social Security Administration and is exhibited in Table 2.4.[11] First, I take the mean of males and females and then calculate the probability that an individual will survive until the next period.

[10] See http://www.ssa.gov/pubs/EN-05-10070.pdf and http://www.ssa.gov/oact/cola/AWI.html for the tables.

[11] See http://www.ssa.gov/OACT/STATS/table4c6.html.

2.8.4 Tables

Table 2.4: Survival probability

Period j	Age	Survival
9	66-72	.90
10	72-78	.73
11	78-84	.54
12	84-90	.30
13	90-96	$0^{\dagger}$

† Imposed by author, not obtained from the data.

Table 2.5: Counterfactual public preschool expenditures

Country	Public preschool %GDP	Earnings persistence		
		Data	Model	$\Delta_{explained}$
Argentina	.43	.49	.465	-0.49
Australia	.06	.26	.389	0.41
Brazil	.44	.58	.465	-0.09
Chile	.53	.52	.447	-0.57
Denmark	.93	.15	.392	0.26
Finland	.4	.18	.466	0.03
France	.68	.41	.415	1
Italy	.44	.5	.465	-0.32
Japan	.1	.34	.397	0.60
New Zealand	.53	.29	.447	0.16
Norway	.43	.17	.465	0.03
Spain	.69	.4	.415	0.86
Sweden	.71	.27	.415	0.30
Switzerland	.19	.46	.472	0.31
United Kingdom	.32	.5	.475	0
United States	.36	.47	.475	

Datasources: Education expenditures from OECD 2013. Earnings persistence from Corak 2013.
Note: The share of a gap between a data moment in the US d_{US} and another country $d_{country}$ explained by moments m_j produced by the model, is calculated using the following definition: $\Delta = \frac{(m_{US} - m_{country})}{(d_{US} - d_{country})}$.

2.8.5 Figures

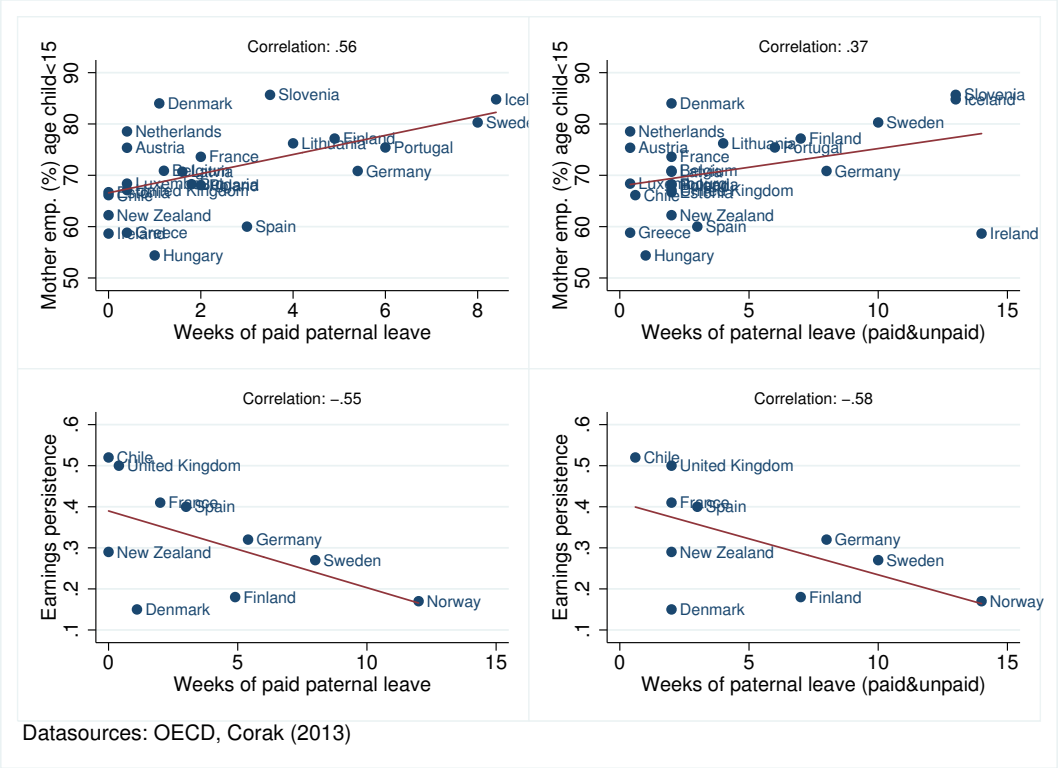


Figure 2.9: Paternal work leave and intergenerational earnings persistence (bottom) and mothers' employment (top)

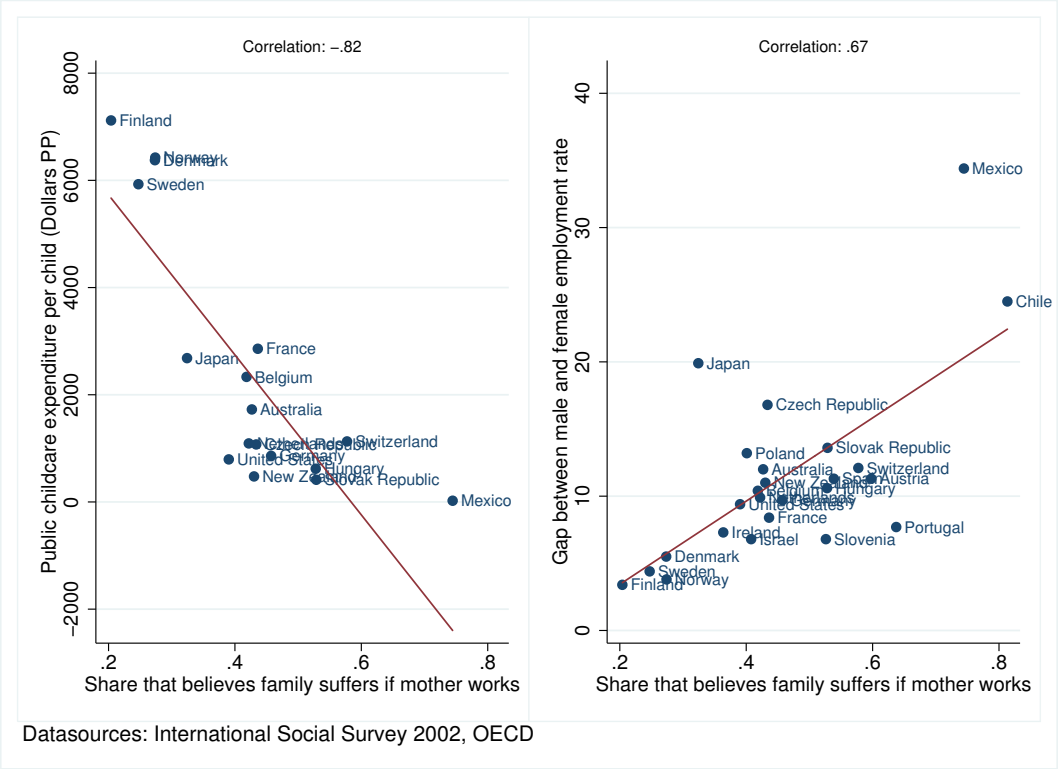


Figure 2.10: Attitudes to mothers working, preschool expenditures, and female employment

References

[1] Abbott, B. (2014). Optimal Policies for Human Capital Investments in Children and Youth. Mimeo.

[2] Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2013). Education policy and intergenerational transfers in equilibrium (No. w18782). National Bureau of Economic Research.

[3] Attanasio, O., Low, H., and Sanchez-Marcos, V. (2008). Explaining changes in female labor supply in a life-cycle model. *American Economic Review*, 1517-1552.

[4] Baker, M., Gruber, J., and Milligan, K. (2008). Universal Child Care, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy*, 116(4), 709-745.

[5] Bernal, R. (2008). The Effect of Maternal Employment and Child Care on Children's Cognitive Development. *International Economic Review*, 49(4), 1173-1209.

[6] Bernal, R., and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1), 164-189.

[7] Bernal, R., and Keane, M. P. (2011). Child care choices and children's cognitive achievement: The case of single mothers. *Journal of Labor Economics*, 29(3), 459-512.

[8] Bick, A. (2013). The quantitative role of child care for female labor force participation and fertility. Mimeo.

[9] Blau, D. M. (1999). The effect of child care characteristics on child development. *Journal of Human Resources*, 786-822.

[10] Blau, D., and Tekin, E. (2007). The determinants and consequences of child care subsidies for single mothers in the USA. *Journal of Population Economics*, 20(4), 719-741.

[11] Brilli, Y., Del Boca, D., and Monfardini, C. (2013). Child care arrangements: determinants and consequences (No. 18). Centre for Household, Income, Labour and Demographic Economics (CHILD)-CCA.

[12] Brooks-Gunn, J., Han, W. J., and Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The NICHD study of early child care. *Child development*, 73(4), 1052-1072.

[13] Card, D., and Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1), 1-40.

[14] Card, D., and Krueger, A. B. (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina. *Journal of Economic Perspectives*, 10(4), 31-50.

[15] Corak, M. (2013). Inequality from Generation to Generation: The United States in Comparison. In Rycroft, R. (ed.). The Economics of Inequality, Poverty, and Discrimination in the 21st Century. ABCCLIO.

[16] Cunha, F., and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-82.

[17] Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of Population Economics*, 15(3), 549-573.

[18] Del Boca, D., Flinn, C., and Wiswall, M. (2013). Transfers to Households with Children and Child Development. Mimeo.

[19] Del Boca, D., Pasqua, S., and Suardi, S. (2012). Childcare, family characteristics and child outcomes: an analysis on Italian data (No. 9). Centre for Household, Income, Labour and Demographic Economics (CHILD)-CCA.

[20] Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 111-134.

[21] Domeij, D., and Klein, P. (2013). Should day care be subsidized?. *The Review of Economic Studies*, 80(2), 568-595.

[22] Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child development*, 74(5), 1454-1475.

[23] El-Attar, M. (2013). Trust, child care technology choice and female labor force participation. *Review of Economics of the Household*, 11(4), 507-544.

[24] Fernandez, R., and Fogli, A. (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics*, 1:1, 146-177.

[25] Fogli, A., and Veldkamp, L. (2011). Nature or nurture? Learning and the geography of female labor force participation. *Econometrica*, 79(4), 1103-1138.

[26] Griffin, D. (2011). Evaluating the Effects of Child Care Policies on Children's Cognitive Development and Maternal Labor Supply. Mimeo.

[27] Guner, N., Kaygusuz, R., and Ventura, G. (2012). Taxation and household labour supply. *The Review of Economic Studies*, 79(3), 1113-1149.

[28] Haley, W. J. (1976). Estimation of the earnings profile from optimal human capital accumulation. *Econometrica*, 44(6), 1223-38.

[29] Hart, B., and Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Paul H Brookes Publishing.

[30] Havnes, T., and Mogstad, M. (2010). Is universal child care leveling the playing field? Evidence from non-linear difference-in-differences (No. 4978). Discussion paper series//Forschungsinstitut zur Zukunft der Arbeit.

[31] Havnes, T., and Mogstad, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, 95(11), 1455-1465.

[32] Hawkinson, L. E., Griffen, A. S., Dong, N., and Maynard, R. A. (2013). The relationship between child care subsidies and children's cognitive development. *Early Childhood Research Quarterly*, 28(2), 388-404.

[33] Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006. *Review of Economic Dynamics*, 13(1), 15-51.

[34] Heckman, J. J. (1976). Estimates of a human capital production function embedded in a life-cycle model of labor supply. In Household production and consumption (pp. 225-264). NBER.

[35] Heckman, J. J., and LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92(2), 244-262.

[36] Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1), 114-128.

[37] Heiland, F. and Price, J. (2012). Maternal Employment and Parent-Child Interaction. Mimeo.

[38] Herbst, C. M., and Tekin, E. (2010). The impact of child care subsidies on child well-being: Evidence from geographic variation in the distance to social service agencies (No. w16250). National Bureau of Economic Research.

[39] Herbst, C. M., and Tekin, E. (2012). Child care subsidies, maternal well-being, and child-parent interactions: Evidence from three nationally representative datasets (No. w17774). National Bureau of Economic Research.

[40] Herrington, C. (2014). Public Education Financing Systems, Earnings Inequality and Intergenerational Mobility. Mimeo.

[41] Holter, H. A. (2014). Accounting for cross-country differences in intergenerational earnings persistence: the impact of taxation and public education expenditure. Mimeo.

[42] Kambourov, G., and Manovskii, I. (2009). Accounting for the Changing Life-Cycle Profile of Earnings. Mimeo.

[43] Love, J. M., Harrison, L., Sagi-Schwartz, A., Van Ijzendoorn, M. H., Ross, C., Ungerer, J. A., Raikes, H., Brady-Smith, C., Boller, K., Brooks-Gunn, J., Constantine, J. Kisker, E.E., Paulsell, D., and Chazan-Cohen, R. (2003). Child care quality matters: How conclusions may vary with context. *Child development*, 74(4), 1021-1033.

[44] Ludwig, J., and Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, 122(1), 159-208.

[45] McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology*, 20(2), 244.

[46] Melhuish, Edward C., Ann Mooney, Eilis Hennesy, and Sue Martin. 1992. Characteristics of child care in early childhood and child development in middle childhood. Paper presented at the European Conference on Developmental Psychology, Seville, September.

[47] OECD (2013). Education at a Glance. Paris.

[48] Rauh, C. (2014). The Political Economy of Early and College Education - Can Voting Bend the Great Gatsby Curve?. Mimeo.

[49] Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language*, 35(1), 185.

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

[51] Schoellman, T. (2014). Early Childhood Human Capital and Development. Mimeo.

[52] Suskind, D., Leffel, K. R., Hernandez, M. W., Sapolich, S. G., Suskind, E., Kirkham, E., and Meehan, P. (2013). An Exploratory Study of "Quantitative Linguistic Feedback" Effect of LENA Feedback on Adult Language Production. *Communication Disorders Quarterly*, 34(4), 199-209.

[53] Tekin, E. (2005). Child care subsidy receipt, employment, and child care choices of single mothers. *Economics Letters*, 89(1), 1-6.

[54] Witte, A. D., and Trowbridge, M. (2004). The structure of early care and education in the United States: Historical evolution and international comparisons (No. w10931). National Bureau of Economic Research.

[55] Youderian, X. (2013). Parental Time Investment in Early Education and Endogenous Labor Supply. Mimeo.

