

THE UNIVERSITY OF CHICAGO

ESSAYS ON INEQUALITY, FAIRNESS, AND TAXATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY
LANCELOT HENRY DE FRAHAN

CHICAGO, ILLINOIS

DECEMBER 2018

Copyright © 2018 by Lancelot Henry de Frahan
All Rights Reserved

To Lorraine, Paps, Mams, Janine and our future little one.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	viii
ACKNOWLEDGMENTS	ix
ABSTRACT	xii
1 PREFERENCE RESPONSIBILITY VS POVERTY REDUCTION IN THE TAXATION OF LABOR INCOMES	1
1.1 Introduction	1
1.2 The Ethics of Responsibility: A Very Short Introduction	6
1.3 The Model and the Basic Axioms	8
1.4 Poverty Reduction	11
1.5 A Local Maximin Result	18
1.6 A General Second-Best Result	22
1.7 Additively Separable Preferences and One-Dimensional Heterogeneity: The Optimal Tax Formula for Low Incomes	26
1.8 Quasilinear and Iso-Elastic Preferences	32
1.9 Calibration to the US Economy	35
1.9.1 Quasilinear Reference Preferences	36
1.9.2 Least Redistributive Social Preferences	41
1.9.3 Comparison to the Current US Tax and Welfare System	42
1.10 Conclusion	46
2 INTERGENERATIONAL ELASTICITY OF CONSUMPTION	49
2.1 Introduction	49
2.2 Data and Variables	54
2.2.1 1940 Census	54
2.2.2 DataQuick	56
2.2.3 CEX	57
2.3 Raw Correlation in Housing Consumption across Last Names	58
2.4 From Surnames to Individual-level Correlations	63
2.4.1 Regional Heterogeneity	74
2.5 Housing Consumption as a Proxy for Total Consumption	77
2.6 Selection into Home Ownership in 2012	83
2.7 Conclusion	86
REFERENCES	89

A	APPENDIX TO CHAPTER 1	97
A.1	Proofs	97
A.1.1	Proof of Proposition 2	97
A.1.2	Proof of Corollary 1	101
A.1.3	Proof of Proposition 3	102
A.1.4	Proof of Corollary 2	105
A.1.5	Derivation of Optimal Income Tax Formulas under Additive Separability and proof of Proposition 4	107
A.1.6	Solving the lower-subproblem	110
A.1.7	Solving the upper-subproblem	114
A.1.8	Deriving the Optimal Income Tax Formulas under Quasilinear and Iso-elastic Preferences and Proof of Proposition 5	116
A.2	Other calibrated optimal tax rates and comparative statics	118
A.3	Calibration details and description of the current US income tax schedule	121
A.4	Least Redistributive SOF: Parametrizing the reference preferences	130
B	APPENDIX TO CHAPTER 2	132
B.1	Additional Figures and Tables	132
B.2	Converting Durable expenditures into Service Flows in the CEX	136
B.3	Derivation of exclusion restrictions in terms of δ_{0i}, δ_{1i}	137
B.4	Choice of the weighting matrix and asymptotic standard errors	140
B.5	Estimating Heterogeneity - Simulations and Placebo tests by size of Surnames	142
B.6	Derivation of OLS and IV coefficient when total consumption is proxied with c^*	147
B.6.1	Potential OLS bias	149
B.6.2	Bounding the IV bias	151
B.7	Estimates of Propensity Scores and Inverse Mills Ratio	153

LIST OF FIGURES

1.1	Average Marginal Tax rate on Income: Optimal vs Current	4
1.2	Illustration of the axioms of <i>poverty-reduction transfer</i> 1, 2 and 3	13
1.3	Illustration of Proposition 1, statement 1.	14
1.4	Illustration of Proposition 1, statement 3.	15
1.5	Illustration of Corollary 1	26
1.6	Illustration of Proposition 4.	30
1.7	How to define a better tax schemes than the one that lifts all agents out of poverty: the role of \tilde{R}	36
1.8	Calibration of the optimal marginal tax rates according to different slopes of the poverty line.	37
1.9	Calibration of the optimal consumption schedule according to different slopes of the poverty line.	38
1.10	Calibration of the optimal marginal tax rates according to different $\bar{\theta}$	40
1.11	Calibration of the optimal marginal tax rates according to different levels of the poverty line.	43
1.12	Calibration of the optimal consumption schedule according to different levels of the poverty line.	44
2.1	Last-name level intergenerational correlation in housing consumption.	59
2.2	Engel Curve for Housing Consumption	78
2.3	Engel Curve for Housing Consumption (Separately by Region)	81
2.4	Engel Curve for Housing Consumption (Separately by Race)	81
2.5	Engel Curve for Housing Consumption (Separately by Decade)	82
2.6	Surname level correlation between 1940 housing consumption and fraction of homeowners in 2012.	83
A.1	Illustration of the existence of R°	98
A.2	Illustration of the proof of Proposition 3.	102
A.3	Calibration of the optimal marginal tax rates according to different slopes of the poverty line.	119
A.4	Calibration of the optimal consumption schedule according to different levels of the poverty line.	120
A.5	Average Marginal Tax Rates - Single Filers - \$0-\$80,000	122
A.6	Average Marginal Tax Rates - Single Filers - \$0-\$700,000	123
A.7	Average Marginal Tax Rates - Joint Filers - \$0-\$700,000	123
A.8	Average Marginal Tax Rates (Zoom) - Single Filers without Children	124
A.9	Average Marginal Tax Rates (Zoom) - Single Filers with 1 Child	124
A.10	Average Marginal Tax Rates (Zoom) - Single Filers with 2 Children	125
A.11	Average Marginal Tax Rates (Zoom) - Joint Filers without Children	125
A.12	Average Marginal Tax Rates (Zoom) - Joint Filers with 1 Child	126
A.13	Average Marginal Tax Rates (Zoom) - Joint Filers with 2 Children	126
A.14	Double Pareto Lognormal estimated on CPS data (2013-2015) versus the empir- ical CDF.	128

A.15	Double Pareto Lognormal estimated on CPS data (2013-2015) versus the empirical CDF.	128
A.16	Double Pareto Lognormal estimated on CPS data (2013-2015) versus empirical histogram.	129
B.1	Intergenerational correlation in housing consumption - mostly white versus mostly black last-names.	132
B.2	Intergenerational correlation in housing consumption - Regional Heterogeneity for mostly white last names	133
B.3	Coefficient of regression of Education on log Housing Consumption - surname-versus individual-level by size of last name	146
B.4	Coefficient of regression of Occupational Score on log Housing Consumption - surname- versus individual-level by size of last name	147

LIST OF TABLES

2.1	Surname-level correlations in Housing Consumption (By race)	59
2.2	Surname-level correlations in Housing Consumption (By region)	60
2.3	Placebo test: Immigration, Geography and race	67
2.4	Placebo test: Education, Work and Family	68
2.5	Placebo Test: Comparison of true OLS with (covariates adjusted - race only) between last-names OLS	71
2.6	Placebo Test: Comparison of true OLS with (covariates adjusted -) between last-names OLS	72
2.7	Long-run Mobility Estimates by Race	73
2.8	Mobility Estimates (β)	75
2.9	Mobility Estimates (α)	76
2.10	Demand estimations	79
B.1	Mobility Estimates adjusted for selection (β)	134
B.2	Mobility Estimates adjusted for selection (α)	134
B.3	Average 1940 Housing Consumption (Per Demography)	135
B.4	Estimates of Average 2012 Housing Consumption (Per Demographics)	135
B.5	Comparison of True and Between last-name regression of Years of Education on Log Housing Consumption (Slope Coefficient by demographic group - 1940)	143
B.6	Simulations: Estimates versus true parameters (IGC)	144
B.7	Simulations: Estimates versus true parameters (α constant)	144
B.8	Simulations: Estimates versus true parameters (IGC) - rare names only	145
B.9	Simulations: Estimates versus true parameters (IGC) - Large names only	146

ACKNOWLEDGMENTS

Any errors are my own. I am grateful to my advisors whose talent, passion, generosity, and drive is an inspiration: To Marianne Bertrand- thank you for always finding the time to meet with me, for asking tough questions, for engaging with me at the very beginning of my journey, and helping shape my research interests. To Stéphane Bonhomme- thank you for your genuine interest in my work, your deep econometric and economic insights, and for always helping me find ways to improve my paper. To Magne Mogstad- thank you for always challenging me, for your sharp advice, and for providing me with a path forward when doors seemed to be closing on me. To all three: thank you for believing in me and always making the time. Over the years, I have also greatly benefited from encounters with brilliant economists that I consider to be mentors. At the very beginning, Michel De Vroey encouraged me to pursue a career in this profession. My first steps in academic research were supervised by François Maniquet whose research on fairness convinced me that economists can look beyond efficiency without sacrificing rigor and intellectual consistency. His mentoring - associated with the stimulating environment of our PoResp team - motivated me and his continued mentorship and friendship after I emigrated to the US were invaluable. Glen Weyl is an inspiration for his creativity and his tackling of the biggest challenges faced by our society. Thank you for believing in me and teaching me to pursue bolder and more creative research agendas. I feel extremely lucky to have shared many long conversations with Raaj Sah, which have enriched my research and whose wisdom is an inspiration. I have learned much from Casey Mulligan and his ability to break down complex problems into simple yet deep price theoretic insights. I am extremely grateful for his continued support. Finally, Erik Hurst has taught me much about the profession and was always available to discuss research ideas. I also thank him for giving me and Jung access to the 1940 Census data used in the second chapter of this thesis. I am indebted to my co-authors that I have learned so much from and who have been incredible friends. To Carolyn- thank you for always pushing us to think big, for your self-discipline, without which we would not be close to finishing the

paper, and for your support through thick and thin. Your passion, tenacity, and talent are an inspiration. To Jung- thank you for the many conversations and coffee chats that have lead to many good (and many bad) ideas, and your ability to lift me up in discouraging moments. Your creativity, passion for economics, ambitious ideas, and intelligence are an inspiration. I am also lucky to have crossed path with many colleagues - and now friends - that I have learned much from and who have made my experience in this profession fun and rewarding. Thank you Ben Griffy, Benoit Decerf, Chiara Fratto, Dan Alexander, Jack Mountjoy, Jan Kozak, Jean-Yves Gérardy, Jorge Luis García-Menéndez, Martin Van der Linden, Mathieu Cloutier, Nick Tsivanidis, Oliver Browne, Paymon Khorrami, Victor Gay and many others. Thank you Manasi Deshpande for helping me gain access to the InfoGroup data and thank you Ingvil Gaarder for the present and future collaboration. I look forward to continuing the work on our project. I also want to thank the great teachers and economists whose classes have inspired me and whose comments have improved my research: Gary Becker, Kevin Murphy, Steve Durlauf, James Heckman, Dan Black and Kerwin Charles. I thank the staff at University of Chicago - in particular Julie Less, Luisana Romero, Otilie Young and Robert Herbst - who helped me through problems, big and small. Finally, I want to thank my two “brothers in arms” David Méndez Yépez and Sevan Holemans who were an integral part of this journey and have stimulated much questioning of the world that is at the root of my journey into academic research.

I would like to thank my “new” American and Lebanese family - Mom, Dad, Anthony, Colette, Taita, and my new aunts and uncles - whose warm welcome and support has created the conditions for me to feel at home in a foreign country and whose understanding that “a good fire requires good wood” has made this life project possible. Colby - thank you for always letting me know where there is free lunch. To my dear friends and family from Belgium, Chicago and everywhere: thank you for always lifting my spirits up, making me laugh, keeping me grounded, and for always having my back.

Finally, I would like to especially thank Lorraine, Jo, Isabelle, Marie-Julie, Jean, and Rosalie - my brothers and sisters - for the love and unconditional support that they have selflessly given me throughout this long road. Growing up alongside you was the greatest privilege and your constant presence, despite the distance and time difference, have powered me through the finish line. To my parents - Paps and Mams - thank you for the unconditional love, for having instilled intellectual curiosity in me, and for always making sure - without counting - that I get every opportunity to realize my dreams. I also want to thank Papy and Nicky for their warm presence at my side and for always believing in me. Janine, thank you for being the happiness of my life. Finding such joy at home was more than necessary to power through the completion of this thesis. Thank you for keeping me laughing even through the most challenging moments- and humoring my poor jokes. Thank you for the sacrifices that you have made and being so committed to us despite the uncertainties we have faced as I complete my academic career. To our future little one: thank you for the immense happiness that you have already given us and for the adventure that lies ahead. This one is for you, as well as, Janine, Lorraine, Paps, Mams, Papy, Nicky, Jo, Isa, Marie-Julie, Jean and Rosalie. Love.

ABSTRACT

This dissertation is a collection of two essays on inequality. The first chapter takes a normative approach to inequality. We start from the normative assumption that not all inequalities are unjust. In particular, we assume that public policy ought to correct inequalities that are caused by differences in skills insofar as they generate poverty but ought to remain neutral towards inequalities generated by different individual preferences over labor choices. We construct social preferences that are consistent with this view in addition to satisfying Pareto efficiency and we derive implications for the taxation of labor income. We calibrate the optimal tax formula to the US and compare it to the current tax and transfer system. One of the most striking features of the optimal tax schedule under these social preferences is the desirability of negative marginal tax rates on low-incomes. Overall, the US tax and transfer system shares many features with the tax scheme that would be designed by a planner whose objective combines reducing poverty with preference responsibility.

The second chapter takes an intergenerational perspective on inequality. We measure intergenerational mobility in housing consumption - across surnames - in the US for the period 1940-2012. We translate, under stated assumptions, the figure into an estimated *family-level* intergenerational elasticity of total consumption of 0.6 – 0.67 between two successive generations. We also document that blacks have much lower IGE than whites, the Northeast has higher IGE than the South and Midwest, and in particular, the black-white gap in IGE is concentrated in the Northeast. Some of these patterns contrast with the previous literature’s analysis of IGE in labor income. We discuss possible mechanisms behind the level of - and the heterogeneity in - the estimated IGE of consumption.

CHAPTER 1

PREFERENCE RESPONSIBILITY VS POVERTY

REDUCTION IN THE TAXATION OF LABOR INCOMES

(Joint work with François Maniquet)

[...] a pivotal part of this economic plan is increasing the earned-income tax which, more than anything else we could do, will reward work and family and responsibility [...] this will be the first time in the history of our country when we will be able to say that if you work 40 hours a week and you have children in your home, you will be lifted out of poverty.”

President Bill Clinton, July 29, 1993 (shortly before expanding the Earned Income Tax Credit)

1.1 Introduction

Income taxation in general, and labor income taxation in particular, is the ultimate policy instrument to reduce income inequality. Since Mirrlees (1971)’s seminal contribution, most of the literature has embraced the view that full income inequality is not a legitimate objective. The literature on optimal taxation is mainly welfarist: the utilitarian objective, especially when individual utility functions are concave, has long been seen appropriate to reduce inequality without eliminating it. Public tax policy discussions, on the other hand, rarely revolve around utilitarianist principles and often invoke some notion of “fairness” or other non-welfarist considerations.¹ In addition, Weinzierl (2014) finds survey-based evidence that most Americans reject important implications of Utilitarianism.

An alternative approach to interpersonal comparisons of utilities has been recently proposed, after the introduction of the ethics of responsibility into normative debates (see de-

1. See, for instance, the quotation of President Bill Clinton above.

tailed accounts in Roemer, 1998, and Fleurbaey, 2008). In a nutshell, the ethics of responsibility is grounded on the assumption that not all inequalities are unjust. As a consequence, the social objective should be to identify and eliminate the unjust inequalities and to remain neutral towards the other inequalities. We provide a very short introduction to this literature and related developments in optimal tax theory in the next section.

In this paper, we define a social objective function compatible with the ethics of responsibility. More specifically, we combine a requisite that the social objective be neutral towards inequalities caused by different labor time choices with the goal of reducing poverty. We also impose that social preferences satisfy Pareto efficiency. After discussing the potential conflicts between these three goals and showing how they can be reconciled, we characterize a family of social welfare functions consistent with these three requirements.² We then derive the associated formula for optimal labor income taxation and calibrate them to the US economy.

There are two main lessons to draw from our paper. First, it is possible to fully characterize the optimal income tax derived from an unconventional social objective that embodies notions of fairness. While the literature on optimal taxation has successfully obtained and calibrated expressions detailing the optimal tax as a function of behavioral responses, distribution of types and a set of weights capturing the normative preferences of society, it has provided very little guidance on how to choose those weights. On the other hand, there is a long tradition in the social choice literature discussing interpersonal comparisons of utilities and the implicit normative judgements inherent in comparing well-being across individuals. This literature provided us not only with social welfare functions respectful of individual preferences but also highlights how different specific definitions of fairness characterize the set of social preferences compatible with each of these definitions (see, among others, Roemer 1998 or Fleurbaey and Maniquet 2011). Though some criteria for the taxation of labor

2. To be precise: we derive a "local" property around the poverty line that all social welfare functions satisfying our axioms have to satisfy.

income have been previously derived from these social objectives, this literature has not defined the precise formula of the tax function leading to somewhat of a disconnect with the optimal tax literature (in addition to not allowing for precise calibrations). As a result, we view this study as an example of how to connect these two literatures. The logical steps that we follow in order to solve for the optimal tax formula can be replicated with a diversity of social welfare functions from the literature in social choice.³

The second main lesson of our paper is that the optimal tax function that is consistent with our three normative goals shares some salient features of the current income tax schedule in the US. The optimal tax schedule exhibits negative marginal tax rates on low incomes and a steep jump upwards in marginal tax rates that are somewhat unconventional in the optimal taxation literature and yet approximate some salient features of the current US income tax.

As can be seen on figure 1.1, the optimal tax corresponding to the combined objective of Pareto efficiency, responsibility and poverty reduction exhibits zero and negative marginal tax rates on low incomes as well as a discrete jump to positive marginal tax rates around \$12,000. This jump approximately corresponds to the phaseout of the Earned Income Tax Credit (EITC) and various welfare programs in the current US tax system. Depending on the specific parametrization of the poverty line, optimal marginal tax rates on high-incomes may be lower than would be implied by setting the marginal welfare weights to zero in the upper-tail (a common normative assumption in optimal tax literature under Utilitarianism).⁴ We discuss more precisely the similarities and differences between our calibrations and the current US tax schedule in section 8. We note here that our approach roughly rationalize the current shape of the tax function in the US. The existence of non-positive marginal tax rates on low-income and the steep jump towards positive rates is a very general result under

3. A lot of these social welfare functions share similarities in the sense that most take the form of a *maximin* in some index of well-being. In theory, our approach can be replicated to these social preferences. It is important to emphasize that though these functions share similarities, they stem from very diverse normative principles and lead to a wide range of aversion to inequality in income.

4. Papers such as Hendren (2014) and Lockwood and Weinzierl (2016) have previously noted that existing tax policies are consistent with less redistributive preferences that is traditionally assumed in the literature.

Average Marginal tax rates on wage income: "Optimal" vs Current US tax

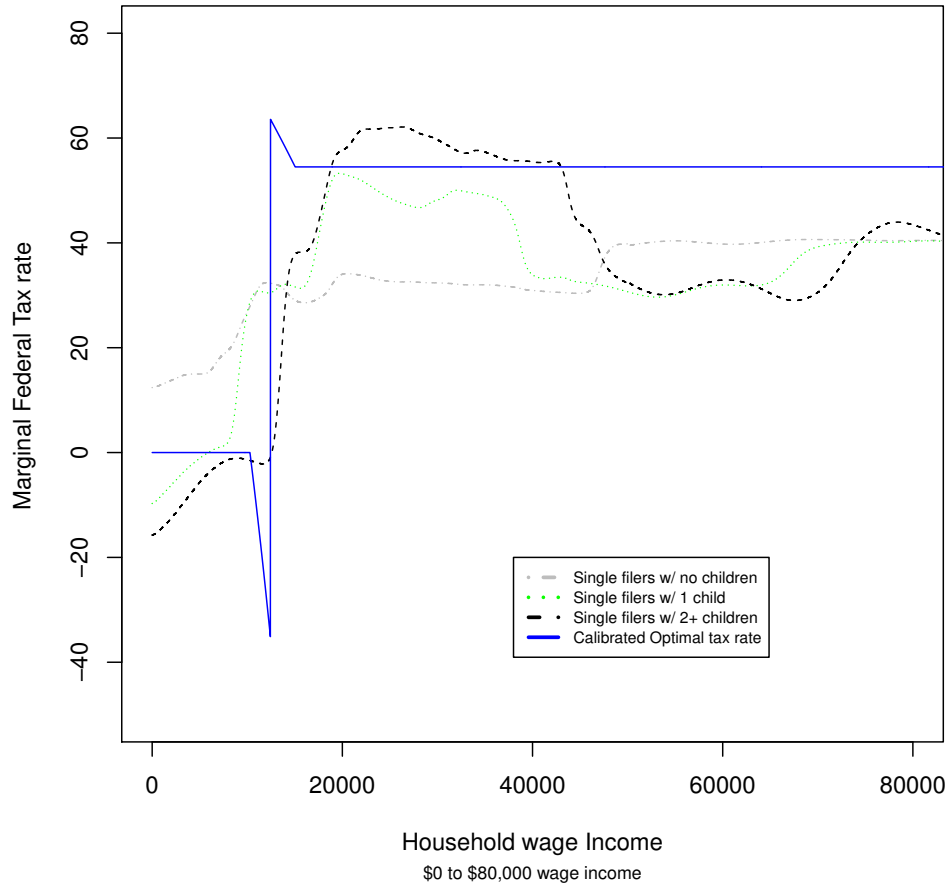


Figure 1.1: Average Marginal Tax rate on Income: Optimal vs Current

Note: The solid line depicts the optimal labor income tax according to our social objective where the poverty line was parametrized as having a slope of \$6/hour. Details on the calibration and estimation of the current tax system using CPS data can be found in Appendix A3.

our normative objective and is independent from the specific parametrizations of the poverty line, of the individuals' preferences and of the distribution of types.

Other authors have reached similar conclusions about the optimality of non-positive tax rates on low earning levels from different viewpoints. First, negative marginal tax rates are optimal when labor responses on the extensive margin are important either because of a fixed cost of labor participation (see Saez, 2002, Diamond, 1980 or Blundell and Shephard, 2012), or a present bias (see Lockwood, 2016). Second, negative marginal tax rates can be justified under the maximization of a utilitarian objective when preferences differ and social weights are a function of these preferences (see Boadway, Marchand, Pestieau and Racionero, 2002). Third, some studies of social preferences that embed fairness principles have also derived the optimality of negative marginal tax rate. Saez and Stantcheva (2016) as well as Maniquet and Neumann (2016) find negative marginal rates on earnings that lead to an after-tax income below the poverty line. Related to our approach, Fleurbaey and Maniquet (2006, 2007) obtain non-positive marginal tax rates from a social objective that combines responsibility for agents' preferences with compensation for their skills. Importantly, Fleurbaey and Maniquet (2006, 2007) do not succeed in deriving the precise formula of the optimal tax function, which we do with our social objective.

Finally, we make an additional contribution to the literature in showing that it is possible to define a social objective combining the three goals of poverty alleviation, responsibility and efficiency. Surprisingly enough, the most basic tension among these three goals is the conflict between poverty alleviation and Pareto efficiency. If being poor means consuming a bundle of goods below some poverty line, the main issue comes from the fact that agents may prefer bundles below the poverty line to bundles above it. This issue echoes analyses of optimal labor income tax that reduces poverty when poverty is merely defined in terms of income (see, for instance, Kanbur, Keen and Tuomala, 1994a and b, and Wane, 2001). Once the definition of "being poor" is adjusted so as to become consistent with Pareto efficiency, it can easily be further adjusted to be also compatible with the requirement that

no redistribution should take place among equally-skilled agents.

To summarize, in this paper, we begin with the definition of a few axioms that embody a specific definition of fairness. We proceed to deriving the associated optimal income tax formula, calibrate and compare it to the current tax policy in the US. We conclude that our set of axioms roughly rationalizes the current tax system. The rest of the paper is organized as follows. In the next section, we give a very brief introduction of the ethics of responsibility and its application to optimal labor income taxation. In Section 3, we define the model and the basic properties. In Section 4, we discuss the possible ways of defining the goal of poverty alleviation and we show that one of them is compatible with both goals of Pareto efficiency and responsibility for agents' preferences. In Section 5, we derive the general shape of any social objective combining the three goals we are interested in. The remaining of the paper consists in studying the second-best tax schemes that maximize our social objective. In Section 6, under extremely general assumptions on the distribution of types in the economy, we derive a useful yet very partial property of the optimal tax. We also derive the optimal tax formula for low incomes under additive separability of preferences and an assumption that disciplines multi-dimensional heterogeneity. In section 7, under specific but customary assumptions on preferences and with a particular focus on two different versions of the social objective, we derive the formula of the optimal tax for the entire income distribution. In Section 8, we estimate the distribution of types using Current Population Survey (CPS) microdata and calibrate the tax formula to the U.S. economy. We also use the CPS data and NBER's TAXSIM to obtain a description of the current tax system and put our results into perspective.⁵ In Section 9, we give some concluding comments.

1.2 The Ethics of Responsibility: A Very Short Introduction

The ethics of responsibility is the social norm that consists in holding individuals responsible for the share of their outcome that depends on specific parameters chosen by society. For

5. Regarding TAXSIM, see Feenberg and Coutts (1993).

instance, it is well-known that agents' probability of finding a job and their pay depend on their beauty. In spite of that, democratic societies refuse to tax handsome people to compensate the ugly. That means that society holds agents responsible for their beauty.

There may be several reasons for which a society wishes to hold people responsible for some individual characteristic. For instance, responsibility for one's beauty is related to the private ownership of self. Another reason to hold agents responsible for some characteristic is that the value of that characteristic is under the control of the individual, such as the way one spends their own money.

In debates on the good taxation system, it is common to argue that agents should be held responsible for the share of their income that is related to their effort. That view may be grounded on the norm of just desert, or on private ownership of self, or on a form of meritocracy. On the other hand, some also argue that agents should not be held responsible for the share of their income that is related to their productivity. That view may be grounded on the norm that productivity is to a large extent a matter of luck and inequality related to luck should be eliminated.

The literature on optimal tax has recently studied how tax rates can be deduced from maximizing a notion of social welfare that combines responsibility for one's labor time choices and compensation for one's productivity. The large spectrum of resulting tax schemes goes from a basic income proposal to the mere *laissez-faire* (absence of redistribution). Between those extreme schemes, a lot of intermediary schemes recommend a marginal tax rate on low income equal to zero or even negative (see Fleurbaey and Maniquet, 2017, for a survey).

When the notion of responsibility is applied to the design of labor income tax schemes, a common feature of many of these approaches is that there should not be any redistribution among equally skilled agents.⁶ The consequences of this objective has recently received a lot of attention (see, among others, Fleurbaey and Maniquet, 2006, 2007, Lockwood and Weinzierl, 2015, Weinzierl, 2014).

6. The approach developed in Roemer (1998), however, does not follow that road.

In spite of having revived the interest in egalitarian views and in spite of its practical achievements, the ethics of responsibility, however, also raises ethical issues. One of them is its potential conflict with the fight against poverty. Indeed, the limit to redistribution among equally skilled agents may imply that some of them end up with a very low consumption level. If society values that nobody lives with material resources below some threshold, that is, if society fights against poverty, there may be a clash between poverty alleviation and responsibility.

1.3 The Model and the Basic Axioms

There are two goods, labor time, denoted ℓ , and consumption, denoted c . A bundle (of goods) is a pair $z = (\ell, c) \in X = \mathbb{R}_+^2$. There is a finite set N of agents. Each agent $i \in N$ is characterized by their productive skill, or wage, $w_i \in [\underline{w}, \infty)$, and their preferences $R_i \in \mathcal{R}$ over bundles. Wages are assumed to be bounded from below. Preferences are assumed to be continuous, increasing in consumption, decreasing in labor time, and convex. An economy is a list of characteristics $E = (w_N, R_N) \in \mathcal{E} = ([\underline{w}, \infty)_+ \times \mathcal{R})^N$, one pair of characteristics for each agent. An allocation is a list of bundles $z_N = (z_i)_{i \in N}$, one bundle for each agent.

We are interested in ranking allocations as a function of the parameters of the economy. Formally, a Social Ordering Functions, in short a SOF, is a function R , such that for each economy $E \in \mathcal{E}$, $R(E)$ is an ordering on the set of allocations X^N . For two allocations $z_N, z'_N \in X^N$, we write $z_N R(E) z'_N$, resp. $z_N P(E) z'_N$, $z_N I(E) z'_N$, to denote that z_N is socially as good as z'_N , resp. strictly better, indifferent.

The following notation will prove useful. For any preferences $R \in \mathcal{R}$ and set $A \subset X$, we write $m(R, A)$ to denote the set of best bundles in A according to R :

$$m(R, A) = \{z \in A \mid \forall z' \in A : z R z'\}.$$

Note that in the case $m(R, A)$ contains more than one bundle, all bundles of $m(R, A)$ are

indifferent for R . For any $z = (\ell, c) \in X$ and $w \in [\underline{w}, \infty)$, we write $B(z, w)$ to denote the budget of slope w going through z :

$$B(z, w) = \{z' = (\ell', c') \in X \mid c' - w\ell' \leq c - w\ell\}.$$

It will be useful below to abuse on notation and write $B(z, w) + \Delta$ to denote the budget set obtained by translating $B(z, w)$ by an amount Δ , that is $\{z' = (\ell', c') \in X \mid c' - w\ell' \leq c - w\ell + \Delta\}$.