Chapter 3

Collective Memories as a Driver of Discrimination

3.1 Introduction

Discrimination against certain social groups over long time periods has been a historical feature of many societies. For instance, in the United States (US) discrimination in the form of slavery officially ended in 1865 after more than two centuries, though racial segregation was maintained in the form of Jim Crow laws until 1965.[1] Starting with the civil rights movements in the early 1960's, one has seen significant advances in the rights of blacks with survey responses also indicating improved racial tolerance. However, today the black population still trails whites in terms of a variety of socio-economic indicators. In India caste, which is inherited by birth, was a marker for social discrimination for centuries. At independence in 1947 the practice of untouchability was made illegal and affirmative action was enshrined in the constitution for the disadvantaged sections. However, the lower castes continue to trail significantly behind the other sections of society on most socio-economic indicators of interest. What contributes to the gap between groups that faced discrimination over long time periods and those that did not? In what outcomes and why might we observe persistent gaps? Could discrimination persist due to more subtle channels than the traditionally assumed chan-

[1] Note that slavery had existed in colonial America since the 17th century but the US as an independent nation state came into existence only in 1774 and slavery ended 91 years hence.

nels of taste based discrimination, statistical discrimination, and discriminatory social norms?

Following Ramachandran and Rauh (2013), we extend their mechanism as to how discrimination can persist by including various levels of taste discriminators and suggest a range of stylized facts the model could explain. They introduce a channel of discrimination, where even under perfect observability of individual ability, the absence of discriminatory social norms, and when taste for discrimination has already died out, to discriminate can be the optimal response. The theoretical mechanism put forth rests on the existence of beliefs about discrimination by others in society, and on distinguishing between activities characterized by the need for *interlinkages* versus *no* need for interlinkages. In the model, activities with interlinkages require coordinated actions. If an individual decides to establish interlinkages she requires the input of two principals to form a productive unit. The success and return for all three, the individual and the two principals, is contingent on the participation of all three in the venture. The coordination failure results from the belief that somebody else might discriminate and the subsequent refusal to participate in the venture, which imposes losses due to the complementarity of inputs in the production process.

Activities involving interlinkages are defined as those that involve the input of more than one individual for the production process to be carried out. The classic example would be the case of entrepreneurs who need to establish multiple interlinkages (productive relations) to be able to start and operate a venture (Basu 2010). In the theoretical model, individuals choose between entering activities which require establishing productive relations and those that do not. Individuals intending to enter activities involving interlinkages are randomly matched with a pair of "principals", for instance a lender and a distributor, with whom they need to establish interlinkages to form a productive unit. The individual cannot produce without capital and cannot sell without a distributor. In case one of the principals agrees to participate and the other does not, the investment of the first principal is held up and imposes a fixed cost. We show how in the presence of beliefs about discrimination against a certain group, principals without a taste for discrimination in equilibrium also discriminate against that group. To clarify the mechanism, picture the following situation: Both the lender and the distributor have no preference for discrimination against the individual and know she has the requisite ability to be an entrepreneur. However, the lender (distributor) does not know whether

the distributor (lender) is a discriminator because historically principals had been discriminating against individuals of her type. He has a belief about the presence of taste discriminators, which he has been updating through Bayes rule based on past observations. If the lender believes with a sufficiently high probability that the distributor has a preference for discrimination, he will reject the loan application. Notice that if the distributor signs the contract to distribute the goods but the individual has obtained no loan to produce, as a consequence the distributor loses out, because by setting up the productive relation he has foregone the chance for his alternative investment. Therefore, in the future the distributor will account for the possibility of the individual being matched with a taste discriminator. The individual also faces a cost because she invested time and effort to become self-employed but did not manage to do so. Consequently, people from her type might also refrain from attempting to become self-employed. This leads us to a persistent equilibrium where able people are not becoming self-employed due to past discrimination and the resulting coordination failure based on beliefs, which leaves everybody worse off.[2]

The model predicts lower participation rates and higher cost of establishing interlinkages for the discriminated group relative to the non-discriminated group in equilibrium. We show that this result is robust to allowing for communication between the two principals as long as there are no reputation effects. The model also establishes conditions under which the steady state equilibrium is characterized by the existence of discrimination due to beliefs about the existence of taste discriminators, although there are no taste discriminators left in society. The discriminatory steady state is characterized by lower participation and payoffs for the discriminated group, leading to an overall welfare loss for society. The persistence of beliefs regarding discrimination in the steady state are rationalized by presenting evidence that these can be interpreted as intergenerational transmission of beliefs as *collective memory*, consistent with utility maximizing or cultural trait preserving strategies.

The theoretical framework identifies occupational choice, such as self-employment, as markets characterized by interlinkages, making it a suitable candidate for empirical scrutiny. We examine the market for self-employment of blacks and whites in the US. Using data from the General Social Survey (GSS) from the years 1972-2012, we create

[2]The example of the distributor believing the lender will discriminate is analog and would exhibit the same outcome.

proxies for beliefs about and tastes for discrimination against blacks for every region and year.

The time trends of taste for discrimination and beliefs about discrimination from the GSS and the self-employment rates for blacks and whites from the Current Population Survey (CPS) for the time period 1972-2012 are shown in Figure 3.1.[3] Taste for discrimination against blacks linearly declines over the period, whereas beliefs about discrimination against blacks as well as the gap between the self-employment rates for blacks and whites remain remarkably constant over the same time period. Figure 3.1 captures the mechanism and the role of *sticky or unchanging* beliefs highlighted by the theoretical model in a snapshot. The unchanging beliefs perfectly correspond to the invariant gap in self-employment rates over the period analyzed, as predicted by the theoretical framework. Using a logit model, we find that our proxies for beliefs about presence of discrimination to be a consistently significant factor in explaining lower levels of self-employment among blacks in the US. The estimate suggests that reducing the share of people who believe that discrimination exists by 10% points would increase self-employment probability of an average black male from 7.3% to 9.3%, an increase of 28%. The results are robust to the inclusion of year and region fixed effects and a variety of controls such as age, years of schooling, parental income at age 16, and whether the father was self-employed.

The literature of the economics of discrimination was pioneered by the seminal work of Becker (1957, 1971). In the setting envisaged, employers hold a taste for discrimination, such that working with members of a particular group imposes a cost on them, and hence these workers have to compensate the employer by either being more productive or accepting lower wages. The class of models of statistical discrimination (Phelps 1972; Arrow 1973; Aigner and Cain 1977; Lundberg and Startz 1983, 1998; Coate and Loury 1993; Rosen 1997) and categorical thinking (Fryer and Jackson 2008) rely on the imperfect observability of worker productivity. In absence of complete information employers base their decision on easily observable characteristics, such as race or gender, to infer the expected productivity of the worker. Mailath, Samuelson, and Shaked (2000) present a model of endogenous discrimination arising from the search decision of firms. The asymmetric discriminatory equilibrium is supported due to the belief that there are more skilled available workers of a particular type, which are borne

[3] See Section 3.3.1 for how the measures of taste and beliefs about discrimination are constructed.

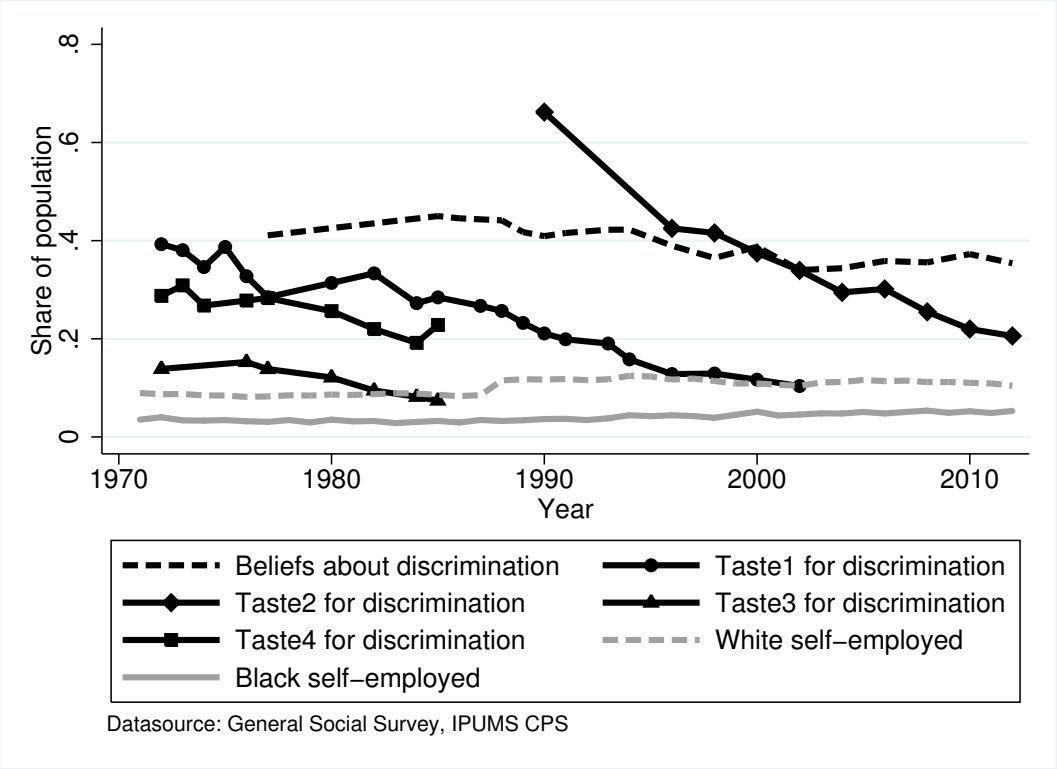


Figure 3.1: Self-employment rates by race and beliefs and taste regarding discrimination in the US

out in equilibrium. The third class of models is that of Akerlof (1976, 1985), Kandori (1992), and Peski and Szentes (2013), where not following the established norm of discrimination against certain groups might result in imposition of social sanctions which cause economic losses, and hence make discrimination a rational response. In contrast to the existing literature, in our setting there is no difference in the distribution of ability between the two groups. It is identical ex-ante and ex-post, there is perfect observability of ability, and no social norms to discriminate. Moreover, the nature of the coordination failure highlighted does not allow for a single principal who does not discriminate to reap the unrealized profits, a possibility traditionally assumed by Becker (1971), therefore providing a theoretical rational as to why discrimination can persist. To our knowledge, we are also the first to provide empirical evidence concerning the channel of discrimination presented theoretically.

Our theoretical model can reconcile many of the recent findings of the empirical

and behavioral studies on discrimination. Alesina, Mistrulli, and Lotti (2013) find that banks in Italy charge self-employed women more than self-employed men for credit. They find that characteristics such as riskiness, type of business, or differential bank choice cannot explain their result. They also find that the effect is not restricted to any particular geographical region and taste based indicators of discrimination cannot explain the pattern. As women businesses need to establish interlinkages, beliefs of banks that potential productive *male* links might discriminate against women, might result in banks discriminating against women too. Consistent with our theoretical model the authors find that that banks discriminate more against women in sectors, where men dominate, and can be interpreted as being more likely to be matched with a discriminatory male link.

The mechanism put forth is also a plausible explanation for features highlighted in data for the market for self-employment in India and Sweden. In India the Schedule Castes and Schedule tribes, both historically discriminated groups in India, are relatively more underrepresented in urban rather than rural areas in terms of non-farm enterprise ownership, even though taste based discrimination is higher in rural areas (Iyer, Khanna, and Varshney 2011). Observe that coordination failures in urban markets are more likely as they are anonymous, so even if taste for discrimination is higher in the rural than urban settings, it could well be the case that the coordination failures in urban areas outweighs the taste for discrimination effect in rural areas, leading to the outcome observed in the data. In Sweden, a country where women's labor force participation rate is very high and according to the wave of 2005-2007 of the World Values Survey only 0.4% of the male population strongly agree that men make better business executives than women, has among the lowest level of self-employment for women in the EU. On the one hand, agreeing that men make better business executives could be interpreted as taste based discrimination, which is very low in Sweden. On the other hand, 43% of the Swedish sample mention "Discrimination against women and girls" as one of the two most pressing problems facing the country, which we interpret as beliefs about about discrimination. Moreover, Swedish males have the lowest taste based bias against women across all 39 countries in the sample, while beliefs about discrimination are the second highest. Our model explains how these stylised facts could lead coordination failures driving the gap in self-employment rates.

Daskalova (2013) documents in a lab experiment that people who do not discrimi-

nate when making decisions individually, discriminate while making joint decisions due to beliefs about what their co-decision maker will do. Albrecht et al. (2013) find that in the lab individuals are conservative in updating their beliefs, which points to another channel through which beliefs regarding discrimination might become sticky over time and be an important determinant of outcomes for the discriminated group.

Our model can also reconcile why some groups are overrepresented in markets that require strategic complementarities. The dominance of particular ethnic groups in certain professions (Greif 1989, 1993; Banerjee and Munshi 2004) could be seen as ethnic enclaves which help secure complementary support from other individuals and overcome coordination failures. In order to avoid discrimination in the open market minority groups might turn to other ethnic group members which are concentrated in few professions. The benefit, however, would have to be weighed against the restriction on occupational choice that might arise due to ethnic enclaves being effective gate keepers to certain professions. A Pakistani immigrant arriving in Barcelona and wanting to become self-employed might not have have a strong preference to run a small convenience store, but given that there exists an infrastructure of Pakistani immigrants that can provide complementary support, while attempting interlinkages with domestic residents potentially exposes the immigrant to discrimination, expected returns based on beliefs might be highest by setting up a convenience store. Bare in mind that again the actual presence of taste based discriminators is not necessary, the belief is sufficient. Card, Mas, and Rothstein (2008) assume that when black people move into a neighborhood, white neighbors with a distaste for blacks will change neighborhoods. People without a distaste for black neighbors will also sell their property and move if they anticipate a drop in property prices. Why might someone fear a drop a property prices? If he believes that amongst his neighbors some have a taste for discrimination and will move away. We show that the presence of neighbors with a distaste for black neighbors is not required to trigger the segregating dynamics, the belief is sufficient, hence providing an alternative explanation for the phenomenon of racial tipping points in the US.

The rest of the paper is organized as follows: Section 3.2 presents the theoretical model, Section 3.3 provides empirical support of our theory, Section 3.4 evaluates policy tools and extends applications to other fields, and Section 3.5 concludes.

3.2 The model

In the following the basic model, which extends the one of Ramachandran and Rauh (2013), is specified. Then we characterize the solution of the static game with and without discrimination, followed by the dynamic game and the belief updating process.

3.2.1 The basic framework

The society consists of individuals i of two types $s_i \in \{A,B\}$. The types A and B form social groups based on visible characteristics which do not influence performance (e.g., race, gender). Individuals of type A and B belong to the finite, large sets $\mathcal{A}$ and $\mathcal{B}$, respectively.[4] The individuals have an ability a_i, where a is distributed uniformly over $[0,1]$. Ability $a_i \in [0,1]$ reflects productive capacity and is perfectly observable to all. For sake of simplicity we are dropping the index i in what follows.

Individuals can opt to engage in one of the two possible kinds of productive activities in the economy ($L \in \{0,1\}$) - those that involve establishing *interlinkages* ($L = 1$) with other agents, who are referred to as "principals", and those that do *not* involve establishing interlinkages($L = 0$). In case the individual i of type s decides to engage in an activity that does not involve establishing interlinkages with other principals in the economy, she earns a net income on her activity equal to the level of her productivity, which is given by:

$$W = G_{NL}(a) = a, \tag{3.1}$$

where G_{NL} is the production function of activities not involving interlinkages. On the other hand individuals have the option of engaging in the activity which involves establishing interlinkages with other principals, and earn a gross income equal to:

$$W = G_L(a,C) \text{ where } \frac{dG_L}{da} \geq 0\,, \frac{dG_L}{dC} \geq 0 \text{ and } \frac{dG_L^2}{dCda} \geq 0, \tag{3.2}$$

where G_L is the production function for activities involving interlinkages. The above production function G_L captures the notion of *interlinkages*. Activities requiring interlinkages imply that the gross income from this activity depends upon not only on the

[4] The assumption of large sets is to ensure that any single individual does not have any market power and collusion cannot take place.

individual's own ability a, but also on C. The component C captures the interlinkages or productive relations individual i is able to establish with the "principals" in the economy, which are necessary for the production process. Interlinkages refer to the fact that the production in such activities is a joint process and requires input from multiple sources.

A simple example to fix ideas for activities that are characterized by the need to establish interlinkages, is the market for self-employment. For example, the entrepreneur might require capital in the form of a loan from a bank (lender) to produce the goods and also may need to have an agreement with a distributor, who will be willing to distribute his goods. The example with the requirement of productive relations with two principals, a distributor and a lender, is only for illustrative purposes and could be extended to n-players or include any other contact necessary to setup a successful enterprise (e.g., supplier, landlord to rent office). Both components (loan, distribution route) are necessary and cannot be substituted through ability.

In order to keep our analysis tractable, we assume that any individual wanting to engage in the activity that involves establishing interlinkages needs to establish productive relations with two principals $P \in \{p_1, p_2\}$ in the economy. In light of the above, the component C in (3.2) is defined as:

$$C = Y(c_{p_1}, c_{p_2}) \; where \; \frac{d^2 Y(c_{p_1}, c_{p_2})}{dc_{p_1} dc_{p_2}} \geq 0, \tag{3.3}$$

where c_{p_1} and c_{p_2} refer to the decisions by the principal $P \in \{p_1, p_2\}$ of whether to establish productive relations, which are given by $c_P \in \{0,1\}$. We only allow for pure strategies, such that the two principals decide whether to establish the relation $(c_P = 1)$ or not $(c_P = 0)$. Moreover, it is assumed that in case the principals decide to establish a productive relationship with the individual, they need to make an investment, which is normalized to unity. The incentive for engaging in activities that involve establishing interlinkages arises as complementary investment by the principals results in a boost in productivity, captured by the factor $\lambda(>1)$ in the production function G_L. We can now represent the productivity or gross income for individual's engaging in activities involving interlinkages by:

$$G_L(a, C) = \lambda c_{p_1} c_{p_2} a. \tag{3.4}$$

The above functional form exhibits an extreme form of complementarity in the actions of the principals P, implying:

$$G_L(a, Y(c_{p_1}, 0)) = G_L(a, Y(0, c_{p_2})) = 0. \tag{3.5}$$

The intuition is that establishing a relationship with both principals $P \in \{p_1, p_2\}$ is required for the individual to produce and be successful.[5]

DEFINITION 3. *Throughout the paper we will refer to those who decide between entering activities involving interlinkages or not as individuals of type $s \in \{A, B\}$, and to the complementary agents necessary for the production process $P \in \{p_1, p_2\}$, as principals.*

3.2.2 The static game

The individuals i decide whether they want to engage in activity that involve no interlinkages or whether to enter into activities involving interlinkages by trying to establish productive relations with the principals. As already noted, individuals who decide to enter activities involving no linkages earn a net income equal to their ability a.

The individuals wanting to enter activities involving interlinkages are randomly matched with a pair of principals, p_1 and p_2, in the market every period to try to establish productive relations. The principals have an outside opportunity of a risk free investment yielding interest r per unit invested. To establish a productive relationship, and in return for the investment in their activity by the principals, they offer an amount σ_P to each of the principals as repayment for the investment. In case this offer is rejected by any one of the principals, the attempt to enter the activity with interlinkages fails and the individual i faces a fixed cost δ from the effort exerted (observe that (3.4) implies production is contingent on the participation of both principals). She then enters the activity not involving interlinkages and earns a net income of $(a - \delta)$.

The offer of σ_P made by the individual to the principal arises as a solution of a Nash bargaining process over the surplus generated from the activity involving interlinkages

[5] For ease of exposition we are assuming perfect complementarity. However, the results presented later qualitatively hold for weaker forms of complementarity as well.

compared to the alternative involving no interlinkages. The bargaining takes place between only one principal and the individual, without communication between p_1 and p_2.[6] In light of the production function given by (3.4), the outcome is dependent on what decision the other principal (henceforth denoted by $-P$) makes. Hence, P has a belief concerning the likelihood of the other principal accepting the offer as well.[7] If P accepts an offer which the other principal rejects, then he is not able to obtain r from the risk free investment in the given period due to his capital being bound and not yielding any interest.

Due to random matching and the absence of perfect competition, the interaction between the individual and the principal is characterized by a monopoly versus monopsony, commonly referred to as bilateral monopoly. Both have one shot at earning a surplus compared to their outside option. Assuming equal bargaining power and linear utility functions in payoffs, the bargaining solution, resulting in offer σ_P, is characterized by the disagreement point $d = (d_i, d_P) = (a - \delta, 1 + r)$ and the maximization of:

$$(\sigma_P - 1 - r)(\lambda a - \sigma_P - \sigma_{-P} - a + \delta) \Rightarrow \sigma_P = \frac{(\lambda - 1)a - \sigma_{-P} + r + \delta + 1}{2}. \quad (3.6)$$

Now assuming i makes the same offer to the two principals, such that $\sigma_P = \sigma_{-P}$, the Nash bargaining solution is given by:

$$\sigma^N(a) = \frac{(\lambda - 1)a + r + \delta + 1}{3}. \quad (3.7)$$

This Nash bargaining solution is a function of a, which we will denote as $\sigma^N(a)$. In order for P to accept this offer, we require the amount she expects to be repaid to be at least what she can earn through the risk free investment, i.e. $(1 + r)$.

Let the state variables $x = (a, s)$ be given by the ability a and type s of an individual i. The utility of the individual is $V_i(x)$, which will either be given by the utility of not establishing interlinkages $V_i^{NL}(x)$, or by the expected utility of attempting to establish

[6] Communication could be allowed between the two principals and would not change our results if we assume costly communication or no reputation effects. Without reputation effects moral hazard problems would arise as there would be no gain from admitting when one was not willing to establish the productive relation.

[7] We disregard higher order beliefs, even though they would additionally speed up the contagion-effect.

interlinkages $E_C[V_i^L(C,x)]$. Formally the static game for the individual can be written as

$$V_i(x) = \max_{L\in\{0,1\},\sigma\in\mathbf{R}^+} \{V_i^{NL}(x), E_C[V_i^L(C,x)]\} \tag{3.8}$$

subject to

$$\begin{cases} E_C[V_i^L(C,x)] = \lambda C(a-2\sigma) + (1-C)(a-\delta) & \text{if } L=1 \\ V_i^{NL}(x) = a & \text{if } L=0, \end{cases} \tag{3.9}$$

$$C = c_{p_1} c_{p_2}, \tag{3.10}$$

and

$$\sigma \geq \sigma^N(a). \tag{3.11}$$

For the principal the static game concerning her decision $c_P \in \{0,1\}$ is represented by

$$V_P(x,\sigma) = \max_{c_P\in\{0,1\}} \{V_P^{c0}, E_{c_{-P}}[V_P^{c1}(c_{-P},x,\sigma)]\} \tag{3.12}$$

subject to

$$\begin{cases} E_{c_{-P}}[V_P^{c1}(c_{-P},x,\sigma)] = c_{-P}\sigma + (1-c_{-P}) & \text{if } c_P = 1 \\ V_P^{c0} = 1+r & \text{if } c_P = 0. \end{cases} \tag{3.13}$$

Given $\sigma^N(a)$ from (3.7) it is now possible to calculate the lowest ability individual, denoted by a', who could possibly offer the principal a share dominating the risk free investment, and is given by:

$$\sigma^N(a) = \frac{\lambda a + r + \delta + 1}{3} \geq 1 + r \Rightarrow a' = \frac{2(1+r)-\delta}{\lambda - 1}. \tag{3.14}$$

Now looking at the participation constraint of the individual, we require that $\lambda a - 2\sigma^N(a) \geq a$, such that

$$\Rightarrow a^* = \frac{2(1+r+\delta)}{\lambda - 1}. \tag{3.15}$$

Since $a^* > a'$ only individuals with $a \geq a^*$ will intend to enter activities involving interlinkages and, as long as there is no taste for or belief about discrimination, will be

accepted.[8]

DEFINITION 4. *Let a^* be the individual with the lowest ability, from both groups $s \in \{A,B\}$, who wants to enter activities involving interlinkages in the absence of discrimination.*

3.2.3 Discrimination in the static framework

The static game outlines the decision making process and identifies the lowest ability type a^* from both groups $s \in (A,B)$ who enter into activities involving interlinkages.

Now assume that individuals believe that a subset of the principals have a taste for discrimination.[9] The taste for discrimination is in the Beckerian sense and can be understood as a cost/disutility which the individuals believe the principals with taste for discrimination face when they decide to establish a productive relation with a B-type individual in society.

The belief regarding the presence of taste discriminators implies that the probability of discrimination occurring has to be taken into account while deciding on the optimal course of action. Due to the random matching the probability of meeting a principal with a taste for discrimination is equal to what is believed to be the share of total taste discriminators in society. As the share is not observable, decisions are conditioned on beliefs based on a common prior about the share of discriminators amongst the principals and individuals. In particular, we assume that indiviudals and the principals believe that φ_h proportion of principals have a high taste for discrimination with taste equal to $b_h > 0$ and φ_l proportion of principals have a low taste for discrimination with taste equal to b_l $(0 < b_l < b_h)$.[10] A simplified extensive form game tree assuming only a single level of taste discriminators can be found in Figure 3.2 in the Appendix.

From the static model, without discrimination, we know that the minimum offer that will be accepted by a principal without a taste for discrimination is $\sigma^N(a) \geq (1+r)$, i.e. a return greater than the risk free investment. Observe that any offer $\sigma^N(a) \geq$

[8] Similar to Lucas (1978) in our model only the most able want to enter the activity involving interlinkages such as entrepreneurship.

[9] How this belief can arise and sustain itself is discussed in the next subsections.

[10] In Ramachandran and Rauh (2013) only a single level of taste discriminators is considered. By extending the model to two levels, high and low, we show that it can be extended to n levels, thereby mimicking a continuous distribution of levels of discriminatory taste when n is sufficiently large.

$(1+r+b_h)$ is always accepted, even by principals with a *high* taste for discrimination, as it compensates for the taste for discrimination. This implies that offers in the range $(1+r+b_l)<\sigma^N(a)<(1+r+b_h)$ could be subject to discrimination only from the principals with the high taste, whereas offers in the range $(1+r)<\sigma^N(a)<(1+r+b_l)$ could be discriminated upon by both high and low taste discriminators. Individuals and principals are expected payoff maximizers. In light of the above, discrimination is defined in the following manner:

DEFINITION 5. *An offer of $\sigma^N(a) > 1+r$, which is rejected by any principal, is defined as a case of discrimination.*

The individuals of the *B*-type now take the probability of meeting a discriminator in the market into account while deciding on their optimal course of action.

DEFINITION 6. *Let us denote by a_h^* the lowest ability type such that $\sigma^N(a_h^*) = 1+r+b_h$, by a_l^* the lowest ability type such that $\sigma^N(a_l^*) = 1+r+b_l$, and by a_b the ability of an individual who is indifferent between offering $(1+r+b_h)$ to each principal and entering an activity not involving interlinkages, i.e. $V_i^{NL}(x) = V_i^{NL}(a_b,B) = a_b = E_C[V_i^L(C,(x)] = E_C[V_i^L(C,(a_b,B))]$.*

All individuals with ability $a \geq a_h^*$ offer $\sigma^N(a) \geq 1+r+b_h$ and will never be discriminated against. The individuals in the ability range $a^* \leq a < a_l^*$ face potential discrimination from both high and low taste discriminators. The individuals in the ability range $a^* \leq a < a_l^*$ compare the expected payoff from each of the four actions available to them, given by:

$$E[V_i(x)] = \begin{cases} (1-\varphi_l+\varphi_h)^2(\lambda a - 2\sigma^N(a)) + (1-(1-(\varphi_l+\varphi_h)^2)(a-\delta) & \text{if } (L=1, \sigma=\sigma^N(a)) \\ (1-\varphi_h)^2(\lambda a - 2(1+r+b_l)) + (1-(1-\varphi_h)^2)(a-\delta) & \text{if } (L=1, \sigma=1+r+b_l) \\ \lambda a - 2(1+r+b_h) & \text{if } (L=1, \sigma=1+r+b_h) \\ a & \text{if } L=0. \end{cases}$$

From applying for the activity involving interlinkages $(L=1)$ and offering the Nash bargaining solution the expected payoff depends on the belief about the share of high (φ_h) and low (φ_h) taste discriminators. By offering each principal $(1+r+b_l)$ the individual escapes the low taste discriminators but still potentially faces discrimination from the high taste discriminators. The strategies of offering $(1+r+b_h)$ and escaping

potential discrimination (from both types) as well as from entering the activity involving no interlinkages provide certain payoffs. The individuals choose whichever profile (L,σ) maximizes their expected payoff.