Finally, for any $z \in X$, $R \in \mathcal{R}$ and $w \in [\underline{w}, \infty)$, we write $IB(z, R, w)$ to denote the implicit budget of an agent (R, w) at z , that is the budget of slope w that leaves this agent indifferent between z and maximizing over that budget:

$$IB(z, R, w) = B(z', w) \text{ such that } z' I z \text{ and } z' \in m(R, I(z, R, w)).$$

We complete this section with the definition of three well-known axioms that we impose on our SOF's. Strong Pareto is the requirement that if everyone considers his bundle in z_N at least as good as in z'_N , then z_N must be socially as good as z'_N . If, in addition, at least one agent strictly prefers his bundle in z_N , z_N must be deemed a strict social improvement.

Axiom 1 *Strong Pareto*

For all $E = (w_N, R_N) \in \varepsilon$, all $z_N, z'_N \in X^N$,

$$[z_i R_i z'_i, \forall i \in N] \Rightarrow [z_N R(E) z'_N]$$

$$\text{and } [z_i R_i z'_i, \forall i \in N, \text{ and } \exists j \in N : z_j P_j z'_j] \Rightarrow [z_N P(E) z'_N]$$

A well-known implication of Strong Pareto will be used in this paper, namely Pareto indifference. It is the requisite that if all agents are indifferent between two allocations, then social preferences must also be indifferent.

Axiom 2 *Pareto indifference*

For all $E = (w_N, R_N) \in \mathcal{E}$, and $z_N = (l_N, c_N), z'_N = (l'_N, c'_N) \in X^N$, if $z_i I_i z'_i$ for all $i \in N$, then $z_N I(E) z'_N$

We now define the axiom capturing the ethics of responsibility. It follows the goal that two agents having the same wage be treated identically by the tax and transfer system, that is, they should ideally receive the same subsidy or pay the same tax. More formally, the requirement is that if two agents having the same wage are maximizing over different budgets, a budget-inequality-reducing transfer among them should be a social improvement.

Axiom 3 *Equal-wage transfer*

For all $E = (w_N, R_N) \in \mathcal{E}$, $z_N, z'_N \in X^N$, if there exists $j, k \in N$ and $\Delta \in \mathbb{R}_+$ such that $w_j = w_k$,

$$\begin{aligned} z'_j &\in m(R_j, B(z'_j, w_j)), z_j \in m(R_j, B(z_j, w_j)), \\ z'_k &\in m(R_k, B(z'_k, w_k)), z_k \in m(R_k, B(z_k, w_k)), \end{aligned}$$

and

$$B(z'_j, w_j) = B(z_j, w_j) - \Delta \supseteq B(z'_k, w_k) = B(z_k, w_k) + \Delta,$$

and for all $i \neq j, k$: $z_i = z'_i$, then $z'_N P(E) z_N$.

In the complete absence of taxation, that is at a *laissez-faire* allocation, all agents are treated identically by the tax system, as they do not receive nor pay anything. Moreover, the resulting allocation is Pareto efficient. This immediately shows that *strong Pareto* and *equal-wage transfer* are compatible with each other. The drawback of *laissez-faire*, however, is that some agents may end up in material deprivation, and the more so if they have small wages. The main goal of this paper is to study the consequences of combining *strong Pareto*

and *equal-wage transfer* with other axioms capturing the idea that society would like to avoid that agents end up at too low a material standard level. We introduce such axioms in the next section.

1.4 Poverty Reduction

The present paper suggests to introduce the objective of poverty reduction as an additional fairness requirement and derive social preferences consistent with this objective. This first requires to introduce a poverty line in the model. We indeed assume that there is a thin, connected and strictly increasing set of bundles in the consumption set that captures the desire by society to let all agents live a materially decent life. Let $PL \subset X$ satisfy the following properties:

- for all $z = (\ell, c), z' = (\ell', c') \in PL$, either $z = z'$ or $\ell < \ell' \Leftrightarrow c < c'$,
- for all $\ell \in \mathbb{R}$, there exists $c \in \mathbb{R}$ such that $(\ell, c) \in PL$,
- the set of bundles above PL is convex.

One comment is in order. Our poverty line is strictly increasing, that is, the minimal consumption level that makes a life materially decent is assumed to increase with the amount of labor performed by the agent. This can be interpreted in two ways. First, as labor time rises the individual's basic needs increase: more food is required by the organism to afford those efforts, clothes wear out faster,... Second, leisure can be considered as a basic need so that less leisure needs to be compensated by more income (we may think for instance of the costs parents incur for hiring someone to take care of their children when they work full time).

To simplify the exposition, we will abuse notation and write $z = (\ell, c) > PL$, for any $z \in X$, to describe the situation in which there is $(\ell, c') \in PL$ such that $c > c'$, and $A > PL$, for any $A \subset X$, to describe the situation in which $z > PL$ for all $z \in A$. It will also be useful

to refer to indifference curves as to $I(z, R)$, that is, for any $z \in X$, $R \in \mathcal{R}$,

$$I(z, R) = \{z' \in X \mid z' I z\}.$$

The poverty line being assumed already specified, we can define an axiom reflecting the objective of poverty reduction. Basically, we state that a lump-sum transfer of consumption from a non-poor to a poor individual must be a social improvement. We call such an axiom:

Axiom 4 *Poverty-reduction transfer 1*

For all $E = (s_N, R_N) \in \mathcal{E}$, all $\Delta \in \mathbb{R}$, all $j, k \in N$, all $z_N = (l_N, c_N)$, $z'_N = (l'_N, c'_N) \in X^N$ such that $l_j = l'_j = l_k = l'_k$, $c'_j = c_j - \Delta$, $c_k = c'_k + \Delta$ and for all $i \neq j, k : z_i = z'_i$,

$$[z'_j > PL > z'_k] \Rightarrow [z'_N R(E) z_N].$$

Axiom 5 *Poverty-reduction transfer 2*

For all $E = (s_N, R_N) \in \mathcal{E}$, all $\Delta \in \mathbb{R}$, all $j, k \in N$, all $z_N = (l_N, c_N)$, $z'_N = (l'_N, c'_N) \in X^N$ such that $l_j = l'_j = l_k = l'_k$, $c'_j = c_j - \Delta$, $c_k = c'_k + \Delta$ and for all $i \neq j, k : z_i = z'_i$,

$$[I(z'_j, R_j) > PL > z'_k] \Rightarrow [z'_N R(E) z_N].$$

Axiom 6 *Poverty-reduction transfer 3*

For all $E = (s_N, R_N) \in \mathcal{E}$, all $\Delta \in \mathbb{R}$, all $j, k \in N$, all $z_N = (l_N, c_N)$, $z'_N = (l'_N, c'_N) \in X^N$ such that $l_j = l'_j = l_k = l'_k$, $c'_j = c_j - \Delta$, $c_k = c'_k + \Delta$ and for all $i \neq j, k : z_i = z'_i$,

$$[I(z'_j, R_j) > PL > I(z'_k, R_k)] \Rightarrow [z'_N R(E) z_N].$$

It should be transparent that the first version of the axiom is logically stronger than the second one, which is logically stronger than the third one. Moreover, we have the following (in)compatibilities with the other axioms.

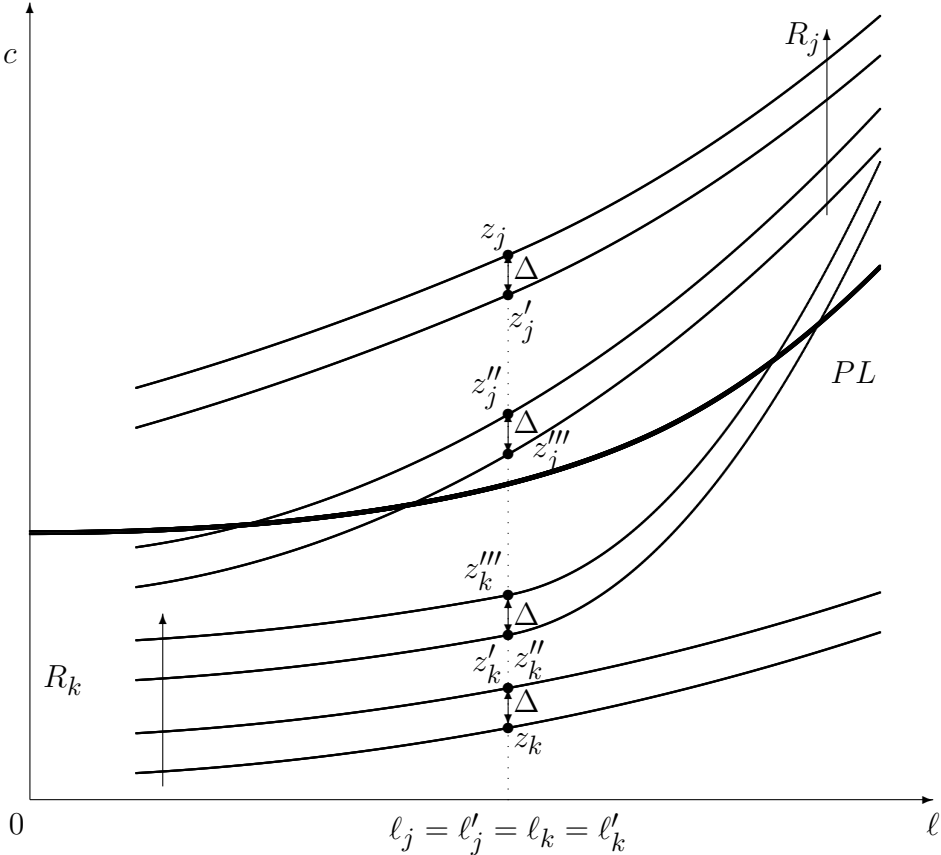


Figure 1.2: Illustration of the axioms of *poverty-reduction transfer 1, 2 and 3*

Note: According to *poverty-reduction transfer 1*, any transfer of Δ from either z_j or z''_j to either z_k or z''_k is a strict social improvement; according to *poverty-reduction transfer 2*, any transfer of Δ from z_j to either z_k or z''_k is a strict social improvement; according to *poverty-reduction transfer 3*, only a transfer of Δ from z_j to z_k is a strict social improvement.

Proposition 1

1. No SOF satisfies *weak Pareto* and *poverty-reduction transfer 1*.
2. There are SOF's satisfying *Pareto indifference* and *poverty-reduction transfer 2*.
3. No SOF satisfies *Pareto indifference*, *equal-wage transfer* and *poverty-reduction transfer 2*.
4. There are SOF's satisfying *strong Pareto*, *equal-wage transfer* and *poverty-reduction transfer 3*.

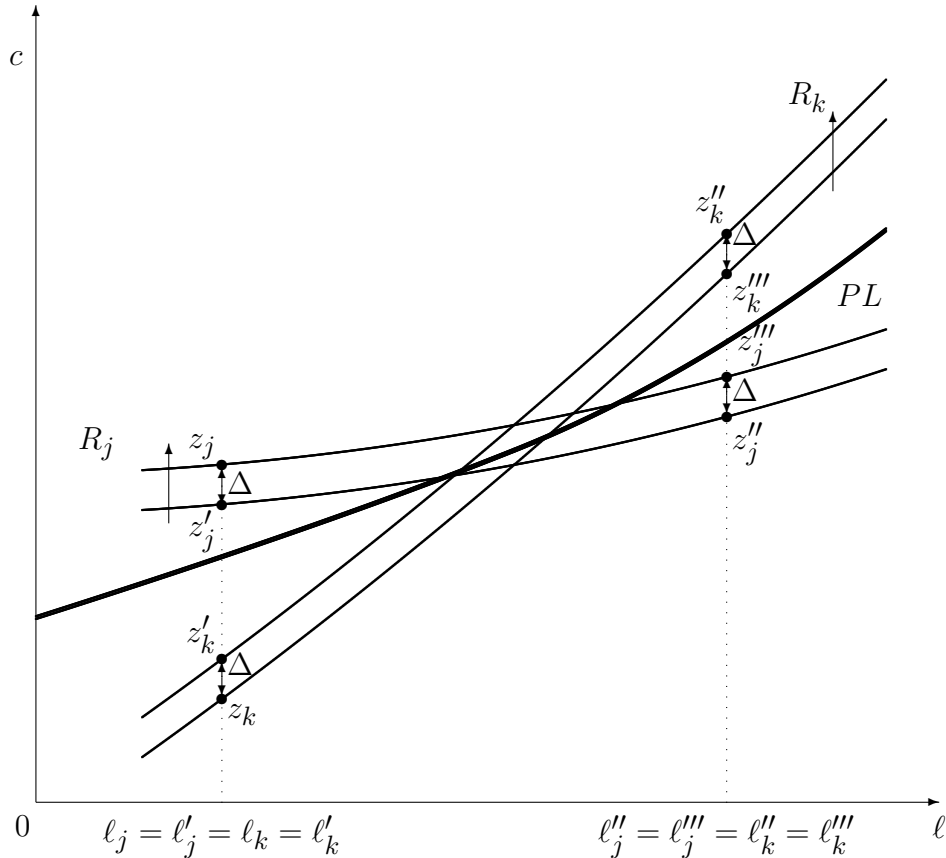


Figure 1.3: Illustration of Proposition 1, statement 1.