On the other hand individuals in the ability range $a_l^* \leq a < a_h^*$ face potential discrimination from only high taste discriminators and compare the expected payoff from each of the three actions available to them, given by:

$$E[V_i(x)] = \begin{cases} (1-\varphi_h)^2(\lambda a - 2\sigma^N(a)) + (1-(1-\varphi_h)^2)(a-\delta) & \text{if } (L=1, \sigma = \sigma^N(a)) \\ \lambda a - 2(1+r+b_h) & \text{if } (L=1, \sigma = 1+r+b_h) \\ a & \text{if } L=0 \end{cases}$$

From applying for the activity involving interlinkages $(L=1)$ by offering the Nash bargaining solution, the expected payoff depends on the belief φ_h about the share of high taste discriminators. From offering each principal $(1+r+b_h)$ and escaping potential discrimination as well as from entering the activity involving no interlinkages the payoffs are certain.

The principals with no taste for discrimination, when facing an individual of the B-type in the ability range $a^* \leq a < a_l^*$ that offers $\sigma^N(a)$, compare their expected payoff from accepting their offer to the risk free investment:

$$E[V_P(x)] = \begin{cases} (1-(\varphi_h+\varphi_l)\sigma^N(a) + \varphi_h + \varphi_l & \text{if } c_P = 1 \\ 1+r & \text{if } c_P = 0, \end{cases}$$

and when facing an individual of the B-type in the ability range $a_l^* \leq a < a_h^*$ that offers $\sigma^N(a)$, compare their expected payoff from accepting their offer to the risk free investment:

$$E[V_P(x)] = \begin{cases} (1-\varphi_h)\sigma^N(a) + \varphi_h & \text{if } c_P = 1 \\ 1+r & \text{if } c_P = 0, \end{cases}$$

The principals with a taste for discrimination equal to b_l reject the Nash bargaining solution when facing an individual in the ability range $a^* \leq a < a_l^*$. When they face an individual with ability in the range $a_l^* \leq a \leq a_h^*$ they calculate their expected payoff

from establishing a productive relationship or their risk free option, i.e.

$$E[V_P(x)] = \begin{cases} (1-\varphi_h)\sigma^N(a) + \varphi_h - b_l & \text{if } c_P = 1 \\ 1+r & \text{if } c_P = 0. \end{cases}$$

Finally the principals with a taste for discrimination equal to b_h always reject any $\sigma^N(a)$ offered by $a < a_h^*$.

Now we have seen how discrimination has spilled over to those principals without a taste for discrimination, as they now account for the probability of the individual being matched with a principal who is a taste discriminator. This occurs, as the production process requires the inputs of the individual and the two principals. In case the principal P accepts and the counterpart $-P$ rejects the individual's offer, the principal P will lose his risk free return r. Thus, the principal P, due to the belief that the individual might face discrimination from principal $-P$, also ends up discriminating against the individual of type-B.

Observe that the taste for discrimination against B-type individuals does not affect the A-types in the market. The A-types play the identical game as in the static framework without discrimination. All individuals of the A-type with $a \geq a^*$ still offer $\sigma^N(a)$, as in the static game without beliefs about discrimination, and are accepted.

For the B-types, first observe that for any given a,

$$\lambda a - 2\sigma^N(a) \geq (1-(\varphi_h+\varphi_l))^2(\lambda a - 2\sigma^N(a)) + (1-(1-(\varphi_h+\varphi_l))^2)(a-\delta), \quad (3.16)$$

such that the expected payoff of establishing interlinkages, ceteris paribus, is always greater for the individual if either or both of φ_h or φ_l are greater than zero . The lowest ability individual, who applies to enter activities with interlinkages in the absence of beliefs about discrimination, is given by a^*, which implies:

$$\begin{aligned} \lambda a^* - 2\sigma^N(a^*) = a^* > \lambda a - 2\sigma^N(a) \geq \\ (1-(\varphi_h+\varphi_l))^2(\lambda a - 2\sigma^N(a)) + (1-(1-(\varphi_h+\varphi_l))^2)(a-\delta). \end{aligned} \quad (3.17)$$

Equation (3.17) and the fact that the payoff from offering the Nash bargaining solution is strictly increasing in a implies some value of $a > a^*$ will solve the equation such that:

$$\lambda a - 2\sigma^N(a) \geq (1-(\varphi_h+\varphi_l))^2(\lambda a - 2\sigma^N(a)) + (1-(1-(\varphi_h+\varphi_l))^2)(a-\delta) = a. \tag{3.18}$$

DEFINITION 7. *Let us denote by a_{db} the lowest ability type from group B who wants to apply for interlinkages in the presence of beliefs about discrimination.*

In the equilibrium of the static game with beliefs about discrimination, all A-types with $a \geq a^*$ apply to enter activities involving interlinkages, offer the Nash bargaining solution, and are successful. For the B-types, if $a_{db} < a_b$, only individuals with $a \geq a_{db} > a^*$ apply to enter activities involving interlinkages, offer the Nash bargaining solution, and are successful.[11] On the other hand, if $a_{db} > a_b$, individuals with $a \geq a_b > a^*$ apply to enter activities involving interlinkages and are successful. The individuals $a > a_{db}$ offer the Nash bargaining solution same as their A-type counterparts. However, individuals in the ability range $a_b \leq a < a_{db}$ offer $(1+r+b_h) > \sigma^N(a)$, an offer strictly greater than the ones offered by the A-types in the same ability range.

Beliefs about discrimination, hence, have two potential effects: (i) if $a_{db} < a_b$, individuals of B-type in the range $a^* \leq a < a_{db}$, enter activities involving no interlinkages whereas A-types of the same ability enter activities involving interlinkages and enjoy the associated surplus, and (ii) if $a_{db} > a_b$, individuals of B-type in the range $a^* \leq a < a_b$, enter activities involving no interlinkages, whereas A-types of the same ability enter activities involving interlinkages and enjoy the associated surplus. On the other hand B-types in the range $a_b \leq a < a_{db}$ enter activities involving interlinkages, but pay a higher price to the principals than the A-types in the equivalent ability range.

3.2.4 The dynamic game and the belief updating process

In the previous subsection we assumed that there was a certain belief regarding the probability of meeting a taste discriminator in society. In light of this we characterized the optimal decisions in a static framework.

We now extend the framework and allow for the game to be repeated every period. We assume that the taste for discrimination arises due to a shock to the taste of a subset

[11] Recall a_b is defined as the ability type from group B who is indifferent between offering $(1+r+b_h)$ to each principal and entering an activity not involving interlinkages.

of principals in society at time t_0. It is assumed that $\pi_{0,h}$ proportion of principals develop a high taste for discrimination with the taste equal to $b_h > 0$ and $\pi_{0,l}$ proportion of principals develop a low taste for discrimination with the taste equal to b_l $(0 < b_l < b_h)$, against establishing a productive relation with B-type individuals. The origins of the shock which result in creating a taste for discrimination among a subset of the principals is not the focus of the paper and can arise due to various reasons. An example could be the incidents of September 11^{th} 2001, which resulted in discriminatory actions against Muslims in the US (e.g., Kaushal, Kaestner, and Reimers 2007).[12]

In the dynamic game, a principal P exits the market with exogenous probability ω every period. The probability ω is not known to anybody in society. A principal without a taste for discrimination replaces the exiting principal, such that at some point no principals with a taste for discrimination will be left. Therefore, if we define the share of principals with a high and low taste for discrimination in period $t = 0$ to be $\pi_{0,j}$ for $j = l, h$, the probability that an individual i is matched with a principal with high or low taste for discrimination in period T is $\pi_{0,j}(1-\omega)^T$.

Since neither $\pi_{t,j}$ nor ω are common knowledge, decisions are conditioned on beliefs about the share of discriminators amongst the principals, which are updated through observations of discrimination in the market.[13] We assume that the event which creates a taste for discrimination also creates a common prior among individuals and principals.[14] The common prior is assumed to have a distribution denoted by η_0^j, capturing the probability of meeting a principal with a taste for discrimination equal to b_j. The common prior η_0^j is modeled as having a beta distribution. More specifically, it is assumed that the individuals and principals believe that the share of principals with taste b_j has a beta distribution with parameters α_0^j and β_0^j. Moreover, we denote the density of the distribution η_0^j by θ^j. The beta distribution captures the belief regarding the probability of meeting a principal with a taste for discrimination through its expected

[12]The assumption that shocks do not work the other way, i.e. people immediately forget the past existence of discrimination due to sudden events today, is justified by the literature on trust and beliefs which shows how persistent past beliefs are in shaping today's actions. See section 3.2.7 for discussion and references.

[13]The definition of a case of discrimination is provided in Definition 5.

[14]The results do not hinge on the prior. Given that in our envisaged setting taste for discrimination is assumed to exist for a long time period we do not want to emphasize the importance of the prior. In an earlier version of this paper, using a frequentist approach based on case based decision theory (Schmeidler and Gilboa 2001), we actually show that no prior is required for the discriminatory equilibrium to hold once taste for discrimination has died out.

value, or the mean of the distribution.

ASSUMPTION 1. *The probability parameter capturing the share of lenders with a taste for discrimination equal to b_j in period t_0 is given by $\theta(\eta_0^j) \sim beta\ (\alpha_0^j, \beta_0^j)$.*

The above distribution implies that the density function associated with facing a discriminator with taste b_j in period t is given by:

$$\theta(\eta_t^j) = \frac{(\eta_t^j)^{\alpha_t^j - 1}(1-\eta_t^j)^{\beta_t^j - 1}}{beta(\alpha_t^j, \beta_t^j)} = \frac{(\alpha_t^j + \beta_t^j - 1)!}{(\alpha_t^j - 1)!(\beta_t^j - 1)!}(\eta_t^j)^{\alpha_t^j - 1}(1-\eta_t^j)^{\beta_t^j - 1} \text{ for } j = h, l\ . \tag{3.19}$$

The beta distribution gives us a density on $[0,1]$, which captures the beliefs held by the individuals and principals regarding η_0^j. As the individuals and principals will need to decide on optimal actions based on their beliefs, and all individuals and principals are assumed to be risk neutral, the individuals and principals use the expected value of the distribution, which is given by $E(\eta_t^j) = \frac{\alpha_t^j}{\alpha_t^j + \beta_t^j}$. This is the point at which the density of the distribution takes its highest value.

The belief updating process of the principals and individuals is characterized by a standard Bayesian approach. Assume that in any period t, n_t individuals of the B-type applied to enter the market characterized by the need for interlinkages and k_t cases of discrimination are observed in the market. Out of the total of n_t cases, assume that n_{tl} made offers such that $1 + r < \sigma^N(a) < 1 + r + b_l$ and let n_{th} denote the total number of individuals who made offers such that $1 + r < \sigma^N(a) < 1 + r + b_h$.

Suppose that out of the total k_t cases of discrimination, k_t^h cases involved offers such that $\sigma^N(a) \geq 1 + r + b_l$, implying that these were offers which could have been only rejected by principals with taste of discrimination equal to b_h. Let k_t^u cases be such that $k_t^h + k_t^u = k_t$ and the offers associated with them be such that $1 + r < \sigma^N(a) < 1 + r + b_l$. This implies that the k_t^u cases of discrimination could have been either due to high or low type discriminators. In order to update their beliefs the individuals and principals allot a proportion of the k_t^u cases of discrimination to high type discriminators and the remaining to low type discriminators based on their period $t-1$ beliefs. More specifically, define k_t^{uh} to be the number of cases in k_t^u, which can be attributed to high type discriminators, and k_t^{ul} to be the number of cases in k_t^u, which can be attributed to

low type discriminators. Therefore,

$$k_t^{uh} = \frac{E(\eta_{t-1}^h)}{E(\eta_{t-1}^h)+E(\eta_{t-1}^l)} k_t^u \text{ and } k_t^{ul} = \frac{E(\eta_{t-1}^l)}{E(\eta_{t-1}^h)+E(\eta_{t-1}^l)} k_t^u. \tag{3.20}$$

The principals and individuals allot the cases of discrimination of which they do not know whether they are due to high or low type discriminators according to their relative expected shares.

ASSUMPTION 2. *It is assumed that all market transactions in terms of the offers made and rejected are common knowledge.*[15]

The above implies that the total number of cases pertaining to high type discriminators are equal to $k_t^h + k_t^{uh}$, while the total number of cases pertaining to low type discriminators is equal to k_t^{ul}. The posterior function for $\theta(\eta_{t-1}^j)$ is given by:

$$\theta(\eta_t^j | k_t^j + k_t^{uj}) \sim beta(\alpha_{t-1}^j + k_t^j + k_t^{uj}, \beta_{t-1}^j + n_{1j} - k_t^j - k_t^{uj}). \tag{3.21}$$

The above outlines the Bayesian belief updating procedure used by individuals to update their beliefs regarding the probability of meeting a high or low type principal with taste for discrimination. The posterior probability densities are given by $\theta(\eta_T^j | \sum_{t=1}^T (k_t^j + k_t^{uj})) \sim beta(\alpha_0^j + \sum_{t=1}^T (k_t^j + k_t^{uj}), \beta_0^j + \sum_{t=1}^T n_{tj} - \sum_{t=1}^T (k_t^j + k_t^{uj}))$ for $j = l, h$. The associated expected value or the point probability estimate used by the individuals and principals to make their optimal decision are given by:

$$E(\eta_T^j) = \frac{\alpha_0^j + \sum_{t=1}^T (k_t^j + k_t^{uj})}{\alpha_0^j + \beta_0^j + \sum_{t=1}^T n_{tj}}.$$

DEFINITION 8. *Let $\varphi_t^j = E(\eta_{t-1}^j)$ for $j = l, h$ be the probability individuals and principals assign to a being matched with a principal with a taste for discrimination equal to b_j.*

The decision-making rules of the individuals and principals imply that the probability of entering an activity involving interlinkages for a B-type individual in any period T

[15] We relax this assumption in Ramachandran and Rauh (2013) to allow the individuals and principals to observe only a subset of all the market transactions and show that the results remain qualitatively unchanged.

will depend upon his ability a, the actual share of taste discriminators $\pi_0^j(1-\omega)^T$, and his beliefs regarding the share of taste discriminators in society φ_T^j. We can thus express the probability of entering an activity involving interlinkages for a B-type individual as a function of the above three factors, i.e.

$$f(a, \pi_0^j(1-\omega)^T, \varphi_T^j). \tag{3.22}$$

It is easy to see that the probability of entering an activity involving interlinkages is increasing in ability and declining in the actual share and the belief regarding the proportion of taste discriminators in society, i.e. $f_1 > 0$, $f_2 < 0$, and $f_3 < 0$, where the subscripts refer to the first, second, and third argument of the function. In the following section we explicitly test for the predictions of our model using the above function f.

3.2.5 Characterization of the dynamic steady state equilibrium under no remaining taste for discrimination

The channel of discrimination that we put forth works on the premise that even once all principals with taste for discrimination have died out, to discriminate against members of group B may remain as the optimal action. In what follows, we address whether discrimination can exist, and if it can, under what conditions does it exist, for how long does it persist, and in what form does it manifest itself. Let us denote by T^* the first period in which no principals with taste for discrimination remain in the economy. The probability density function, given the beliefs and the Bayesian updating rule used, for meeting a principal with a taste for discrimination is given by:

$$beta(\alpha_0^j + \sum_{t=1}^{T^*}(k_t^j + k_t^{uj}), \beta_0^j + \sum_{t=1}^{T^*} n_{tj} - \sum_{t=1}^{T^*}(k_t^j + k_t^{uj})). \tag{3.23}$$

The associated probability point estimates for meeting a high and low taste discriminator is given by:

$$\varphi_{T^*}^h = \frac{\alpha_0^h + \sum_{t=1}^{T^*}(k_t^h + k_t^{uh})}{\alpha_0^h + \beta_0^h + \sum_{t=1}^{T^*} n_{th}} \text{ and } \varphi_{T^*}^l = \frac{\alpha_0^l + \sum_{t=1}^{T^*}(k_t^{ul})}{\alpha_0^l + \beta_0^l + \sum_{t=1}^{T^*} n_{tl}}. \tag{3.24}$$

It is clear that all $B-$type individuals with $a \geq a_h^*$, will offer their Nash bargaining solution and be accepted.[16] The form of discrimination and the length for which it will persist will depend on $\varphi_{T^*}^h$ and $\varphi_{T^*}^l$.

The point probability estimates in period T^* are a function of the initial beliefs (α_0^j, β_0^j), the actual shares of taste discriminators $(\pi_{0,h}, \pi_{0,l})$ and the rate at which principals with a taste for discrimination exit the market in every period ω. If we assume that the initial beliefs are a function of the actual share of taste discriminators, i.e. $\alpha_0^j(\pi_{0,h}, \pi_{0,l})$ and $\beta_0^j(\pi_{0,h}, \pi_{0,l})$, then we can write $\varphi_{T^*}^j = z(\pi_{0,h}, \pi_{0,l}, \omega)$ for $j = l, h$. In the proposition that follows we highlight the various forms in which discrimination manifests itself and persists in the multiple steady state equilibria depending on φ_{T^*} after no principals with a taste for discrimination are left.

PROPOSITION 1. *1. Let $\varphi_{T^*}^h$ be such that at time T^* no individual of the B type in the range $a_l^* \leq a \leq a_h^*$ prefers the Nash bargaining solution to entering the activity involving no interlinkages. In such a scenario discrimination manifest itself in two forms:*

(a) B-types are underrepresented relative to A-types in activities involving interlinkages at the lower tail of the ability distribution.

(b) B-types, in the middle ability ranges, pay a strictly higher fee to establish interlinkages, and hence earn strictly less than the A-types with equal ability.

2. Let $\varphi_{T^*}^h$ and $\varphi_{T^*}^l$ be such that some individual of B-type in the range $a^* \leq a \leq a_l^*$ strictly prefers the Nash bargaining solution to entering the activity involving no interlinkages. This implies that in the long run, discrimination will not persist. However, B-types are penalized in the form of lower participation rates relative to A-types for a finite duration before beliefs about discrimination disappear from society.

3. Let $\varphi_{T^*}^h$ and $\varphi_{T^*}^l$ be such that some individual of the B-type in the range $a_l^* < a < a_h^*$ strictly prefers the Nash bargaining solution to entering the activity involving no interlinkages and no individual in the range $a^* \leq a < a_l^*$ prefers the Nash bargaining solution to entering the activity involving no interlinkages. This implies in the long run that either:

[16] Recall $\sigma^N(a_h^*) = 1 + r + b_h$ and $\sigma^N(a)$ is increasing in a.

(a) *All individuals with ability $a > a_l^*$ enter activities involving interlinkages and pay $\sigma^N(a)$, whereas individuals in the ability range $a^* \leq a < a_l^*$ enter the activity involving no interlinkages.*

(b) *All individuals with ability $a \geq a^*$ become self-employed and discrimination will not persist in the long run.*

Proof. Proof in Appendix 1. ■

The intuition for the equilibrium outcome in case (1) where discrimination persists forever crucially depends on who is the lowest B-type deciding to apply for activity involving interlinkages when the last principal with taste for discrimination dies out. This occurs when all individuals of the B-type, whose Nash bargaining solution is not sufficiently high to compensate the taste for discrimination, i.e. $a < a_h^*$, decide to enter activities involving no interlinkages rather than seek to establish productive relations. Even if an offer, which could be subject to potential discrimination, were to be made, it would be rejected due to beliefs about discrimination being prohibitively high. Therefore, this equilibrium is even stable under the trembling hand. This in turn implies that beliefs remain frozen at the current level and hence all individuals with ability levels $a^* \leq a < a_b$ will always prefer entering activities involving no interlinkages.[17]

However, if when the last principal with taste for discrimination dies out, as in case (2), the lowest B-type who decides to enter an activity involving interlinkages is one whose Nash bargaining solution is not sufficient to compensate the taste for discrimination, i.e. $a < a_l^*$, implies discrimination will not persist in the long run. Now that all principals with taste for discrimination have died out and beliefs are not prohibitively high, all offers made by individuals seeking to form productive relations are accepted. As this includes individuals whose offers could have been subject to potential discrimination, but are not (as no taste discriminators are left), the next period beliefs about discrimination will be lower after beliefs have been updated. As in every period all offers are accepted, in the long run the belief about discrimination will converge to zero.[18]

[17] Recall that a_b is the individual who is indifferent between offering $(1+r+b_h)$ to each principal and entering an activity not involving interlinkages.

[18] Case (3) outlined in the proposition is an intermediate situation where only individuals facing discrimination from both high and low type discriminators are penalized in the long run. Refer to the proof in the Appendix for further details.

3.2.6 Welfare effects

How does discrimination affect welfare when there is no more taste for discrimination remaining in the economy? One the one hand, there is an efficiency loss due to productive individuals not entering the more remunerative activities involving interlinkages. On the other hand, discrimination gives rise to a redistribution effect. The efficiency loss not only affects the B-type individuals, who become involved in activities involving no interlinkages instead of establishing productive relations (which would be the social optimum), but also the principals, who lose out on opportunities of receiving offers yielding more than the return r on their risk free investment.

The wealth redistribution takes place because B-types offering $(1+r+b_h)$ are paying a higher price than the equivalent A-type, from which the principals are profiting, as they are receiving more than the Nash bargaining solution.[19] Assuming that the necessary and sufficient condition of proposition 1.1 hold and discrimination persists, we can quantify the deadweight loss of each period to the B-type as:

$$\sum_{a \geq a^*}^{a_b} [(\lambda a - 2\sigma^N(a)) - a], \tag{3.25}$$

while the deadweight loss to the principals is:

$$\sum_{a \geq a^*}^{a_b} 2[\sigma^N(a) - (1+r)]. \tag{3.26}$$

The redistribution from the B-types to the principals is:

$$\sum_{a \geq a_b}^{a_h^*} 2[(1+r+b_h) - \sigma^N(a)]. \tag{3.27}$$

PROPOSITION 2. *In an equilibrium as in proposition 1.1, on average both principals earn lower profits by discriminating.*

[19] This theoretical prediction is consistent with the finding of the US Department of Justice that Countrywide charged more than 200,000 black and Hispanic borrowers higher fees and interest rates than comparable whites with similar credit histories between 2004 and 2008 leading to the Bank of America paying a settlement of 335M$.

Proof. Proof in Appendix 1. ■

It is obvious that the B-type suffers from discrimination. But proposition 2 implies that even for the principals in the aggregate the efficiency loss due to the untapped source of talent dominates the higher prices some principals receive from those B-types that try to escape discrimination by making greater offers.

3.2.7 Persistence of beliefs as collective memories

The model presented above assumes that once the equilibrium set of beliefs have been established they will persist over time. The crucial question then arises as to how and why beliefs regarding the presence of discrimination might tend to persist? We interpret transmission of beliefs in our model as happening through intergenerational transmission of collective memory regarding discrimination.

The contemporary usage of the term collective memory can be traced back to Emile Durkheim (1859-1917), and his student Maurice Halbwachs (1877-1945), who published the seminal study titled *The social framework of memory* in 1925. The concept of memory has been constructed in the literature as to how the mind works in a society and how their operations are structured by social arrangements. Halbwachs argues: "It is in society that individuals normally acquire their memories. It is also in society that they may recall, recognize and localize their memories" (Halbwachs 1992, p. 38). Formulation of memories regarding the past are hence affected by transmission of cultural beliefs and norms in society.

Beliefs regarding discrimination can be seen to fullfill the two important criteria to be categorized as collective memories. First, events which influence collective memory are widely documented and recorded in these societies (Griffin and Bollen 2009). For the case of discrimination against blacks in the US or Dalits in India, these events have been widely recorded and recollected. Second, collective memories require a consensual view of the recollected past. The presence of affirmative action policies in countries like the US and India, serve as clear signal to the consensus among the policy makers and the public at large on the need to readdress previous wrongs.

Beliefs regarding discrimination being transmitted as collective memory through generations can also be rationalized by economic models of cultural transmission such as Bisin and Verdier (2001) and Desai (2008). They show that transmission of existing

beliefs by parents to their offspring would be consistent with maximizing the utility of the children or preserving their cultural traits.

Finally, the importance of history, culture, and past events such as discrimination in shaping today's beliefs, behavior, and outcomes has also been demonstrated in the empirical literature (Alesina, Giuliano, and Nunn 2011; Nunn and Wantchekan 2011; Argenziano and Gilboa 2012; Voth and Voigtlaender 2012).

The above discussion highlights the fact that beliefs regarding discrimination could be understood as collective memory that are passed on from one generation to another, which are likely to be remarkably similar from one generation to another for long stretches of time.

3.3 Data and empirics

In the following we investigate whether our theoretically presented channel of discrimination can explain worse outcomes of a historically discriminated group in activities characterized by interlinkages. The fact that blacks have lower self-employment rates than whites in the US by exploiting differences in taste for and beliefs about discrimination across regions and time has been shown in Ramachandran and Rauh (2013). We extend their analysis by increasing the sample size by including the most recent survey, adding two proxies for taste for discrimination, and looking at occupational prestige scores verifying that high skilled potential entrepreneurs suffer from discrimination.

The empirical literature dealing with discrimination and self-employment in the US documents the differences in participation and returns between ethnic groups (Moore 1983; Bailey and Waldinger 1991; Borjas 1986; Fairlie 1996; Fairlie and Meyer 1996, 2000; Blanchflower 2009). In line with our theoretical predictions, the above mentioned studies find lower rates of self-employment for black males, blacks are more likely to have loan applications rejected, and they pay higher interest rates on loans than white males with similar observables. Neal and Johnson (1996) find that for wage employment premarket skills measured by the Armed Forces Qualification Test (AFQT) score can explain most of the black-white wage gap. However, for self-employment Fairlie (2002) demonstrates that controlling for AFQT scores does not significantly reduce the black-white gap in the probability of becoming self-employed, indicating that discrimination might play a role in the observed differences in self-employment rates.

3.3.1 Data

We use the General Social Survey (GSS) from 1972-2012 with more than 50,000 observations along 29 questionnaires to test the predictions of the theoretical framework. We construct proxies for the taste for discrimination parameters in our model by taking the share of white people, for each year and region, answering the following questions:[20]

1. Whites who answer "yes" to "Do you think there should be laws against marriages of Blacks and Whites?"

2. Whites who are "very" or "somewhat opposed" when asked "What about having a close relative marry a Black person?"

3. Whites who prefer "separate school" when asked "Do you think white students and Black students should go to the same schools or to separate schools?"

4. Whites who "object" or "mildly object" when asked "How strongly would you object if a member of your family wanted to bring a Black friend home to dinner?"

In order to construct a proxy for beliefs regarding discrimination, we take the share of the sample, for each year and region, answering the following question with "yes":

- "On the average Blacks/African-Americans have worse jobs, income, and housing than White people. Do you think these differences are mainly due to discrimination?"

Unfortunately, neither of these questions is asked throughout the entire period, but depending on the model specifications we can make us of sample sizes of up to nearly 20,000 observations. In Figure 3.1 the four constructed measures of taste for discrimination, beliefs about discrimination, and the self-employment rates of blacks and whites are plotted from 1972-2012. The discrimination measures are obtained from the GSS dataset. The self-employment rates by race we obtain from the Current Population Survey (CPS) March supplement provided by the Integrated Public Use Microdata Series (IPUMS). On the one hand, all constructed measures of taste for discrimination decline linearly over the observed time period. On the other hand, beliefs about discrimination

[20] The reason for segmentation of beliefs and taste for discrimination by regions in the US can be found in Becker (1971).

have remained remarkably stable. Similarly, the gap in self-employment rates between blacks and whites has remained stable over observed time period.

In the aggregate beliefs about discrimination among whites peak in 1985 at 45% and reach the minimum in 2004 at 34%. Our first measure for taste for discrimination among whites declines from 39% in 1972 to 10% in 2002. The second measure declines from 66% in 1990 to 21% in 2012. The third prix for taste for discrimination takes its highest value in 1976 at 15%, while reaching its minimum in 1985 at 7%. The fourth measure for taste for discrimination ranges from 31% in 1973 to 19% in 1984.

Figure 3.4 in the Appendix shows the trends in beliefs and taste for each region. When decomposing by year and region, disparities across the US become very clear. Beliefs about discrimination take their highest value in 1993 in New England at 67% and the lowest in West South Central in 2002 at 20%.