The proof of the four statements in Proposition 1 is rather intuitive. We present it in the text. Statement 1 is reminiscent of the proof of the impossibility of a Paretian egalitarian in Fleurbaey and Trannoy (2003). It is illustrated in Fig. 1.3. The key feature of the figure

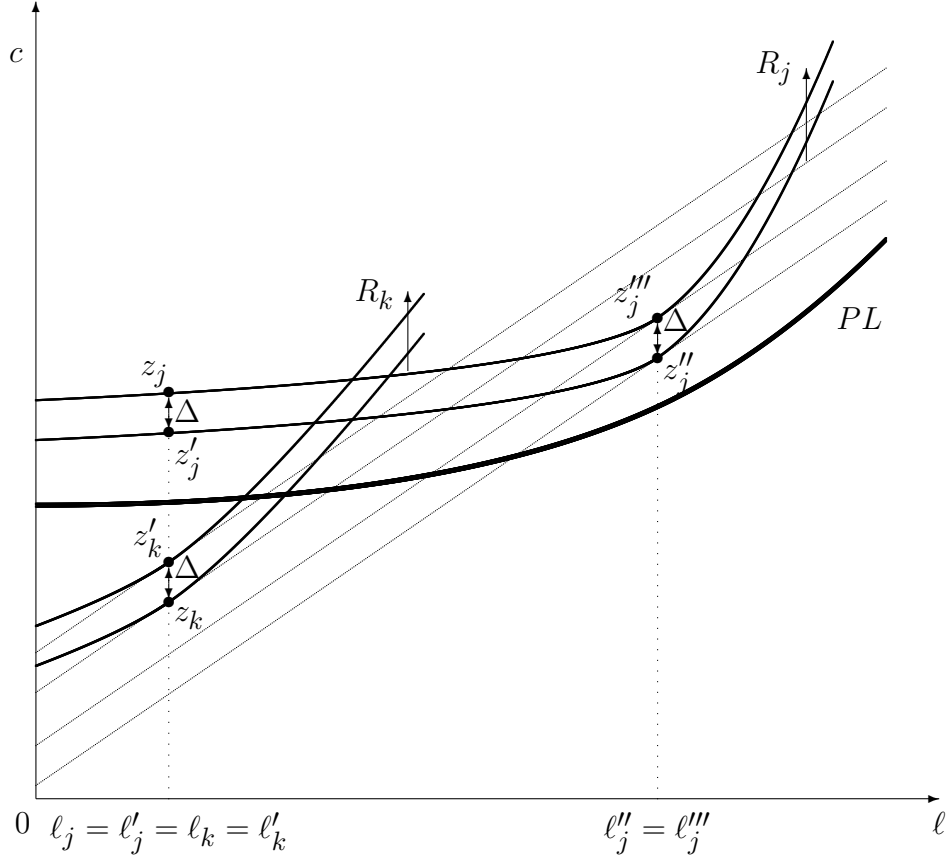


Figure 1.4: Illustration of Proposition 1, statement 3.

is that both agents j and k have preferences that cross the poverty line, but in opposite direction. As a result, when they both choose a low labor time, say $\ell_j = \ell_k$, then agent j has a consumption level above the poverty line whereas k 's consumption is below the poverty line. As a result, (z'_j, z'_k) is socially preferred to (z_j, z_k) . When they both choose a large labor time, say $\ell''_j = \ell''_k$, at the same satisfaction levels, k 's consumption level is above the poverty line contrary to j 's consumption level. As a result, (z'''_j, z'''_k) is socially preferred to (z''_j, z''_k) . By Pareto indifference, which is implied by strong Pareto, (z_j, z_k) is socially indifferent to (z'''_j, z'''_k) and (z'_j, z'_k) is socially indifferent to (z''_j, z''_k) , creating an intransitivity.

The proof of statement 3 is illustrated in Fig. 1.4. Poverty-reduction transfer 2 requires that (z'_j, z'_k) be socially preferred to (z_j, z_k) . Equal-wage transfer requires that (z'''_j, z_k) be socially preferred to (z''_j, z'_k) . Finally, Pareto indifference requires that (z_j, z_k) be socially indifferent to (z'''_j, z_k) and (z'_j, z'_k) to (z''_j, z'_k) , creating an intransitivity.

Statement 3 is at the heart of the ethical dilemma that this paper addresses. Agents j and k have the same wage, represented in the figure by the slope of their budgets. The responsibility goal pushes towards no redistribution among them two. However, agent k enjoys a bundle below the poverty line, whereas agent j 's indifference curve is everywhere above the poverty line. This is why even the weak poverty-reduction transfer 2 requires to prefer a redistribution between the two agents when they don't work hard, that is, when their labor times are $\ell_j = \ell_k$. The two transfers go in opposite directions, whereas agent j is indifferent between the two involved bundles, hence the incompatibility.

Statements 2 and 4 are proven by way of examples. We begin with statement 2. We first need to introduce some terminology. Let \mathcal{U} denote the set of utility functions representing preferences in \mathcal{R} . This example works by choosing a utility representation for each individual preference relation and apply the leximin aggregator to the utility vector corresponding to the evaluated allocations.⁷ Let $U : \mathcal{R} \rightarrow \mathcal{U}$ be a representation function satisfying the property that for each $R \in \mathcal{R}$, $U(R)$ is a utility function representing R and if we write $u = U(R)$ then for all $z \in X$

$$u(z) = 1 \Leftrightarrow z \text{ Im}(R, PL), \quad (1.1)$$

that is, all agents have the same utility level at their preferred bundle on the poverty line (and this utility level is equal to 1). Social preferences R^U are defined as follows: for all $E = (w_N, R_N) \in \mathcal{E}$, all $z_N, z'_N \in X^N$,

$$z_N R^U z'_N \Leftrightarrow (u_i(z_i))_{i \in N} \geq_{lex} (u_i(z'_i))_{i \in N}$$

where for all $i \in N$, $u_i = U(R_i)$. The fact that R^U satisfies *strong Pareto* follows from u_i

7. The leximin criterion consists in applying the maximin criterion lexicographically, that is a vector weakly dominates another vector according to the leximin if the minimal component of the latter vector is larger than the minimal component of the former vector, or they are equal and the second minimal component of the latter vector is larger than the second minimal component of the former vector, etc., or they are equal.

being a utility representation of R_i and the leximin aggregator being strictly increasing in all its arguments. The fact that R^U satisfies *poverty-reduction transfer 2* follows property 1.1 satisfied by the utility representation. Indeed, as long as $I(z'_j, R_j) > PL > z'_k$, we have $u_j(z'_j) > 1 > u_k(z'_k)$, so that the leximin aggregator gives priority to agent k , with the consequence that z'_N is strictly preferred to z_N .

Statement 4 is also proven by way of an example. Among the many SOF's satisfying the three axioms listed in the statement, we focus on a specific family of SOF's, which will turn out to play a crucial role in the next sections. This family is parameterized by some individual reference preferences $\tilde{R} \in \mathcal{R}$. These reference preferences need to be chosen among the preferences that contain the poverty line as an indifference curve: for all $z = (\ell, c), z' = (\ell', c') \in PL$,

$$z \tilde{I} z'.$$

The SOF's we are interested in work like this: similarly as in the above example, they begin by computing a well-being level for each agent at their assigned bundle, then they apply the leximin criterion \geq_{lex} to vectors of well-being levels. Here is how the well-being levels are computed. That is where the reference \tilde{R} comes into play. The well-being level of an agent (w_i, R_i) consuming bundle z_i is computed by looking at the bundle that \tilde{R} prefers in the implicit budget of (w_i, R_i) at z_i . Formally, Let $\tilde{u} : X \rightarrow \mathbb{R}$ be a utility representation of \tilde{R} . Then, for $z_N, z'_N \in X^N$,

$$z_N R^{\tilde{R}lex} z'_N \Leftrightarrow$$

$$(b(z_i, w_i, R_i))_{i \in N} \geq_{lex} (b(z'_i, w_i, R_i))_{i \in N},$$

where for all $z \in X$, $w \in [\underline{w}, \infty)$, and $R \in \mathcal{R}$,

$$b(z, w, R) = \tilde{u}(m(\tilde{R}, IB(z, w, R))).$$

The fact that $R^{\tilde{R}lex}$ satisfies *strong Pareto* follows again from $b(\cdot, w, R)$ being a utility repre-

sentation of R_i and the leximin aggregator being strictly increasing in all its arguments. The fact that $R^{\tilde{R}lex}$ satisfies *equal-wage transfer* comes from the fact that the well-being index $b(\cdot, w, R)$ is computed using implicit budgets. As a result, as soon as $z_j \in m(R_j, B(z_j, w_j))$ and $z_k \in m(R_k, B(z_k, w_k))$, we have

$$b(z_j, w_j, R_j) \geq b(z_k, w_k, R_k) \Leftrightarrow B(z_j, w_j) \supset B(z_k, w_k),$$

and the leximin aggregator will give priority to the agent with the smallest budget.

The fact that $R^{\tilde{R}lex}$ satisfies *poverty-reduction transfer 3* is related to the specific property satisfied by \tilde{R} . Indeed, let us assume that $I(z_j, R_j) > PL > I(z_k, R_k)$. It is sufficient to prove that $b(z_j, w_j, R_j) > b(z_k, w_k, R_k)$ because the leximin aggregator gives priority to the lowest well-being level. First, $I(z_j, R_j) > PL$ implies that $IB(z_j, w_j, R_j)$ contains a part of PL . As a result, maximizing \tilde{R} over $IB(z_j, w_j, R_j)$ must yield a utility level larger than at PL (remember PL is one of the indifference curves of \tilde{R}). On the contrary, $PL > I(z_k, R_k)$ implies that $IB(z_k, w_k, R_k)$ is everywhere below PL , so that maximizing \tilde{R} over $IB(z_k, w_k, R_k)$ must yield a utility level lower than at PL , the desired result.

Our examples used to prove statements 2 and 4 are of the leximin type, that is, they exhibit an infinite aversion towards inequality in their respective well-being index. It is not impossible to define other SOF's, but the result we present in the next section will prove that the leximin SOF are the only reasonable ones.

1.5 A Local Maximin Result

The previous section has proven that it is possible to combine the objectives of Pareto efficiency, responsibility for one's preferences and poverty reduction, provided the definition of being poor is sufficiently weakened. We proved this compatibility, at the end of the section, by providing an example of a family of SOF's satisfying the axioms. In this section, we prove that all the SOF's that are to satisfy our axioms will look like these examples, at least as

long as social preference over allocations around the poverty line is concerned. This is why we call this result a *local* maximin result. An important feature of these SOF's is that they are of a *maximin* type, in the sense that they need to give absolute priority to these agents who are identified as the poorest according to the well-being index $\tilde{u}(m(\tilde{R}, IB(z, w, R)))$ satisfying the axioms.

This property is obtained when one adds the following separability axiom, which gives some robustness to the social preference. Separability is reminiscent to the idea that indifferent agents should not influence social preference. More precisely, assume social preference is set between two allocations, z_N and z'_N , whereas each agent in a subset M of the agents is allocated exactly the same bundle in both allocations. The requirement is that agents in M should not matter in the social preference, which is captured by the requirement that social preference should remain unaffected if the preferences of agents in M change and if their bundles change in such a way that they are still exactly the same in the two resulting allocations.

Axiom 7 *Separability*

For all $E = (s_N, R_N), E' = (s'_N, R'_N) \in \mathcal{E}$, all z_N, z'_N, z''_N and $z'''_N \in X^N$, if $s_i = s'_i$ for all $i \in N$ and there exists $M \subset N$ such that

$$\forall i \in M : z_i = z'_i \quad \text{and} \quad z''_i = z'''_i, \quad (1.2)$$

$$\forall i \in N \setminus M : R_i = R'_i, \quad (1.3)$$

$$\forall i \in N \setminus M : z_i = z''_i \quad \text{and} \quad z'_i = z'''_i, \quad (1.4)$$

then

$$z_N R(E) z'_N \Leftrightarrow z''_N R(E') z'''_N.$$

The result is best explained by comparison with the axiom of *poverty-reduction transfer*

3. If the axiom is satisfied then a transfer from a non-poor agent to a poor agent is a social improvement, provided an agent qualifies as poor, respectively non-poor, when his entire indifference curve lies below, respectively above, the poverty line. When this axiom is combined with the other axioms, the possibility of poverty reducing reallocation is strengthened in two ways. First, it is no longer needed that the non-poor agent transfers some resources to the poor one. It is sufficient that the non-poor agent loses some amount of resources, however large this amount, and the poor agent receives some resources, however small this gain. This is the maximin aspect of the resulting SOF.

Second, the definition of poor and non-poor changes, to encompass many more situations. It is now sufficient to look at the implicit budget of the agents. As soon as the implicit budget of an agent is below the poverty line, this agent qualifies as poor, whether or not this agent's entire indifference curve lies below it. As soon as the implicit budget of an agent is not below the poverty line, this agent qualifies as non-poor, whether or not this agent's entire indifference curve lies above the poverty line. Observe that with this new definition, agents are either poor or non-poor, depending on their consumption. This considerable strengthening of the definition of poverty yields a complete ordering of the allocations, but a local one, because only allocations in which the same set of agents are below and above the poverty line are ordered on this basis.

Slightly abusing on notation, we write $IB(w, PL)$ to denote the implicit budget of slope w that is tangent from below to the poverty line.

Proposition 2

If a SOF satisfies *Pareto indifference*, *equal-wage transfer*, *poverty-reduction transfer 3* and *separability*, then for all $E = (w_N, R_N) \in \mathcal{E}$ such that $|N| \geq 4$, all $z_N, z'_N \in X^N$, if there exist $j, k \in N$ such that

- $IB(z_j, w_j, R_j) \supset IB(z'_j, w_j, R_j) \supset IB(w_j, PL)$,
- $IB(w_k, PL) \supset IB(z'_k, w_k, R_k) \supset IB(z_k, w_k, R_k)$,

and for all $i \neq j, k$: $z_i = z'_i$, then $z'_N P(E) z_N$.

The proof of Proposition 2 is left for the appendix. The fact that combining Pareto efficiency, fairness and separability conditions leads to SOF's of the maximin type is typical in the literature, and the mechanics behind these results is now well understood (see, especially, Fleurbaey and Maniquet, 2011, chapter 3). What is specific of the above proposition is that *poverty-reduction transfer 3* is a very weak axiom that only justifies transfers of very specific types, and yet the combination with the other axioms leads to an extreme form of inequality aversion, although this aversion only plays around the poverty line.

A simple corollary of Proposition 2 is that poverty is eliminated as soon as the implicit budget of no agent is below the poverty line, with the consequence that any allocation in which poverty is eliminated is socially preferable to any allocation in which at least one agent's implicit budget is still below the poverty line.

Corollary 1 If a SOF satisfies *strong Pareto, equal-wage transfer, poverty-reduction transfer 3* and *separability*, then for all $E = (w_N, R_N) \in \mathcal{E}$ such that $|N| \geq 4$, all $z_N, z'_N \in X^N$, if there exist $j \in N$ such that

$$IB(w_j, PL) \supset IB(z_j, w_j, R_j)$$

whereas for all $i \in N$

$$IB(z'_i, w_i, R_i) \supset IB(w_i, PL),$$

then $z'_N P(E) z_N$.

The conclusion of this section is the following. If we were to reduce poverty based on *poverty-reduction transfer 3* alone, we would be limited at increasing the well-being of agents whose indifference curve lies below the poverty line. Now, reducing poverty and satisfying all

the relevant axioms means that we need to increase the well-being of all agents whose implicit budget lies below the poverty line. To say it differently, poverty is eliminated according to the social preferences axiomatized in Proposition 2 as soon as all agents' implicit budgets intersect with the poverty line or lie above it. Of course, this does not guarantee that all agents' *bundles* are above the poverty line, as would be required if we had imposed *poverty-reduction transfer 1* or *2*, but these axioms are not compatible with our main requirements.

1.6 A General Second-Best Result

In the remaining of the paper, we study the properties of the tax schemes that maximize the SOF's we have justified in the previous sections. We adopt the classical Mirrlees setting. The planner does not observe the characteristics of the agents. She only observes their income. She has to design a redistributive tax scheme $\tau : \mathbb{R}_+ \rightarrow \mathbb{R}$. As a result, an agent with wage $w \in [\underline{w}, \infty)$ and a labor time of ℓ earns income $y = w\ell$, pays income tax $\tau(y)$ and consumes $c = y - \tau(y)$.

Studying tax schemes forces us to study the income/consumption space instead of the labor time/consumption space. That requires some rewriting. As a general rule, we underline variables when they are converted from the former to the latter space. Let $E = (w_N, R_N)$ be an economy. For each $i \in N$, we redefine preferences R_i as \underline{R}_i as follows:

$$(y, c) \underline{R}_i (y', c') \Leftrightarrow \left(\frac{y}{w_i}, c\right) R_i \left(\frac{y'}{w_i}, c'\right).$$

We now refer to (second-best) budgets as $\underline{B}(\tau)$, defined as

$$\underline{B}(\tau) = \{z = (y, c) \in \underline{X} \mid c \leq y - \tau(y)\}.$$

All implicit budgets in the (y, c) space have slopes equal to 1. We now refer to them as

follows:

$$\underline{IB}(z_i, R_i) = \{z = (y, c) \in X \mid \exists t \in \mathbb{R} : c - y \leq t \text{ and } z_i \underline{I}_i \underline{m}(R_i, \underline{IB}(z_i, R_i))\}.$$

Let τ be a tax scheme. Let \underline{z}_N be the allocation generated by the tax scheme, that is for all $i \in N$,

$$z_i \in \underline{m}(R_i, \underline{B}(\tau)).$$

We say that τ is feasible if it generates $\underline{z}_N = (y_i, c_i)_{i \in N}$ such that

$$\sum_{i \in N} c_i \leq \sum_{i \in N} y_i.$$

We say that a feasible tax scheme τ is the optimal tax scheme in economy $E = (w_N, R_N)$ if τ generates allocation \underline{z}_N corresponding to allocation z_N in the labor time/consumption space, and for all other feasible tax scheme τ' , generating allocation \underline{z}'_N corresponding to allocation z'_N in the labor time/consumption space we have

$$z_N R^{\tilde{R}lex} z'_N.$$

In this section, we derive a property of the optimal tax scheme under general assumptions about the characteristics of the agents in the economy. We have up to now assumed that there is a finite type of agents. As a result, we face the risk of being only interested in a finite number of points in the income spectrum. Of course, we think of economies with a finite but large number of taxpayers. As a result, we make the following assumption to guarantee that no interval of incomes is irrelevant for the tax scheme.

Assumption 1: For all $y \in \mathbb{R}_+$, there exists $i \in N$ such that $(y, y - \tau(y)) \underline{I}_i z_i$.

Let us note that Assumption 1 is equivalent to assuming that the function $y - \tau(y)$ follows the lower envelope of the union of the upper contour sets at all bundles z_i .

The second assumption is about the existence of an agent in the economy, we will call her 1, having a low wage but a high willingness to work. To define this agent formally, we introduce the partial ordering \geq over preferences. For $R, R' \in \mathcal{R}$, we write $R \geq R'$ to denote that whatever the set of bundles among which to choose, R chooses a larger labor time than R' .

Assumption 2: There exists agent $1 \in N$ such that for all $i \in N$, $w_1 \leq w_i$ and $R_1 \geq R_i$.

Assumption 3 is more substantial. It requires that in the interval of incomes that are likely to be chosen by low wage agents, that is agents with wage equal to w_1 , as soon as we find one agent earning income y , there is also one agent with wage equal to w_1 earning income y .

Assumption 3: Let y_1 be the earning level of agent 1. For all $y \leq y_1$, if there exists $i \in N$ such that $z_i = (y, y - \tau(y))$, then there exists $j \in N$ such that $w_j = w_1$ and $z_j = (y, y - \tau(y))$.

Assumption 3 is less restrictive than it looks like. Let us assume it does not hold. Then, it means that there exists an interval of incomes around y for which the planner knows for sure that only high wage agents earn incomes in that interval. As a result, in spite of y being a low earning level in the sense that $y < y_1$, the planner is likely to impose a high tax rate simply because of the certainty that it cannot harm low skill agents. Justifying such high tax rate on low incomes sounds extremely implausible. Assumption 3 excludes that possibility.

We are now equipped to prove the main result of this section. It says that agent 1, identified in Assumption 2, is necessarily the worst-off agent according to the well-being index associated to $R^{\tilde{R}lex}$ at the allocation generated by the optimal tax scheme.

Proposition 3 Let Assumptions 1, 2 and 3 hold. If tax scheme τ is optimal in economy $E = (w_N, R_N)$ for a Social Ordering Function $R^{\tilde{R}lex}$, then allocation \underline{z}_N generated by τ is such that the corresponding allocation z_N in the labor time/consumption space is such that agent 1 has the lowest well-being level:

$$b(z_1, w_1, R_1) \leq b(z_i, w_i, R_i), \forall i \in N.$$

The proof of Proposition 3 is left for the appendix. An immediate corollary of that proposition, which proof is essentially contained in the proof of the proposition, is that the optimal tax scheme τ satisfies the following property, the first part of which is illustrated in Fig. 1.5. Again, we have included the proof of this corollary in the appendix.

Corollary 2 Let Assumptions 1, 2 and 3 hold. A) If tax scheme τ is optimal in economy $E = (w_N, R_N)$ for a Social Ordering Function $R^{\tilde{R}lex}$, then allocation \underline{z}_N generated by τ with $\underline{z}_1 = (y_1, c_1)$ is such that for all income $y \leq y_1$:

$$\tau(y_1) \leq \tau(y) \leq t,$$

where $t \leq 0$ is defined by

$$\underline{IB}(\underline{z}_1, \underline{R}_1) = \{(y, c) \in \underline{X} \mid c \leq y - t\},$$

that is, agent 1 receives the largest transfer and marginal tax rates are on average non-positive below y_1 . B) Moreover, if there exists $j \in N$ such that $w_j = w_1$, $\underline{IB}(\underline{z}_1, \underline{R}_1) = \underline{IB}(\underline{z}_j, \underline{R}_j)$ and $\underline{R}_j < \underline{R}_1$, then

$$\tau(y_j) = t,$$

and for all $y \leq y_j$,

$$\tau(y) = t.$$

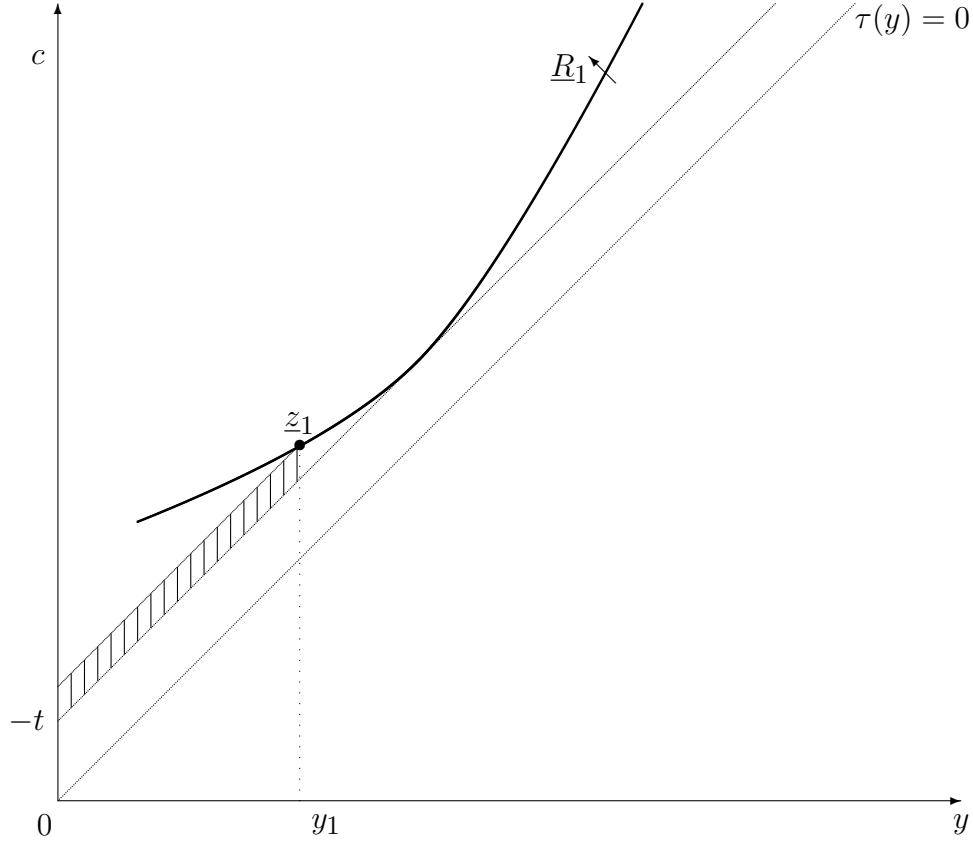


Figure 1.5: Illustration of Corollary 1

Note: The graph of $y - \tau(y)$ must lie in the shaded area for all $y \leq y_1$.

1.7 Additively Separable Preferences and One-Dimensional Heterogeneity: The Optimal Tax Formula for Low Incomes

In this section, we assume that individuals have additively separable utility functions:

$$u_i(c, \ell) = v(c) - h\left(\frac{\ell}{\theta_i}\right)$$

with $v'(\cdot) > 0$, $v''(\cdot) \leq 0$ and $h'(\cdot) > 0$, $h''(\cdot) \geq 0$. Individuals differ in their taste for leisure only through parameter $\theta_i \in [\underline{\theta}, \bar{\theta}]$. This assumption, also used by Lockwood and Weinzierl (2015), reduces the two dimensions of heterogeneity to a single index in the in-

come consumption space: $n_i = w_i\theta_i$.⁸ All individuals with the same parameter n_i will be behaviorally equivalent (that is they all have the same \underline{R}_i) even if they differ in their skill:

$$\underline{u}_i(c, y) = v(c) - h\left(\frac{y}{w_i\theta_i}\right) = v(c) - h\left(\frac{y}{n_i}\right)$$

We assume that $n \sim G(n)$ in the population and the associated density is denoted by $g(n)$.

The constraints of the social preferences maximization are well-known and we review them quickly. We begin with the incentive compatibility constraints. The planner chooses income-consumption bundles $(y(n), c(n))$ for each “behavioral” type n . The income and consumption profiles will define an implicit marginal tax rate:

$$T'(y(n)) = 1 - \frac{h'\left(\frac{y}{n}\right)\left(\frac{1}{n}\right)}{v'(c)} \quad (1.5)$$

We write $u(n)$ the utility of agents of type n at bundles $(y(n), c(n))$. Incentive compatibility constraints require:

$$u(n) = \max_{n' \in [\underline{n}, \infty)} v(c(n')) - h\left(\frac{y(n')}{n}\right) = v(c(n)) - h\left(\frac{y(n)}{n}\right) \quad (1.6)$$

The first-order conditions of the latter maximization problem being satisfied at $n' = n$ implies:

$$u'(n) = h'\left(\frac{y(n)}{n}\right)\left(\frac{y(n)}{n^2}\right) \quad \forall n \in [\underline{n}, \infty) \quad (C1)$$

It also follows from Mirrlees (1976) that the second-order condition for incentive-compatibility is

$$y''(n) \geq 0 \quad \forall n \in [\underline{n}, \infty) \quad (C2)$$

In addition to incentive compatibility, the planner maximizes the SOF under the con-

8. More generally, collapsing multiple dimensions of heterogeneity to one parameter in the context of optimal taxation has also been used by Brett and Weymark (2003) and Choné and Laroque (2010).

straint that at least some revenue R be raised to finance public goods, which, using Eq. 1.6, can be written

$$\int_{\underline{n}}^{\infty} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n) \geq R \quad (\text{C3})$$

We now turn to the definition of the social objective. Individuals with the same n may not all have the same well-being index $b(z_i, w_i, R_i)$, because it depends on their underlying skill w_i and preference parameter θ_i . For any social preferences $R^{\tilde{R}lex}$, however, we can restrict our attention to the minimum well-being among agents of type n at utility level $u(n)$, which we denote $b(u(n))$.

Among individuals of type n , individuals with the lowest-skill are always the worst-off according to b . Indeed, let us consider two agents, j and k with $n_j = n_k$, so that $\underline{R}_j = \underline{R}_k$, but $w_j < w_k$. They choose $\underline{z}_j = (y_j, c_j) = \underline{z}_k = (y_k, c_k)$, with the resulting property that $\underline{IB}(\underline{z}_j, \underline{R}_j) = \underline{IB}(\underline{z}_k, \underline{R}_k)$. Therefore, shifting our attention from the (y, c) -space to the (ℓ, c) -space, we note that $w_j < w_k$ implies $IB(z_j, w_j, R_j) \subset IB(z_k, w_k, R_k)$, where $z_j = (y_j/\theta_j, c_j)$ and $z_k = (y_k/\theta_k, c_k)$. Consequently, $b(z_j, w_j, R_j) < b(z_k, w_k, R_k)$.

The objective of the planner is then

$$\max_{\{u(n), y(n)\}} \left\{ \min_{n \in [\underline{n}, \infty)} b(u(n)) \right\} \text{ subject to (C1), (C2) and (C3).}$$

By Proposition 3 we know that

$$\min_{n \in [\underline{n}, \infty)} b(u(n)) = b(u(n_1))$$

where $n_1 = \underline{w}\bar{\theta}$. Let us denote the solution to the latter maximization problem by $\{u^*(n), y^*(n)\}$ and the corresponding value of the objective by b^* . The problem can equally well be stated as one of budget surplus maximization under the constraint that $b(u(n)) \geq b^*$ for all n . As $b(u(n))$ is a utility representation for lowest-skilled agents within type n , it is strictly increasing. As a result, the constraint can be rewritten directly in terms of a lower-bound

on utilities:

$$u(n_1) = \underline{u}_1 \tag{C4}$$

$$u(n) \geq \underline{u}(n) \tag{C5}$$

where $\underline{u}_1 = u^*(n_1)$ and $\underline{u}(n)$ is the utility assignment as a function of n such that $b(u(n)) = b^*$ for all n .

The social planner's problem becomes:

$$V \equiv \max_{\{u(n), y(n)\}} \int_{\underline{n}}^{\infty} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n)$$

subject to (C1),(C2),(C4) and (C5)

Solving this problem allows us to derive the following proposition, characterizing the optimal tax scheme on incomes earned by agents of type $n \leq n_1$.