3.3.2 The method

From the theoretical model we derive the factors influencing the probability of individuals being self-employed to be ability a, the proportion π_{tqs} of principals with a taste for discrimination at time t in region q against group s, the proportion φ_{tqs} with beliefs about discrimination at time t in region q against group s, and a vector of individual characteristics X_i with the associated parameter vector γ. Years of schooling serve us as a proxy for ability. Therefore, the probability of individuals i at time t in region q and of group s being self-employed $se_{itqs} \in \{0,1\}$ we define as

$$Prob(se_{itqs} = 1|a, \varphi_{tqs}, \pi_{tqs}, X_i) = f(a, \varphi_{tqs}, \pi_{tqs}, X_i). \tag{3.28}$$

With a logit regression we can define the estimated probability as

$$Prob(se_{itqs}|a, \varphi_{tqs}, \pi_{tqs}, X_i) = \frac{e^{g(a_i, \varphi_{tqs}, \pi_{tqs}, X_i)}}{e^{g(a, \varphi_{tqs}, \pi_{tqs}, X_i)} + 1} \tag{3.29}$$

where $g(a, \varphi_{tqs}, \pi_{tqs}, X_i) = \beta_0 + \beta_1 a_i + \beta_2 \varphi_{tqs} + \beta_3 \pi_{tqs} + \gamma X_i + \varepsilon_{itqs}$ and ε_{itqs}, the error term, is a binomially distributed random variable.

The share of principals with a taste for discrimination π_{tqs} and the share with beliefs

about discrimination φ_{tqs} take the value zero for white individuals, i.e. for $s = A$. We restrict our sample to white and black respondents who are neither students nor retired, while assuming no differences in preferences to become self-employed.[21]

In the logit regression estimating self-employment we control for gender, age, age squared, self-categorized family income when the respondent was 16, and whether the father was self-employed. Gender we include because of the well documented inequality between men and women in the job market. It would be interesting to see whether it is belief about discrimination against women leading to lower self-employment rates among women, but unfortunately the GSS does not provide the appropriate questions to construct the necessary measures. We add self-categorized family income in order to control for possible inherited wealth, which could help overcome credit constraints. The possibility of taking over a family business is controlled for by including a dummy, specifying whether the father was self-employed.

3.3.3 Results

Years of schooling, our proxy for ability, and all controls are significant and have the expected sign. We find that either proxy for taste for discrimination significantly reduces the likelihood of a black person being self-employed in model (1), (2), (3), and (4) of Table 3.2 in the Appendix. However, taste for discrimination only is significant as along as the proxy for beliefs about discrimination does not enter the estimation. Once beliefs about discrimination are included either proxy for taste for discrimination becomes insignificant, as can be seen in model (5), (6), (7) and (8). The proxy representing beliefs about discrimination is significant at the 1% level in (5) and (6), and at the 10% level in (7). In (8) it is insignificant. However when including region and year fixed effects beliefs about discrimination become significant at the 10% level as can be seen in Table 3.1.

Computing the mean effect of a decrease in beliefs about discrimination for an average black male in the sample for model (5), we find that by reducing average beliefs about discrimination by 10% points from 24% to 14%, the probability of the average

[21] In the International Social Survey on Work Orientation in 2005 we find that 71% of blacks versus 58% of whites in the labour force in the US would choose self-employment if they could choose between different kinds of jobs, suggesting that our estimates might even be underestimating effects of discrimination.

black male being self-employed increases from 7.3% to 9.3%, which is an increase of 28%. The magnitude of the effect of beliefs about discrimination becomes clear when we calculate the probability of the average black male being self-employed at zero taste for discrimination, but at the remaining 24% beliefs about discrimination. Here the probability of self-employment increases from 7.3% to 7.9%, which is an increase of 8%. The estimation indicates that abolishing taste for discrimination will not be enough to close the gap in self-employment rates. These findings suggest that when beliefs about discrimination are well established these are enough to lower the probability of black people being self-employed significantly, without even requiring a taste for discrimination to be present, just as predicted by our model. In Table 3.3 in the Appendix we control for regional or year fixed effects and find that beliefs about discrimination remain significant at the 1% level in most model specifications. When including both year and region fixed effects beliefs about discrimination are significant in all specifications, as can be seen in Table 3.1.

We construct a proxy for statistical discrimination by coding as statistical discrimination the share of whites by year and region that assign a value from one to three on a scale from one to seven to the question "Do people in these groups [blacks] tend to be unintelligent or tend to be intelligent?", where one signifies unintelligent to seven signifying intelligent.[22] Here we only include the second measure for taste for discrimination because only that question coincides for sufficient waves of the survey. When not paired with beliefs, statistical discrimination enters with the expected negative sign and is highly significant, but is insignificant when paired with beliefs about discrimination. Beliefs about discrimination remain significant at a 1% level whether including year and/or regional fixed effects (Table 3.4 in the Appendix).

In order to check whether the results are driven by unskilled individuals, who become entrepreneurs due to a lack of labor market opportunities, we test whether beliefs about discrimination are a significantly negative estimator for the self-employed excluding farmers, and excluding the self-employed whose occupation has a prestige score of no more than 40 points on the 1970 and 1980 prestige score scale.[23] As can be seen in Table 3.5, beliefs about discrimination are significant at the 1% level for all chosen

[22] This measure suffers from the problem that it most likely captures taste based discrimination, as there exists no scientific evidence of racial differences in intelligence.

[23] This somewhat arbitrary value is what we felt excluded low-skilled blue collar occupations. The results are robust for a range of cut-off points and are available upon request.

Table 3.1: Logistic regression with year and region fixed effects

Dependent variable: Self-employment				
	(1)	(2)	(3)	(4)
Belief about discrimination	-2.616***	-3.345**	-2.779**	-2.423*
	(0.875)	(1.537)	(1.214)	(1.405)
Taste for discrimination 1	-1.009			
	(1.502)			
Taste for discrimination 2		-0.792		
		(1.386)		
Taste for discrimination 3			1.715	
			(2.527)	
Taste for discrimination 4				0.158
				(2.054)
Years of schooling	0.030***	0.015	0.048**	0.047**
	(0.012)	(0.012)	(0.024)	(0.024)
Female	-0.662***	-0.648***	-0.960***	-0.959***
	(0.067)	(0.067)	(0.147)	(0.147)
Age	0.100***	0.098***	0.092***	0.092***
	(0.014)	(0.014)	(0.029)	(0.029)
Age squared x 1,000	-0.795***	-0.703***	-0.756**	-0.756**
	(0.140)	(0.143)	(0.297)	(0.297)
Family income at age 16	0.178***	0.124***	0.135	0.134
	(0.042)	(0.040)	(0.089)	(0.089)
Father was self-employed	0.589***	0.602***	0.746***	0.746***
	(0.069)	(0.071)	(0.148)	(0.148)
Year fixed effects	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes
Pseudo R^2	0.066	0.068	0.098	0.098
Observations	9417	8812	2211	2211

All regressions include a constant. Standard errors are in parentheses.
*, ** and *** significant at 10, 5 and 1 % significance level, respectively.
Datasource: General Social Survey

specifications of self-employment.

3.4 Policy considerations

The discriminatory equilibrium can persist despite its inefficiency. This naturally raises the question of the existence of effective affirmative action, which could move the economy to the "good" equilibrium as a focal point in the coordination game. The following analysis is restricted to the long run equilibrium where no taste for discrimination remains, but discrimination persists due to beliefs.

One way to overcome discrimination, would be to simply provide a financial subsidy to the B-type individuals endowed with sufficiently high ability to become entrepreneurs, but who is discriminated upon. With the subsidy the individual could buy himself out of the discriminatory range and their offers would be accepted with certainty. This measure would overcome the problem that beliefs are prohibiting both the principals from accepting and the individuals from applying. The caveat of this measure is that it would only be effective as long the subsidy is in place, as this solution does not alter beliefs. Moreover, the welfare effect would be negative, as the additional value creation attributed to self-employment sums up to less than the subsidy.[24]

Another method of achieving equality among the A and the B-type would be to discriminate against the discriminator. By imposing a fine F on principals who reject a B-type that has the same ability and offers the same amount as an A-type that has been accepted in the same period, one could facilitate equal treatment of A- and B-types.[25] This equal treatment might come at a high cost, though. If one principal interacts with various individuals in a given period there exists the possibility that principals begin discriminating against the A-type as well. In order to avoid the fine when rejecting the B-type, discrimination could spill over to the A-type. Imagine a principal receiving the same offer $\hat{\sigma}$ by two individuals with identical ability $\hat{a}$, but types A and B, in the same period. Now if he accepts the A-type and rejects the B-type he will receive

[24] Observe that if it was not the case, the individuals themselves would pay extra to compensate the belief about discrimination.

[25] This solution assumes that the authorities could estimate ability through information on observables such as education, years of experience, age, etc.

$\hat{\sigma}+1+r-F$, given that $-p$ accepts the A-type offer as well. This would only be rational if $\hat{\sigma}-F \geq 1+r$, because otherwise he would be better off by also rejecting the A-type.

One possibility would be for the government to step in as a lender of last resort to the discriminated. This intervention would have to be communicated publicly, such that it would serve as a signal and would spillover to the beliefs of the B-type and the distributors. To see this in terms of our model, imagine the government announcing publicly and credibly the implementation of this policy. Now there would be no reason for the distributor or the individual of the B-type to assign $\varphi > 0$. The great advantage of this intervention would be that intervening in one market would be enough to correct beliefs in other markets. Once the measure were to be removed, beliefs about discrimination would have vanished and no further discrimination would take place (given no taste for discrimination). This hinges on requiring only two principals for the production process. For n principals the government would have to promise to act as $n-1$ principals of last resort.

A further possibility to overcome the coordination failure would be the introduction of an institution acting as coordination device providing the service of linking prescreened non-discriminatory lenders and distributors to able blacks wanting to become entrepreneurs. As this could even be a profitable exercise such institutions might automatically arise and be provided by the market itself.

In the above we saw that schemes, such as subsidies or equal treatment regulations, might only address the problem myopically or, even worse, have undesirable consequences (like discrimination of A-types in equilibrium).

3.5 Conclusions

In this paper we show that the legacy of past discrimination can cast a long shadow. Even once no more taste for discrimination or statistical discrimination exists in a society, discrimination can persist due to remaining beliefs making discrimination the best-response. Discrimination can take place despite individuals' ability being common knowledge and ex-ante and ex-post identical between a discriminated and non-discriminated individual. This form of discrimination is sustained under much weaker conditions than traditionally assumed in the literature.

The theoretical mechanism put forth is relevant for markets characterized by the need to establish productive relations or interlinkages with other agents in society in order for the production process to be carried out. It is shown that in such markets the presence of beliefs regarding the existence of taste discriminators, even when no agents with taste for discrimination exist in society, can result in agents exhibiting discriminatory behavior in equilibrium. Discrimination arises as a rational response to the belief that other agents might discriminate, which would impose losses due to the complementarity in the production process. The model shows lower participation and payoff to the discriminated group in markets characterized by the presence of interlinkages.

Empirical evidence in support of the theoretical framework is provided by analyzing the market for self-employment, a market characterized by the need to establish productive relations to be able to operate and be successful. The outcomes predicted by the model, in terms of participation rates and incomes for the self-employed for the discriminated group being lower, are validated using data from the General Social Survey 1972-2012 of the US and creating proxies for taste and beliefs regarding discrimination. A simple logit model suggests that beliefs about discrimination are significant in explaining lower self-employment rates of blacks compared to whites, even after controlling for individual and family characteristics, as well as region and year fixed effects.

In the model overall welfare is lower and on average nobody is better off in the discriminatory equilibrium. We show that the nature of discriminatory coordination failure does not enable competitive market forces to overcome discrimination, but may require alternative policy tools. The various mechanisms through which discrimination manifests its dynamic linkages in terms of cross market and intergenerational effects, and the tendency to persist through cumulative and belief based channels, need to be understood and explored in order to develop policies aimed at eradicating discrimination and achieving equal treatment and opportunities.

3.6 Appendix

3.6.1 Proofs

3.6.1.1 Proposition 1

Proof. 1. First, observe that by definition individuals in the ability range $a_l^* \leq a \leq a_h^*$ offer $1+r+b_l \leq \sigma^N(a) \leq 1+r+b_h$, and hence can be rejected only by the high type taste discriminators. Therefore, they care only about the probability of meeting a high taste discriminator. The fact that no individual in the range $a_l^* \leq a \leq a_h^*$ prefers the Nash bargaining solution to entering the activity involving no interlinkages implies $(1-\varphi_{T^*}^h)^2(\lambda a - 2\sigma^N(a)) + (1-(1-\varphi_{T^*}^h)^2)(a-\delta) < a$. The left hand side of the expression is strictly increasing in a, which implies that if it is not satisfied for a_h^* then it is not satisfied for all $a \leq a_h^*$. This implies that all individual with a, such that $\lambda a - 2(1+r+b_h) \geq a$ (or all $a_b \leq a \leq a_h^*$) offer a share equal to $(1+r+b_h)$ and are accepted to enter the activity involving interlinkages. All individuals with $a^* \leq a \leq a_b$ are unable to offer a share to compensate the taste of high type discriminators and hence enter the activity involving no interlinkages. Moreover, note as now either all individuals of the B-type with $a \geq a_h^*$ offer $\sigma^N(a) \geq (1+r+b_h)$, and all individuals with $a_b \leq a \leq a_h^*$ offer $(1+r+b_h)$, implies that from period T^* onwards there will be no offers made within the range of $(1+r)$ to $(1+r+b_h)$, and hence the beliefs remain frozen at the current level implying the discriminatory equilibrium will persist for ever.