Proposition 4 Under the assumptions of this section, if a planner maximizes some $R^{\tilde{R}lex}$, there exist $n_u \in [\underline{n}, n_1]$ and $n_b \in [n_u, n_1]$ such that

1. $T'(y(n)) = 0 \forall n \in [\underline{n}, n_u]$
2. $T'(y(n)) \leq 0 \forall n \in [n_u, n_b]$
3. $y^*(n) = y^*(n_1) \forall n \in [n_b, n_1]$
4. $\frac{T'(y(n))}{1-T'(y(n))} = - \left[\frac{1+\epsilon^u}{\epsilon^c} \right] \frac{v'(c(n))}{ng(n)} \left[\int_{\underline{n}}^n \frac{g(t)}{v'(c(t))} dt - \int_{\underline{n}}^{n_u} \frac{g(t)}{v'(c(t))} dt \right] \forall n \in [n_u, n_b]$

where ϵ^c and ϵ^u are, respectively, the compensated and uncompensated elasticity of labor earnings.

The intuition of this result goes as follows. The reasoning is illustrated in Fig. 1.6. Because the optimal path for $u(n)$ is anchored in the interior of $[\underline{n}, \infty)$, we can divide

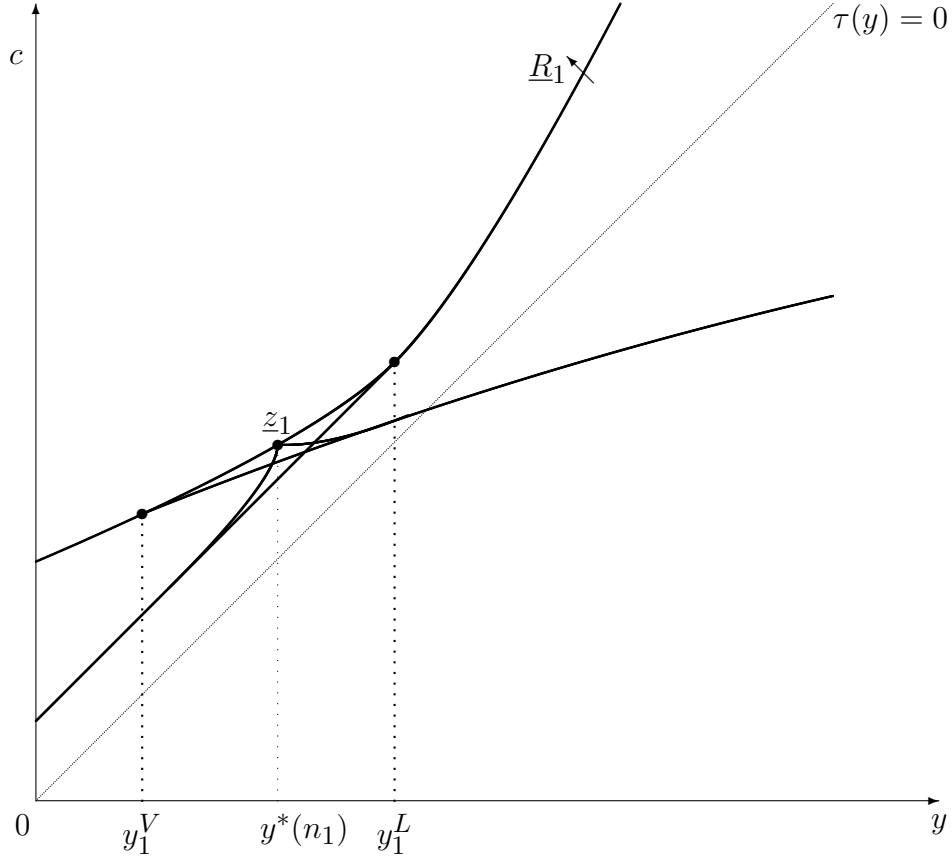


Figure 1.6: Illustration of Proposition 4.

this original constrained problem of maximizing revenue over $[\underline{n}, \infty)$ into two independent subproblems: maximizing revenue over $[\underline{n}, n_1]$ and over $[n_1, \infty)$.

We define:

$$V^L(y_1) \equiv \max_{\{u(n), y(n)\}} \int_{\underline{n}}^{n_1} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n)$$

subject to (C1),(C2),(C4), (C5) and

$$y(n) \leq y_1 \quad \forall n$$

and

$$V^U(y_1) \equiv \max_{\{u(n), y(n)\}} \int_{n_1}^{\infty} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n)$$

subject to (C1),(C2),(C4), (C5) and

$$y(n) \geq y_1 \quad \forall n$$

Denoting by $y^*(n_1)$ the income of agent 1 at the solution, the optimality principle implies

$$V = V^L(y^*(n_1)) + V^U(y^*(n_1))$$

The requirement that $y(n) \leq y^*(n_1) \quad \forall n \in [\underline{n}, n_1]$ in the lower subproblem and $y(n) \geq y^*(n_1) \quad \forall n \in [n_1, \infty)$ in the upper subproblem follows from the constraint that $y(n)$ be increasing in all n (and therefore in n_1).

Let us assume as a preliminary step, that we solve V^L and V^U without taking care of the last constraint on $y(n)$. Then, the solution to V^L would be that $b(u(n)) = b^*$ and $T'(y(n)) = 0$ for all $n \leq n_1$, and $y(n_1) = y_1^L$. Indeed, a zero marginal tax rate below y_1^L guarantees that $\underline{IB}(z_j, R_j) = \underline{IB}(z_k, R_k)$ for all j, k such that $w_j = w_k = \underline{w}$, so that $b(z_j, w_j, R_j) = b(z_k, w_k, R_k) = b(z_1, w_1, R_1)$. Moreover, all agents getting the same $T(y(n))$ is clearly the cheapest way of providing them with well-being level b^* .

However, this tax cannot be optimal if one adds the constraint that $y(n) \leq y^*(n_1)$. Indeed, if we look now at the V^U program without the constraint on $y(n)$, the solution would be that from some y_1^U the optimal T would guarantee that $b(u(n)) = b^*$ for all n in some interval $[n_1, n^*]$.⁹ As illustrated in the figure, as soon as $y_1^U \neq y_1^L$, that cannot be the global solution.

The optimal $y^*(n_1)$ that links the solution to both subproblems must be such that $y_1^U \leq y^*(n_1) \leq y_1^L$ and the optimal T is like in the figure: all agents of type $n \in [\underline{n}, n_u]$ face a zero marginal tax rate, and, among them, all agents with the minimal wage have exactly the same well-being as agent 1. All agents of type $n \in [n_u, n_b]$ face a negative marginal tax

9. The upper bound of this interval would be the limit above which the well-being constraint is no longer binding and the optimal tax is derived from efficiency considerations only.

rate, and they all have a strictly larger well-being level than agent 1. All agents of type $n \in [n_b, n_1]$ earn the same income as agent 1 but all have a strictly larger well-being.

To complete this section, we need to add that the result that marginal tax rates should be zero or negative below $y^*(n_1)$ is more related to the fact that social preferences satisfy equal-wage transfer than poverty reduction. Indeed, it is the combination of Pareto and equal-wage transfer that forces us to measure well-being by using implicit budgets, with the result that well-being is equalized among minimal skill agents if they all maximize their utility over the same real (first-best) budget, as it is the case with a zero marginal tax rate. Surprisingly, the fact that social preferences also satisfy poverty reduction is not seen in the shape of the marginal tax rates on low incomes. It is seen in the level of transfers to the low-skill agents, as well as in the shape of the optimal tax above $y^*(n_1)$, as we show in the next section.

1.8 Quasilinear and Iso-Elastic Preferences

In order to characterize marginal tax rates on incomes above $y^*(n_1)$ we impose additional assumptions on the reference preferences and on the preferences of the agents. We assume that individuals have quasilinear and iso-elastic preferences:

$$u_i(c, \ell) = c - \frac{\epsilon}{1 + \epsilon} \left(\frac{\ell}{\theta_i} \right)^{\frac{1+\epsilon}{\epsilon}}$$

We further assume that the poverty line itself, PL , has the shape of an iso-elastic indifference curve.

$$PL = \left\{ z = (c, \ell) \in \underline{X} = \mathbb{R}_+^2 \mid \bar{P}_0 = c - \frac{\epsilon}{1 + \epsilon} (\ell/\tilde{\theta})^{\frac{1+\epsilon}{\epsilon}} \right\}$$

Any $R^{\tilde{R}lex}$ satisfying our axioms have reference preference \tilde{R} that contain PL as one indifference curve. In Section 1.9.2, below, we assume that \tilde{R} leads to the least redistributive tax scheme consistent with PL . In this section, we assume that \tilde{R} are quasi-linear. They can be

represented by the following utility function:

$$\tilde{u}(c, \ell) = c - \frac{\epsilon}{1 + \epsilon} \left(\frac{\ell}{\tilde{\theta}} \right)^{\frac{1+\epsilon}{\epsilon}}$$

Under these assumptions, we are able to explicitly define the well-being index. An individual with utility u_i , wage rate w_i and preferences θ_i has

$$b(u_i, w_i, \theta_i) = u_i + \frac{1}{1 + \epsilon} \left[(w_i \tilde{\theta})^{1+\epsilon} - (w_i \theta_i)^{1+\epsilon} \right].$$

Under these assumptions on preferences, the optimal marginal tax rate formula become:

$$\frac{T'(y(n))}{1 - T'(y(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{\int_{\underline{n}}^n \mu(t) dt - G(n)}{ng(n)} \quad \forall n \in [\underline{n}, n_b] \quad (1.7)$$

$$\frac{T'(y(n))}{1 - T'(y(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{1 - G(n) - \int_n^\infty \mu(t) dt}{ng(n)} \quad \forall n \in [n_{\bar{b}}, \infty) \quad (1.8)$$

We can solve these equations for the μ 's, and derive the following proposition, which completely characterizes the optimal tax scheme under the assumption that the distribution of types has an upper tail Pareto index α .

Proposition 5 If $G(n)$ has a Pareto index α in the upper-tail and if the lower-bound on utilities binds on one interval above y_1^* , then for $n_u \leq n_b \leq n_{\bar{b}} \leq n_l \leq n' \in [\underline{n}, \infty)$, $\tau'(y(n))$ is such that

1. $\tau'(y(n)) = 0 \quad \forall n \in [\underline{n}, n_u]$
2. $\frac{\tau'(y(n))}{1 - \tau'(y(n))} = \left(\frac{1 + \epsilon}{\epsilon} \right) \frac{G(n_u) - G(n)}{ng(n)} \quad \forall n \in [n_u, n_b]$
3. $y(n) = y_1^* \quad \forall n \in [n_b, n_{\bar{b}}]$

$$4. \frac{\tau'(y(n))}{1-\tau'(y(n))} = \left(\frac{1+\epsilon}{\epsilon}\right) \frac{[G(n_l)-G(n)] + \frac{1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right) \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}} n_l g(n_l)}{ng(n)} \quad \forall n \in [n_{\bar{b}}, n_l]$$

5. If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \geq 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$: then

$$\tau'(y(n)) = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \quad \forall n \in [n_l, \infty]$$

6. Otherwise,

$$\tau'(y(n)) = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \quad \forall n \in [n_l, n']$$

and

$$\frac{\tau'(y(n))}{1 - \tau'(y(n))} = \frac{1 - G(n)}{ng(n)} \left(\frac{1+\epsilon}{\epsilon}\right) \quad \forall n \in [n', \infty)$$

and the marginal tax rate converges to $\tau'(y(n)) \rightarrow \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ as $n \rightarrow \infty$.

This proposition gives us the formula that will be used in the next sections for the calibrations. Before we move on to the calibrations, we add three remarks. First, it is well known that efficiency alone may lead to marginal tax rates that converge to zero when the distribution of types has a thin tail $\alpha \rightarrow \infty$. In that case, the lower bound on utility is never binding in the tail and the tax rate converges to zero.

Second, let us mention that for $\frac{\tilde{\theta}}{\theta}$ large enough (i.e. a sufficiently flat poverty line), the lower-bound on utilities is never binding and the optimal marginal tax rates are again the same as under a Rawlsian maximin over $[n_{\bar{b}}, \infty)$. For instance, this is always the case when the poverty line is perfectly flat.

Third, we can also see the role of the slope of the poverty line relative to the work-aversion of the most hardworking individuals in the economy. A steeper poverty line (a lower $\tilde{\theta}$) leads to lower marginal tax rates over the interval of incomes larger than $y^*(n_1)$ over which the lower-bound is binding.

1.9 Calibration to the US Economy

The combination of the axioms we analyse leads us to defining a poor agent as one whose implicit budget is entirely below the poverty line. As a consequence, poverty in this sense is eliminated if the implicit budget of the worst-off agent is just tangent to the poverty line. Even in the absence of poverty, then, some agents may still consume bundles that are under the poverty line. This situation is illustrated in Fig. 1.7. We know that agent 1 is the worst-off. In the figure, this agent's implicit budget is tangent to the poverty line. As it is defined in the (ℓ, c) space, it is then calibrated to the (y, c) space to match agent 1's wage, \underline{w} . The well-being is equalized as much as possible below and above agent 1's earning, which means that then tax scheme has the property we derived in the previous section. For instance, agents who do not earn any income have a consumption level of $-\tau(0)$ and their bundle is below the poverty line.

Such an allocation is unlikely to be efficient, especially if the poverty line is low (as it is the case for the US). Some freedom is left about how the possible welfare gains should be distributed. That is the role played by the choice of the reference preferences \tilde{R} , and, more precisely, by the shape of the indifference curves of these preferences above the poverty line. In the previous section, we assumed that \tilde{R} are quasi-linear. That means that the welfare gains above the poverty line are distributed uniformly over the entire population. Our first set of calibrations, in the first subsection, follow this assumption.

We may assume, however, that social preferences are such that no more redistribution is legitimate as soon as poverty is eliminated. That means that redistribution should be calibrated so that the optimal tax scheme does not implement more redistribution than what is necessary to lift all implicit budgets above the poverty line, but, at the same, time, redistribution should still be consistent with the maximization of some social preferences $R^{\tilde{R}lex}$. Our second set of calibrations, in the second subsection, follows this line.

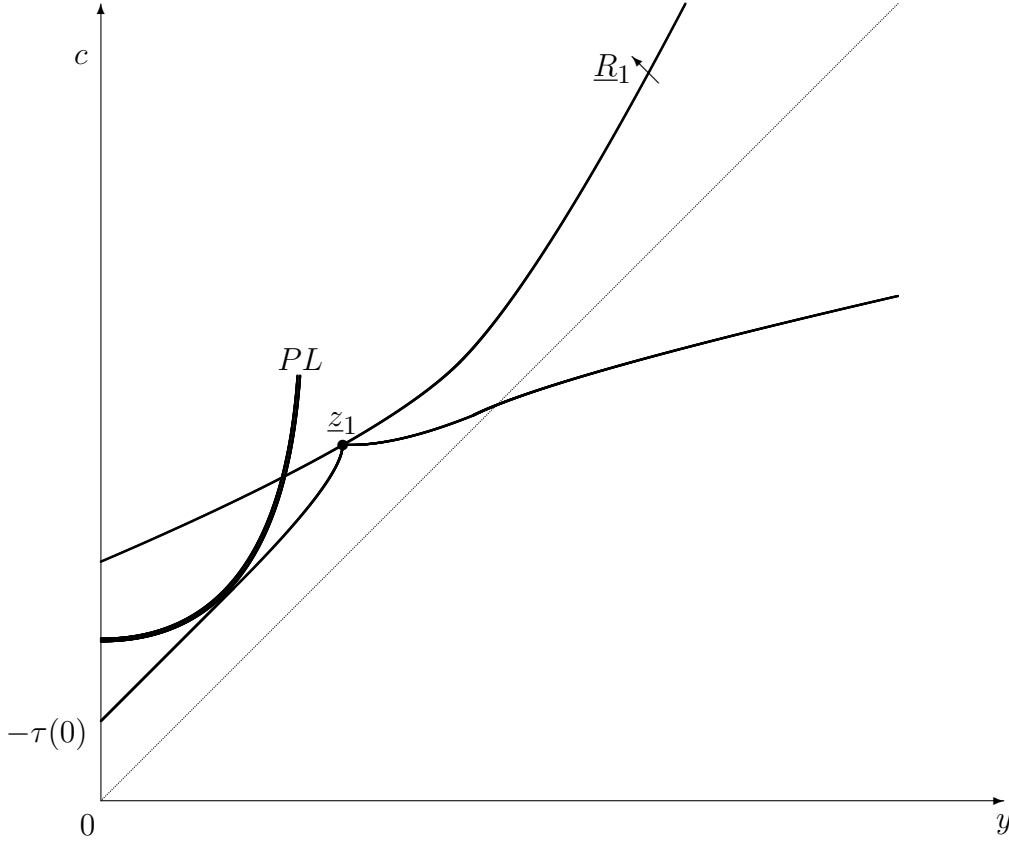


Figure 1.7: How to define a better tax schemes than the one that lifts all agents out of poverty: the role of \tilde{R} .

1.9.1 Quasilinear Reference Preferences

We calibrate the optimal tax formula and compare them to the current US income tax and welfare system. The distribution of types is estimated via Maximum Likelihood using Current Population Survey (CPS) microdata.¹⁰ We estimate a four parameter double Pareto Lognormal distribution.¹¹ For each household in the data, we simulate the marginal federal tax rate they face using NBER’s TAXSIM.^{12 13} Conditional on a calibrated labor supply

10. IPUMS-CPS, University of Minnesota, www.ipums.org

11. The Double Pareto Lognormal distribution was introduced by Reed(2003). Also see Hajargasht and Griffiths (2013) for an empirical application.

12. See Feenberg and Coutts (1993).

13. In addition, we impute an additional implicit marginal tax rate from the phaseout of government assistance. Details on the estimation of the phaseout of government assistance and imputation of implicit marginal tax rates are given in appendix A3.

elasticity parameter, the marginal tax rate and wage income of the households, we can recover individual types n_i by inverting the first-order conditions of the individuals' problem. The distribution is then estimated by maximizing the log-likelihood of this sample of n_i .¹⁴ We pick a labor supply elasticity $\epsilon = 0.33$ which is the preferred estimate in Chetty (2012).

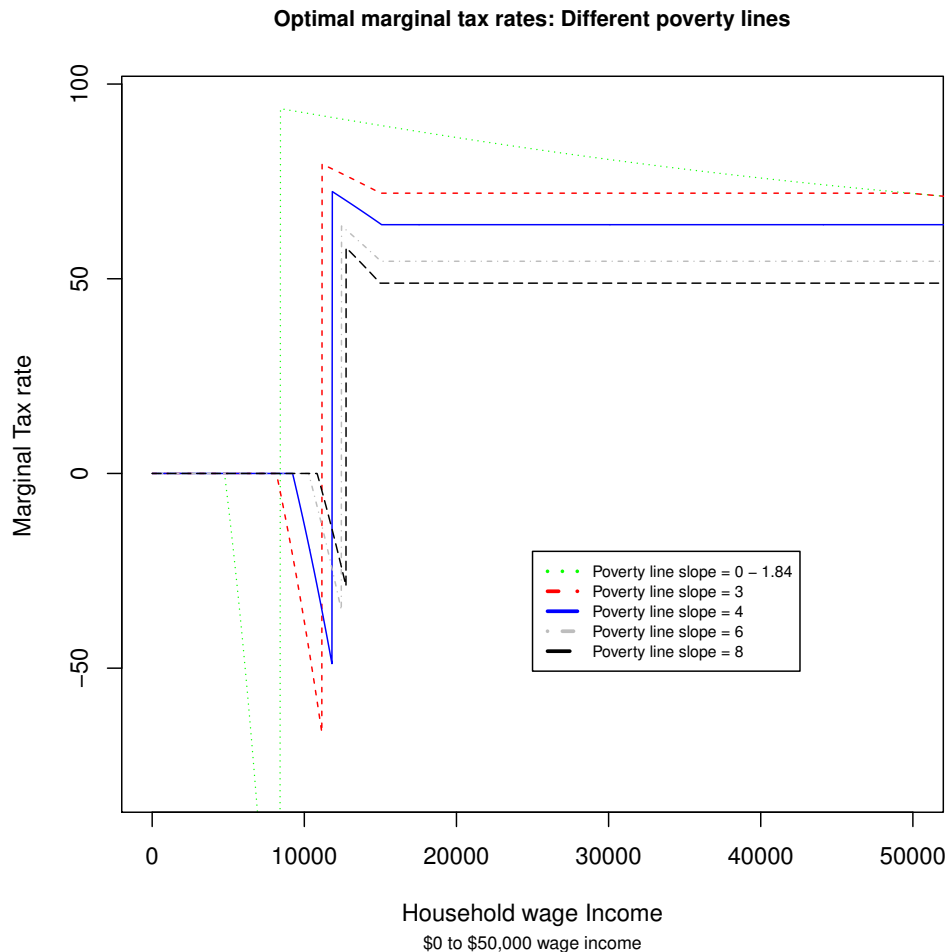


Figure 1.8: Calibration of the optimal marginal tax rates according to different slopes of the poverty line.

We present the optimal marginal tax rates for different calibrations of the poverty line in figure 1.8. We vary the average slope of the poverty line from \$0 to \$8/hour. Bunching

14. It is worth mentioning that we obtain a Pareto tail parameter $\alpha = 3.418$ for the distribution of n_i . Because the marginal rate in the upper tail is approximately constant and given the first-order conditions of individual behavior, this translates into a Pareto tail of about $\alpha_{income} = \frac{\alpha}{1+\epsilon} = 2.57$ well within the range of previous estimates in the literature. Most estimates lie between 1.5 and 3. Our estimate is closer to the upper bound of this range. It was expected since we calibrate the distribution of wage and salary income only (exclude other sources of income such as dividends, capital gains,...).

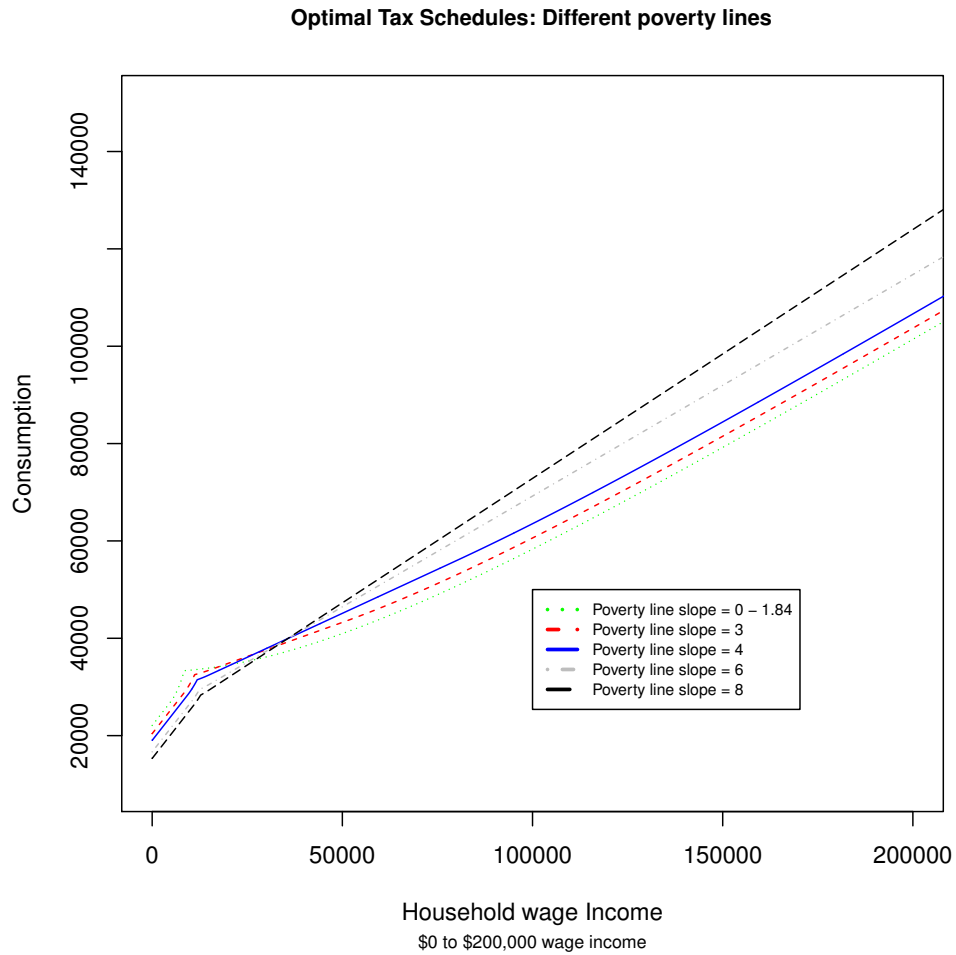


Figure 1.9: Calibration of the optimal consumption schedule according to different slopes of the poverty line.

happens at $y^*(n_1)$, the level of pre-tax income chosen by the most hardworking individuals among low skills. This level of income varies between \$8,431 for a flat poverty line and \$12,758 for an average slope of \$8/hour. The intuition for this result is that tax systems are more and more generous towards the hardworking low-skill individuals as the poverty line becomes flatter, but this goes with an increase in the incentives of higher types to mimic this individual who is then incentivized to earn a lower income.

For very flat poverty lines, marginal tax rates are extremely large and negative just before the bunching point. They are also very large (over 93%) right after. Notice that this is true for all poverty lines with a slope below \$1.84/hour. At these values, the marginal tax rates along the lower-bound on utilities are so prohibitive that optimal rates are driven by efficiency consideration alone after $y^*(n_1)$. Marginal tax rates take less extreme values as the slope of the poverty line increases. For a slope of \$8/hour, marginal tax rates are never higher than 58% and are 48.9% in the tail. This is lower than the rates implied by zero welfare weights in the tail. Most social welfare functions in the literature apply welfare weights of zero on very large incomes in which case $\tau'(y(n)) \rightarrow \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} = 54\%$. Figure A.3 in Appendix A2 shows the marginal tax rates over a larger range of income.

Figure 1.9 shows the corresponding optimal consumption schedules. The subsidy to individuals with \$0 of earnings vary between \$15,318 and \$22,090. This is likely an upper bound for this estimate as it is sensitive to the assumptions on the revenue for public goods that needs to be raised by the income tax.¹⁵ The thought experiment is to consider an optimal income tax according to the social objective that is revenue neutral relative to the current tax and welfare system. Also notice that these calibrations assume that reference preferences are quasi-linear - as a result the level of the poverty line does not matter and

15. We estimate that the average federal tax liability net of welfare receipt in the CPS sample is \$12,152 per household but this estimate does not include disability payments and excludes households with zero income. 14.9% of households in the original sample have an income of \$0. They receive an average of \$4,086 in welfare (this number includes only food stamps, SSI, TANF and disability benefits). Assuming that 85.1% of working households also need to finance welfare payments for the non-working households raises the constraint on government revenue to \$12,867 per working households. The sample also excludes households whose head is above 60 years old.

social preferences lead to redistribution beyond what is necessary to lift everybody out of poverty. We consider and calibrate “least redistributive” social preferences compatible with our axioms in the next subsection.

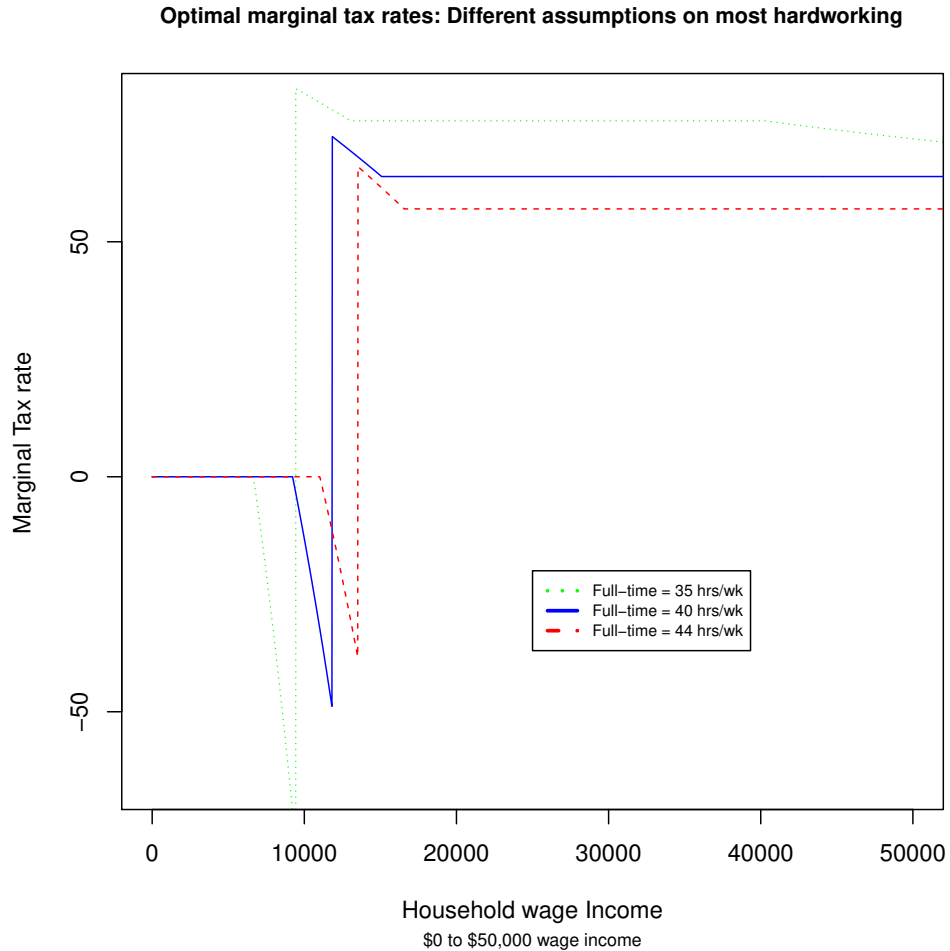


Figure 1.10: Calibration of the optimal marginal tax rates according to different $\bar{\theta}$

Note: $\bar{\theta}$ is the preference parameter for the most hardworking among minimum skill. All calibrations in this figure are made under the assumption that the poverty line has an average slope of \$4/hour.