2. The individual with $a^* \leq a \leq a_l^*$, while comparing the option of offering the Nash bargaining solution and entering the activity involving no interlinkages, calculates the expected payoff from offering the Nash bargaining solution given by $(1-(\varphi_{T^*}^h + \varphi_{T^*}^l))^2(\lambda a - 2\sigma^N(a)) + (1-(1-(\varphi_{T^*}^h + \varphi_{T^*}^l))^2)(a-\delta)$. Their $\sigma^N(a) < 1+r+b_l$ implies that they need to take into account both, the share of high and low taste discriminators. Let us denote by a_{lN} as the lowest type individual in the range $a^* \leq a \leq a_l^*$ who prefers offering the Nash bargaining solution to entering the activity involving no interlinkages at T^*. As the above expression is strictly increasing in a, it implies that all individuals with $a > a_{lN}$ offer the Nash bargaining solution in period T^*. This means that all individu-

als in the ability range $a_{lN} \leq a \leq a_h^*$ offer the Nash bargaining solution and are accepted. Let the total number of cases for potential discrimination by low and high type discriminators be n_{pot}^l and n_{pot}^h, respectively. The point estimates in the next period T^*+1 for meeting a high and a low type discriminator are given by $\varphi_{T^*+1}^h = \frac{\alpha_0^h + \sum_{t=1}^{T^*}(k_t^h + k_t^{uh})}{\alpha_0^h + \beta_0^h + \sum_{t=1}^{T^*} n_{th} + n_{pot}^h}$ and $\varphi_{T^*+1}^l = \frac{\alpha_0^l + \sum_{t=1}^{T^*}(k_t^{ul})}{\alpha_0^l + \beta_0^l + \sum_{t=1}^{T^*} n_{tl} + n_{pot}^l}$, respectively. Therefore, $\varphi_{T^*+1}^l < \varphi_{T^*}^l$ and $\varphi_{T^*+1}^h < \varphi_{T^*}^h$, implying the lowest type who applies in T^*+2 is such that $a < a_{lN}$, and thus $\varphi_{T^*+t}^j < \varphi_{T^*}^j$ for all $t > 0$, or $\frac{d\varphi_t^j}{dt} < 0$ for all $t > T^*$. This implies at some point $\varphi_T^j \to 0$, wherefore all $a \geq a^*$ apply and enter the activity involving interlinkages and discrimination does not persist in society. The number of periods for which discrimination will persist could be calculated as a function of $\varphi_{T^*}^j = f(\pi_1, \pi_2, \omega)$ and the lowest a who still prefers the Nash solution to entering the activity involving no interlinkages at T^*.

3. Let us denote by a_{hN} the lowest ability type in the range $a_l^* < a < a_h^*$ who prefers to offer the Nash bargaining solution over entering the activity involving no interlinkages in period T^*. Therefore, all individuals in the range $a_{hN}^* \leq a \leq a_h^*$ offer their Nash bargaining solution and are accepted. This implies that $\varphi_{T^*}^h > \varphi_{T^*+1}^h$ or $\frac{d\varphi_t^h}{dt} < 0$ for all $t > T^*$. This in turn implies that the ability of the lowest ability type, who applies, is decreasing in time for $t > T^*$.

   (a) If the individual with the lowest a, who prefers the Nash bargaining solution to entering the activity involving no interlinkages, is such that $a > a_l^*$ when $\varphi_T^h \to 0$, then individuals with ability $a > a_l^*$ will become self-employed and pay $\sigma^N(a)$ and all individuals in the ability range $a^* \leq a < a_l^*$ will enter the activity involving no interlinkages.

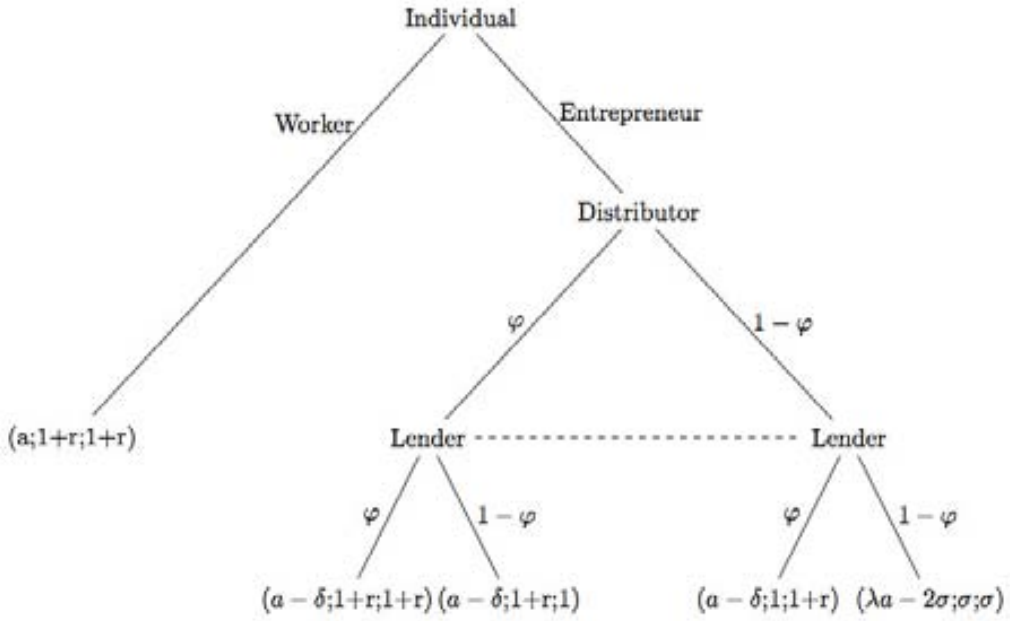
   (b) If the individual with the lowest a, who prefers to offer Nash bargaining solution to entering the activity involving no interlinkages as $\varphi_T^h \to 0$, is such that $a < a_l^*$, then all individuals with ability $a \geq a^*$ will become self-employed and discrimination will not persist in the long run.

■

3.6.1.2 Proposition 2

Proof. Let us assume that the set $\mathcal{B}$ is large enough that individuals' abilities can be approximated by a continuous distribution on the interval $[0,1]$. Proposition 1.1 shows that all individuals of the B-type with $a^* \leq a < a_b$ enter activities not involving interlinkages, whereas *A*-types of the same ability enter activity involving interlinkages. Moreover, *B*-type individuals with $a_b \leq a < a_h^*$ offer $1+r+b_h > \sigma^N(a)$.

Now we can write the transfer, which the principals receive from individuals wanting to escape the discrimination as an integral over the ability range who pay the higher fee: $\int_{a_b}^{a_h^*} 2(1+r+b_h-\sigma^N(a))da = \int_{a_b}^{a_h^*} 2(1+r+b_h-\frac{(\lambda-1)a+r+\delta+1}{3})da$. Therefore, the gain to the principals would be $\frac{1}{3}((a_h^*-a_b)(4+4r+6b_h-2\delta)-(\lambda-1)((a_h^*)^2-a_b^2)))$. Now the loss to principals, due to able individuals entering activities not involving interlinkages, can be written as: $\int_{a^*}^{a_b} 2[\sigma^N(a)-(1+r)]da = \int_{a^*}^{a_b} 2[\frac{(\lambda-1)a+r+\delta+1}{3}-(1+r)]da$ $\Rightarrow$ Loss$= \frac{1}{3}((a_b^2-(a^*)^2)(\lambda-1)+(a_b-a^*)(2\delta-4r-4))$. In order for the loss to be at least as big as the gain we require: $((a_h^*-a_b)(4+4r+6b_h-2\delta)-(\lambda-1)((a_h^*)^2-a_b^2) \leq ((a_b^2-(a^*)^2)(\lambda-1)+(a_b-a^*)(2\delta-4r-4))$. Rearranging we get $(4+4r-2\delta)(a_h^*-a^*)+6b(a_h^*-a_b) \leq ((a_h^*)^2-(a^*)^2)(\lambda-1)$. Now substituting $a^* = \frac{2(1+r+\delta)}{\lambda-1}$, $a_h^* = \frac{2(1+r)+3b_h-\delta}{\lambda-1}$, and $a_b = \frac{2(1+r+b_h)}{\lambda-1}$ we find that for this to hold the condition is $b_h \geq \delta$. But since no discrimination exists when $\delta > b_h$ (because in that case the offer $\sigma^N(a) > 1+r+b_h$ for all $a \geq a^*$) gains can never be greater than losses if we are in an equilibrium as in proposition 1.1. ■



Note: The payoffs are in the following order: Individual, Distributor, Lender.

Figure 3.2: Extensive game form of decision of B-type individual

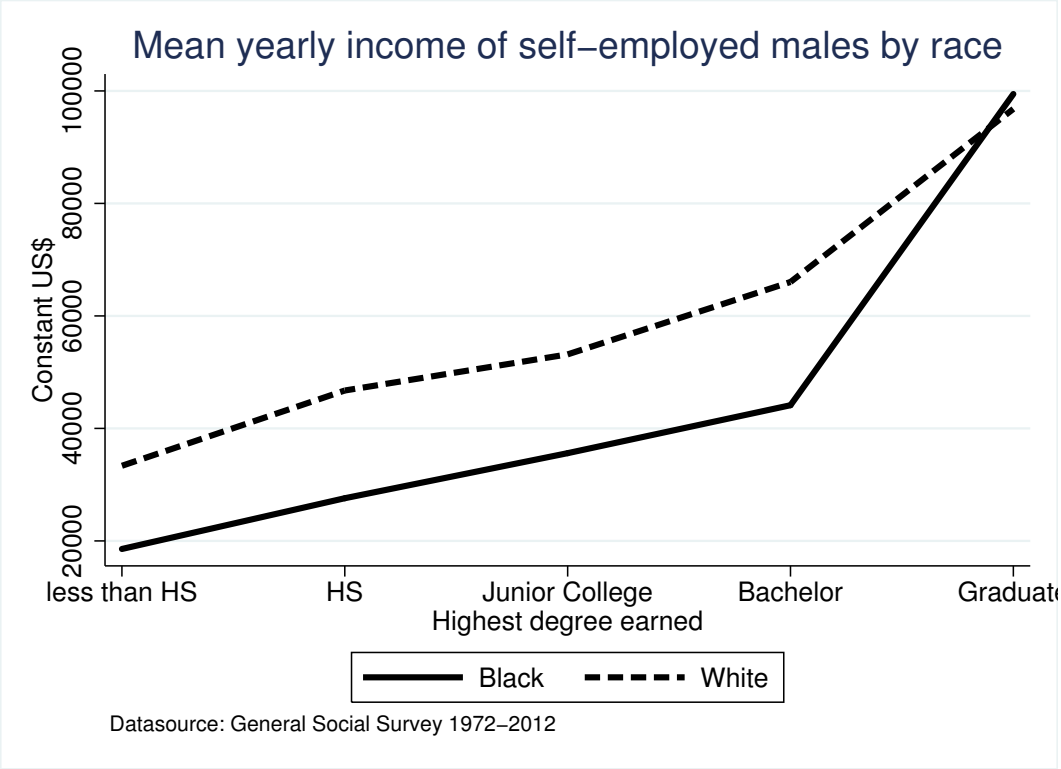


Figure 3.3: Mean yearly income of black and white self-employed by educational degree

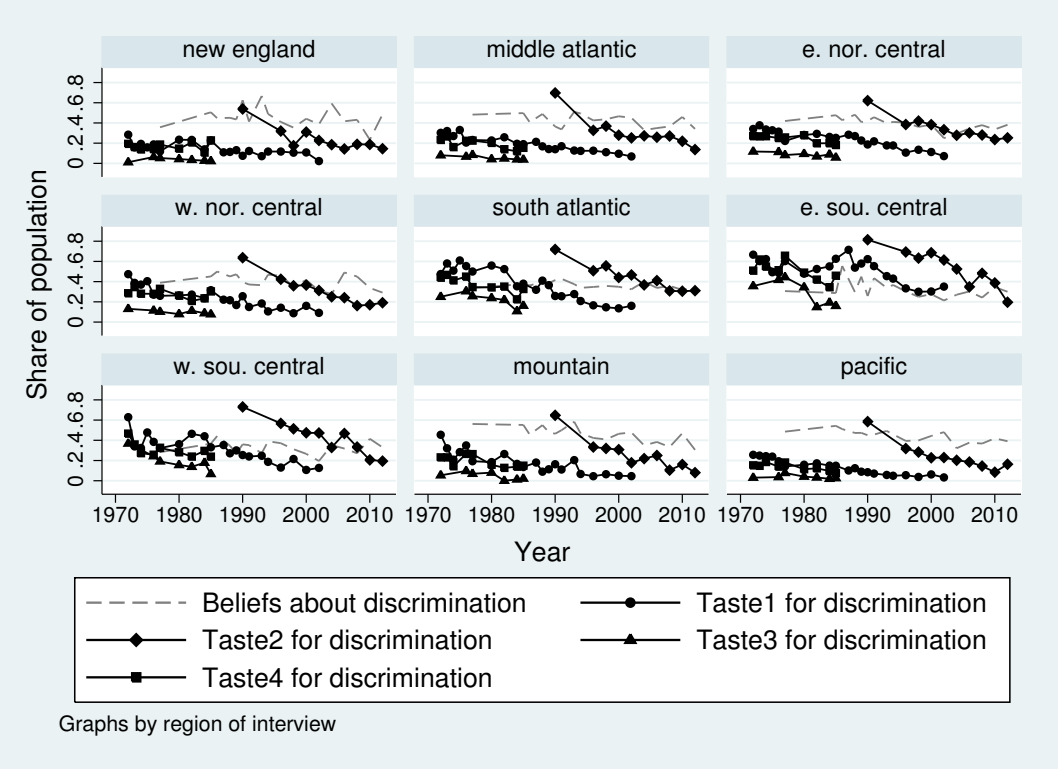


Figure 3.4: Beliefs and taste regarding discrimination by region in the US

Table 3.2: Logistic regression benchmark

Dependent variable: Self-employment									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Belief about discrim.					-2.547***	-3.734**	-2.214*	-1.486	-3.305***
					(0.883)	(1.501)	(1.225)	(1.418)	(0.453)
Taste for discrim. 1	-1.773***				-1.202				
	(0.404)				(1.511)				
Taste for discrim. 2		-3.760***				-0.506			
		(0.756)				(1.342)			
Taste for discrim. 3			-2.302**				-0.208		
			(0.902)				(2.547)		
Taste for discrim. 4				-1.488***				-1.469	
				(0.464)				(2.099)	
Years of schooling	0.028***	0.015	0.036***	0.033***	0.037***	0.015	0.066***	0.065***	0.022***
	(0.008)	(0.012)	(0.013)	(0.011)	(0.011)	(0.012)	(0.023)	(0.023)	(0.008)
Female	-0.845***	-0.655***	-0.942***	-0.997***	-0.657***	-0.656***	-0.939***	-0.937***	-0.719***
	(0.049)	(0.067)	(0.079)	(0.071)	(0.067)	(0.067)	(0.146)	(0.146)	(0.049)
Age	0.091***	0.098***	0.080***	0.081***	0.099***	0.098***	0.083***	0.082***	0.094***
	(0.009)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.028)	(0.028)	(0.010)
Age$^2x1,000$	-0.689***	-0.708***	-0.621***	-0.599***	-0.774***	-0.710***	-0.641**	-0.640**	-0.730***
	(0.096)	(0.142)	(0.148)	(0.133)	(0.139)	(0.142)	(0.293)	(0.293)	(0.102)
Family income age 16	0.186***	0.123***	0.199***	0.177***	0.180***	0.123***	0.139	0.138	0.149***
	(0.030)	(0.040)	(0.049)	(0.044)	(0.041)	(0.040)	(0.089)	(0.089)	(0.030)
Father was self-empl.	0.636***	0.617***	0.753***	0.679***	0.575***	0.613***	0.681***	0.684***	0.612***
	(0.048)	(0.070)	(0.077)	(0.069)	(0.068)	(0.070)	(0.143)	(0.143)	(0.050)
Pseudo R^2	0.066	0.062	0.074	0.075	0.060	0.063	0.079	0.080	0.061
Observations	19584	8812	7833	10037	9417	8812	2211	2211	16990