Figure 1.10 shows the sensitivity of our calibrations to assumptions on preferences of the most hardworking individuals $\bar{\theta}$. This is an important parameter in our expressions for optimal taxes. We calibrate it by assuming that the most hardworking individuals with the minimum wage work 40 hours a week and 48 weeks a year. First, it is clear from the formula that larger values of $\bar{\theta}$ (more hardworking) lead to lower marginal tax rates. Actually, the

ratio $\frac{\bar{\theta}}{\theta}$ enters the expressions. As a result, varying $\bar{\theta}$ has a similar effect as varying the slope of the poverty line. Conceptually, we can think of these comparative statics this way: we hold the distribution of n fixed and vary the extent to which inequality in n (and thus income) are due to differences in preference parameters θ . As we increase $\bar{\theta}$, we increase the range of the preference parameter in the economy and conceptually load more of the income inequality on heterogeneity of preferences as opposed to differences in skills. This yields lower marginal tax rates above $y^*(n_1)$. This result is similar in spirit to Lockwood and Weinzierl (2015), in which more heterogeneity of preferences also lead to less redistribution. They however consider a different measure of heterogeneity of preferences. Because our social ordering function is of the leximin type, an increase in heterogeneity of preferences that leave $\bar{\theta}$ and the distribution of incomes unchanged, also leaves the optimal tax unaffected. This is an important difference with Lockwood and Weinzierl (2015).

1.9.2 *Least Redistributive Social Preferences*

An odd feature of the social preferences defined in section 1.9.1 is that the level of the poverty line does not matter to their definition. This is entirely due to our choice of quasi-linear reference preferences. If reducing poverty is the only redistributive motive of the planner, we may want to consider preferences that do not involve redistribution towards the worst-off if he is above the poverty line. With the same logic according to which a steeper poverty line in figure 1.8 leads to less redistribution, we can see that more labor averse reference preferences imply less redistribution. In this section, we are looking to define reference preferences that become labor averse as quickly as possible above the poverty line (remember that the poverty line itself needs to be an indifference curve of the reference preferences) so that redistribution between agents above the poverty line decreases as quickly as possible as we are further away from poverty.

There are multiple ways of defining such preferences. We keep the parametrization of the poverty line from section 1.8 and decide to have the following: preferences are quasi-

linear below the poverty line but indifference curves rotate around $(c, \ell) = (\bar{P}_0, 0)$ to define preferences above PL . Again, this rotation of preferences may happen in different ways. We choose to let $\tilde{\theta}$ get smaller. The resulting preference relation is well-defined everywhere except in $(c, \ell) = (\bar{P}_0, 0)$. We define these preferences more formally in appendix A4.

Under these assumptions, we can directly use our results from section 1.8 in order to derive the optimal tax. The only difference is that we are now looking for $\tilde{\theta}$ such that $b(u(n_1)) = \bar{P}_0$. Figures 1.11 and 1.12 show the optimal allocation for different values of \bar{P}_0 .¹⁶ Unsurprisingly, a higher poverty line leads to more redistribution and larger marginal tax rates on large incomes. It also involves lower marginal tax rates on incomes below $y^*(n_1)$. The values for the poverty lines used in this calibration are weighted average poverty lines in 2014 in the US for households of size one to five members respectively.¹⁷ The subsidy to non-working poor households is, in each case, a little below the line - meaning that poverty is not completely eradicated.¹⁸ This is a consequence of our weakening of the poverty-reduction axiom. Though individuals with income below the poverty line in those calibrations are poor according to the traditional definition of poverty, their indifference curve does not lie entirely under the poverty line. Under our axiom of poverty reduction, a poverty-reducing transfer is required to be a social improvement only if the indifference curve of the poor individual lies entirely under the line.

1.9.3 Comparison to the Current US Tax and Welfare System

Figure 1.1 in the introduction plots the calibrated optimal tax rate for an average slope of the poverty line of \$6/hour alongside the current marginal tax rate for single filers with 0, 1 and 2 or more children respectively. We show the marginal tax rates for joint filers in Appendix A3. In the US, households with at least one child face negative marginal

16. Figure A.4 in appendix A2 shows the optimal consumption schedule on a larger range of incomes.

17. Source: Census Bureau report by DeNavas-Walt and Proctor (2015).

18. The subsidies to non-working poor households are \$7,747, \$10,388, \$13,117, \$17,052 and \$20,074 for poverty lines of \$12,071, \$15,379, \$18,850, \$24,230 and \$28,695 respectively.

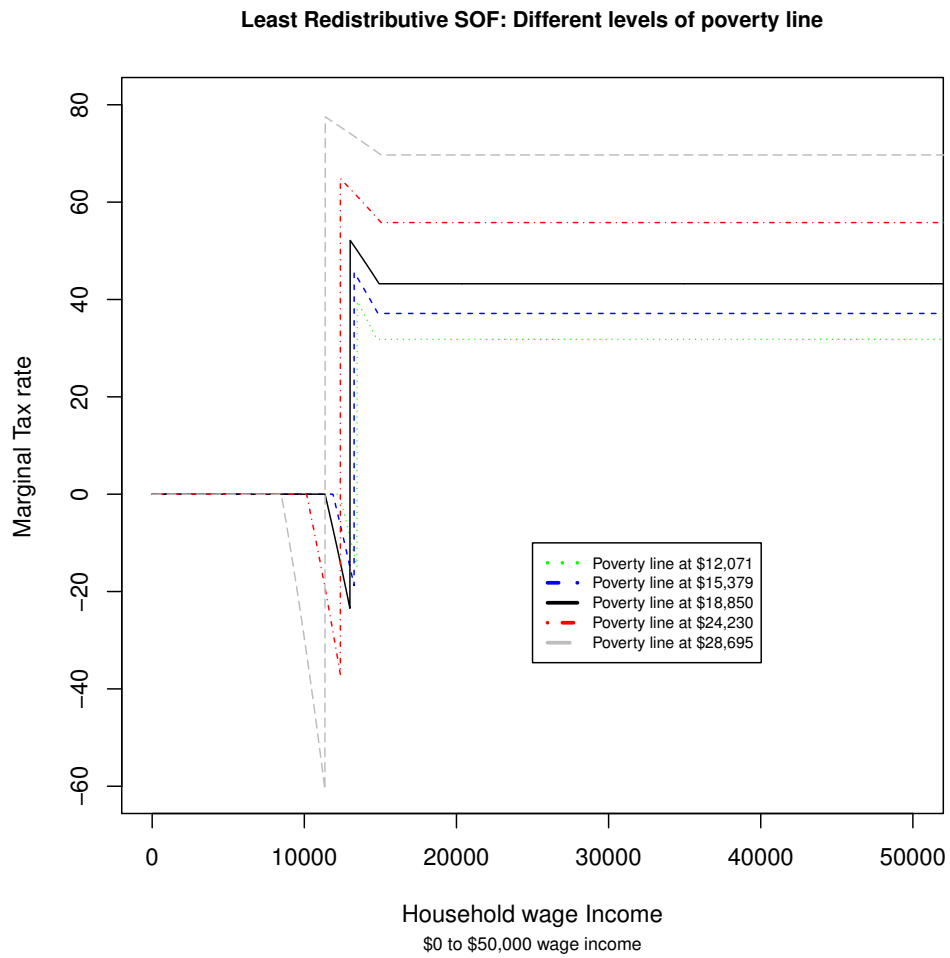


Figure 1.11: Calibration of the optimal marginal tax rates according to different levels of the poverty line.

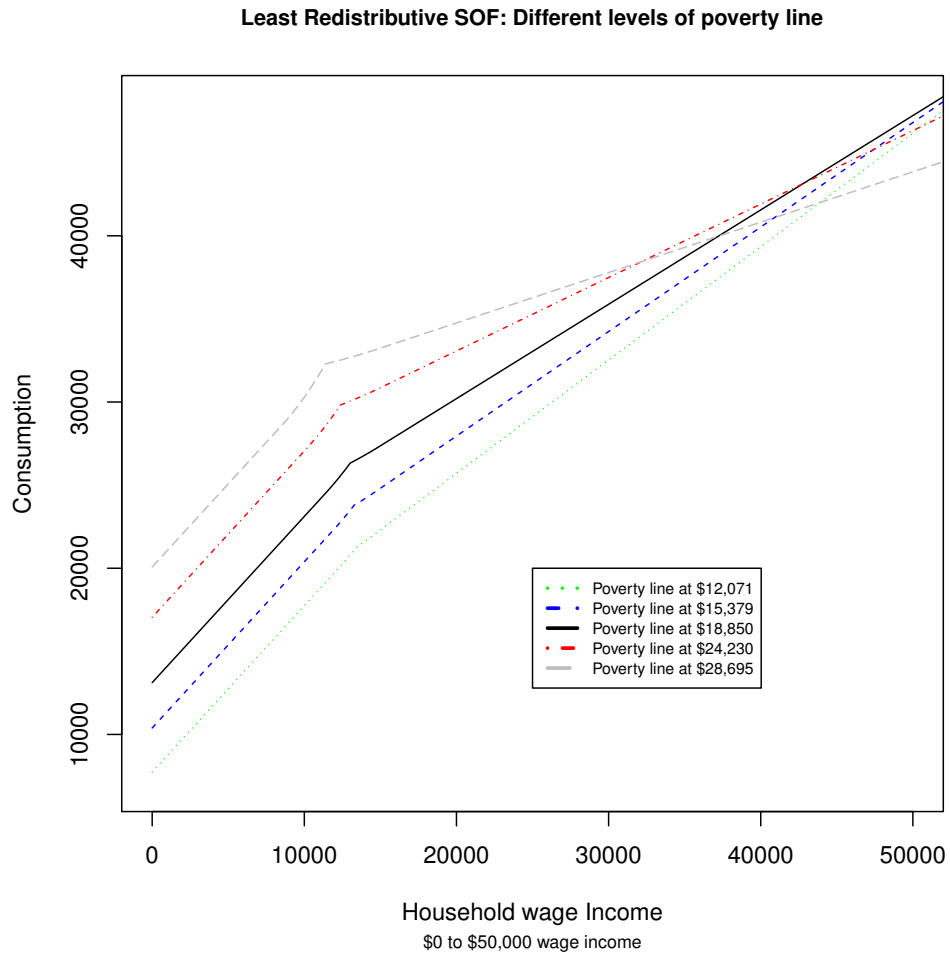


Figure 1.12: Calibration of the optimal consumption schedule according to different levels of the poverty line.

tax rates on income below \$6,000 to \$13,000 depending on the number of children and filing status. This is driven by the Earned Income Tax Credit - though some of the work incentives from the EITC are undone by the implicit positive marginal tax rate generated by the phaseout of welfare programs such as SSI, TANF and SNAP. Marginal rates tend to increase steeply roughly around \$10,000 and reach up to 60% for households with more than 2 children. The increase is mainly driven by the phaseout of the Earned Income Tax Credit and other welfare programs. In appendix A3, we show, for each household type, the break down of the marginal tax rate between the federal income tax and the phaseout of welfare payments. Both contribute to the steep increase in marginal rates. Marginal rates then tend to decline and stabilize around 40% for high incomes. Households with no children do not face negative marginal tax rates mainly because the EITC is much more generous for families with children. As a result, the phaseout of the EITC is not as important and they do not face as high positive marginal tax rates around \$10,000 to \$45,000 as households with children. This is also the case because the welfare system is less generous towards families without children.

Interestingly, our calibrated optimal tax rates resemble the tax schedule faced by households with at least one child: zero and negative rates on low incomes then a steep increase to large positive marginal rates around \$8,000 to \$12,500. The calibrated formulas corresponding to an average slope of the poverty line of \$6 to \$8 per hour look the closest to the current US schedule. Marginal tax rates are similar in magnitude though they are lower (more negative) on low income. They are also larger on middle- to high income earners (above \$50,000) implying more redistribution towards everybody else.¹⁹

19. An important exception is the “least redistributive SOF” considered in 8.2 for low levels of the poverty line. A poverty line at \$15,379 leads to marginal rates of 37% on high incomes for that particular calibration.

1.10 Conclusion

This paper is the first one to start an analysis of optimal taxation at the level of the axioms that social preferences should satisfy and to complete the analysis at the level of the complete characterization of the optimal tax formula, a formula that has been calibrated to the US economy.

The axioms that we started with embed fairness principles. For this reason, this paper follows a current trend of the literature in which the emphasis shifts from the classical utilitarian objective, in which the social optimum is defined in terms of allocation of subjective utility, to the fairness of the resulting allocation of resources, in this case labor time and consumption.

The two main fairness norms that we have studied are responsibility for one's preferences and poverty reduction. The former is related to the requirement that no redistribution should take place among agents having the same ability to earn income, a requirement that seems to be central to the many approaches to fairness in labor income taxation (see Fleurbaey and Maniquet, 2018, for a survey). The latter, poverty reduction, has already received attention in the literature, but in a way that is incompatible with Pareto efficiency, with the consequence, that, contrary to what we obtain here, an increase in the social objective can be concomitant with a decrease in the utility of the poor. The axiom of poverty reduction that we have studied here avoids this paradox, but at the price of weakening the definition of being poor. Rather than defining poverty as consuming a bundle below the poverty line, we have been left to define it as consuming a bundle on an indifference curve entirely below the poverty line.

As a result, a tax scheme may be optimal in our sense and, yet, leave some agents consume bundles below the poverty line. What the optimal tax scheme accomplish, though, is to avoid redistribution among the low-skill agents. As a result, the marginal tax rates are zero or even negative up to the earning level of the low-skill agent choosing the longer labor time. Consequently, the budget of these agents cross the poverty line from below, so that

agents end up being lifted out of poverty, that is consuming bundles above the poverty line, provided their labor time is sufficient.

Another consequence of our axioms is that, in spite of our social preference taking the shape of an extremely egalitarian one, the optimal tax rates on high incomes may differ