All regressions include a constant. Standard errors are in parentheses.
*, ** and *** significant at 10, 5 and 1 % significance level, respectively.
Datasource: General Social Survey

Table 3.3: Logistic regression with year or region fixed effects

Dependent variable: Self-employment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Belief about discrimination	-2.738***	-3.385**	-2.676**	-2.203	-2.404***	-3.734**	-2.273*	-1.642
	(0.882)	(1.534)	(1.220)	(1.411)	(0.874)	(1.503)	(1.219)	(1.411)
Taste for discrimination 1	-0.875				-1.377			
	(1.508)				(1.499)			
Taste for discrimination 2		-0.765				-0.496		
		(1.371)				(1.355)		
Taste for discrimination 3			1.174				0.183	
			(2.514)				(2.551)	
Taste for discrimination 4				-0.324				-1.110
				(2.066)				(2.080)
Years of schooling	0.031***	0.015	0.053**	0.052**	0.036***	0.015	0.062***	0.062***
	(0.012)	(0.012)	(0.024)	(0.024)	(0.011)	(0.012)	(0.024)	(0.024)
Female	-0.659***	-0.655***	-0.948***	-0.946***	-0.661***	-0.649***	-0.957***	-0.956***
	(0.067)	(0.067)	(0.147)	(0.147)	(0.067)	(0.067)	(0.147)	(0.147)
Age	0.101***	0.098***	0.091***	0.091***	0.099***	0.098***	0.084***	0.084***
	(0.014)	(0.014)	(0.029)	(0.029)	(0.014)	(0.014)	(0.028)	(0.028)
Age squared x 1,000	-0.796***	-0.702***	-0.737**	-0.736**	-0.773***	-0.710***	-0.665**	-0.665**
	(0.140)	(0.142)	(0.295)	(0.295)	(0.139)	(0.142)	(0.294)	(0.294)
Family income at age 16	0.180***	0.126***	0.141	0.140	0.178***	0.121***	0.135	0.133
	(0.041)	(0.040)	(0.089)	(0.089)	(0.042)	(0.040)	(0.089)	(0.089)
Father was self-employed	0.588***	0.613***	0.708***	0.709***	0.575***	0.601***	0.702***	0.704***
	(0.068)	(0.070)	(0.144)	(0.144)	(0.069)	(0.071)	(0.147)	(0.147)
Year fixed effects	Yes	Yes	Yes	Yes	No	No	No	No
Region fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Pseudo R^2	0.063	0.064	0.089	0.089	0.063	0.067	0.088	0.088
Observations	9417	8812	2211	2211	9417	8812	2211	2211

All regressions include a constant. Standard errors are in parentheses.
*, ** and *** significant at 10, 5 and 1 % significance level, respectively.
Datasource: General Social Survey

Table 3.4: Logistic regression with statistical discrimination

Dependent variable: Self-employment									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Belief about discrim.		-3.677**	-4.519***	-3.325**	-4.313***	-3.602**	-4.662***	-3.205**	-4.441***
		(1.470)	(1.200)	(1.502)	(1.221)	(1.464)	(1.180)	(1.497)	(1.200)
Statistical discrim.	-25.208***	11.117	2.198	11.444	1.240	14.056	3.461	14.409	2.386
	(6.008)	(11.354)	(7.183)	(11.447)	(7.321)	(10.977)	(6.897)	(11.078)	(7.039)
Taste for discrim. 2		-1.996		-2.298		-2.446		-2.795	
		(2.053)		(2.083)		(2.072)		(2.106)	
Years of schooling	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Female	-0.657***	-0.656***	-0.657***	-0.655***	-0.655***	-0.649***	-0.650***	-0.647***	-0.648***
	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
Age	0.098***	0.098***	0.098***	0.098***	0.098***	0.098***	0.098***	0.098***	0.098***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Age squared x 1,000	-0.705***	-0.710***	-0.710***	-0.701***	-0.702***	-0.710***	-0.710***	-0.701***	-0.702***
	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.143)	(0.143)
Family income age 16	0.127***	0.122***	0.123***	0.125***	0.126***	0.120***	0.121***	0.123***	0.124***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Father was self-emp.	0.620***	0.612***	0.612***	0.613***	0.613***	0.600***	0.601***	0.601***	0.602***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.071)	(0.071)	(0.071)	(0.071)
Year fixed effects	No	No	No	Yes	Yes	No	No	Yes	Yes
Region fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes
Pseudo R^2	0.060	0.063	0.063	0.064	0.064	0.067	0.067	0.068	0.068
Observations	8812	8812	8812	8812	8812	8812	8812	8812	8812

All regressions include a constant. Standard errors are in parentheses.
*, ** and *** significant at 10, 5 and 1 % significance level, respectively.
Datasource: General Social Survey

Table 3.5: Logistic regression with different definitions of self-employment

Dependent variable: Self-employment				
	(Benchmark)	(Prestige80)	(Prestige70)	(No farmers)
Belief about discrimination	-3.180***	-2.480***	-4.745***	-3.180***
	(0.455)	(0.801)	(0.930)	(0.455)
Years of schooling	0.019**	0.099***	0.129***	0.019**
	(0.009)	(0.017)	(0.013)	(0.009)
Female	-0.717***	-0.975***	-0.607***	-0.717***
	(0.049)	(0.108)	(0.072)	(0.049)
Age	0.096***	0.092***	0.101***	0.096***
	(0.010)	(0.021)	(0.016)	(0.010)
Age squared x 1,000	-0.744***	-0.656***	-0.714***	-0.744***
	(0.103)	(0.217)	(0.163)	(0.103)
Family income at age 16	0.148***	0.197***	0.206***	0.148***
	(0.030)	(0.064)	(0.044)	(0.030)
Father was self-employed	0.612***	0.626***	0.519***	0.612***
	(0.051)	(0.106)	(0.075)	(0.051)
Year fixed effects	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes
Pseudo R^2	0.067	0.097	0.088	0.067
Observations	16990	5758	12262	16990

All regressions include a constant. Standard errors are in parentheses.
The column "Benchmark" includes all self-employed, "Prestige80" and "Prestige70" exclude those with occupations with a prestige score of no more than 40 according to the 1980 definition, and "No farmers" excludes farmers.
*, ** and *** significant at 10, 5 and 1 % significance level, respectively.
Datasource: General Social Survey

References

[1] Akerlof, G.A. (1976). The Economics of Caste and of the Rat Race and Other Woeful Tales, *Quarterly Journal of Economics*, vol. 90(4), pp. 599-617.

[2] Akerlof, G.A. (1985). Discriminatory, Status-based Wages among Tradition-oriented, Stochastically Trading Coconut Producers, *Journal of Political Economy*, vol. 93(2), pp. 265-76.

[3] Albrecht, K., Von Essen, E., Parys, J., and Szech, N. (2013). Updating, self-confidence, and discrimination. *European Economic Review*, 60, pp. 144-69.

[4] Alesina, A.F., Giuliano, P. and Nunn, N. (2011). On the Origins of Gender Roles: Women and the Plough, *Quarterly Journal of Economics*, vol. 128(2), pp. 469-530.

[5] Alesina, A. F., Lotti, F., and Mistrulli, P. E. (2013). Do women pay more for credit? Evidence from Italy, *Journal of the European Economic Association*, vol. 11(1), pp. 45-66.

[6] Argenziano, R. and Gilboa, I. (2011). History as a Coordination Device, *Theory and Decision*, 73(4), pp. 501-12.

[7] Arrow, K. (1973). The Theory of Discrimination, in (O. Ashenfelter and A. Rees, eds.), *Discrimination in Labor Markets*, pp. 3-33, Princeton: Princeton University Press.

[8] Bailey, T. and Waldinger, R. (1991). Primary, Secondary, and Enclave Labor Markets: A Training System Approach, *American Sociological Review*, vol. 56, pp. 432-45.

[9] Banerjee, A. and Munshi, K. (2004). How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur, *Review of Economic Studies*, vol. 71(1), pp. 19-42.

[10] Basu, K. (2010). *Beyond the Invisible Hand: Groundwork for a New Economics*, Princeton: Princeton University Press, and New Delhi: Penguin.

[11] Becker, G.S. (1957). *The Economics of Discrimination*, Chicago: University of Chicago Press.

[12] Becker, G.S. (1971). *The Economics of Discrimination*, Chicago: University of Chicago Press.

[13] Bisin, A. and Verdier, T. (2001). The Economics of Cultural Transmission and the Evolution of Preferences, *Journal of Economic Theory*, 97, 298-319.

[14] Blanchflower, D.G. (2009). Minority Self-Employment in the United States and the Impact of Affirmative Action Programs, *Annals of Finance*, vol. 5(3-4), pp. 361-96.

[15] Borjas, G.J. (1986). The Self-Employment Experience of Immigrants, Working Paper 1942, National Bureau of Economic Research, Inc.

[16] Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the Dynamics of Segregation, *Quarterly Journal of Economics*, vol. 123(1), pp. 177-218.

[17] Coate, S. and Loury, G.C. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes?, *American Economic Review*, vol. 83(5), pp. 1220-40.

[18] Daskalova, V. (2013). Discrimination, Social Identity, and Coordination: An Experiment. mimeo.

[19] Dessi, R. (2008). Collective Memory, Cultural Transmission and Investments, *American Economic Review*, 98(1), pp. 534-560.

[20] Fairlie, R.W. (1996). The Absence of African-American Owned Business: An Analysis of the Dynamics of Self-Employment, *Journal of Labor Economics*, vol. 17(1), pp. 80-108.

[21] Fairlie, R.W. (2002). Drug Dealing and Legitimate Self-Employment, *Journal of Labor Economics*, vol. 20(3), pp. 538-67.

[22] Fairlie, R.W. and Meyer, B.D. (1996). Ethnic and Racial Self-Employment Differences and Possible Explanations, *Journal of Human Resources*, vol. 31(4), pp. 757-93.

[23] Fairlie, R.W. and Meyer, B.D. (2000). Trends in Self-Employment Among White and Black Men During the Twentieth Century, *Journal of Human Resources*, vol. 35(4), pp. 643-69.

[24] Fryer, R. and Jackson, M. (2008). A Categorical Model of Cognition and Biased Decision-Making, *B.E. Journal of Theoretical Economics*, vol. 8(1), Article 6.

[25] Gilboa, I., and Schmeidler, D. (2001). *A theory of case-based decisions*. Cambridge University Press.

[26] Greif, A. (1989). Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders, *Journal of Economic History*, vol. 49(4), pp. 857-82.

[27] Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade, *American Economic Review*, vol. 83(3), pp. 525-49.

[28] Griffin, L.J., and Bollen, K. (2009). What Do These Memories Do? Civil Rights Remembrance and Racial Attitudes, *American Sociological Review*, 74(4), 594-614.

[29] Halbwachs, M. (1992). *On Collective Memory*, Trans. and ed. Lewis A. Coser. Chicago: University of Chicago Press.

[30] International Social Survey Programme (ISSP): Work Orientations III, 2005, Distributor: GESIS Cologne Germany ZA4850, Data Version 2.0.0 (2009-10-29).

[31] Iyer, L., Khanna, T. and Varshney, A. (2011). Caste and Entrepreneurship in India, Working Paper 12-028, Harvard University.

[32] Kandori, M. (1992). Social norms and community enforcement. *Review of Economic Studies*, 59(1), 63-80.

[33] Kaushal, N., Kaestner, R., and Reimers, C. (2007). Labor market effects of September 11th on Arab and Muslim residents of the United States. *Journal of Human Resources*, 42(2), 275-308.

[34] King, M., Ruggles, S., Alexander, S.T., Flood, S., Genadek, K., Schroeder, M.B., Trampe, B., and Vick, R. (2010). Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota.

[35] Lucas, R.E., Jr. (1978). On the Size Distribution of Business Firms, *Bell Journal of Economics*, vol. 9(2), pp. 508-23.

[36] Lundberg, S. and Startz, R. (1983). Private Discrimination and Social Intervention in Competitive Labor Markets, *American Economic Review*, vol. 73(3), pp. 340-47.

[37] Lundberg, S. and Startz, R. (1998). On the Persistence of Racial Inequality, *Journal of Labor Economics*, vol. 16(2), pp. 292-323.

[38] Mailath, G.J., Samuelson, L. and Shaked, A. (2000). Endogenous Inequality in Integrated Labor Markets with Two-Sided Search, *American Economic Review*, vol. 90(1), pp. 46-72.

[39] Neal, D.A. and Johnson, W.R. (1996). The Role of Premarket Factors in Black-White Wage Differences, *Journal of Political Economy*, vol. 104(5), pp. 869-95.

[40] Nunn, N. and Wantchekon, L. (2011). The Trans-Atlantic Slave Trade and the Evolution of Mistrust in Africa: An Empirical Investigation, *American Economic Review*, vol. 101(7), pp. 3221-52.

[41] Peski, M. and Szentes, B. (2013). Spontaneous discrimination, *American Economic Review*, vol. 103(6), 2412-36.

[42] Phelps, E.S. (1972). The Statistical Theory of Racism and Sexism, *American Economic Review*, vol. 62(4), pp. 659-61.

[43] Ramachandran, R. and Rauh, C. (2013). Discrimination Without Taste - How Discrimination Can Spillover and Persist. Chapter 2 of Phd thesis (Ramachandran, R.). Essays on Education, Discrimination and Development. Universitat Autonoma de Barcelona.

[44] Smith, T.W, Marsden, P., Hout, M. and Kim, J. (2012). General social surveys, 1972-2012 Chicago: National Opinion Research Center [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut.

[45] Voth, J. and Voigtlaender, N. (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany, *Quarterly Journal of Economics*, 127(3), 1339-92.

[46] World values survey (2009). 1981-2008 Official aggregate v.20090901 (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid.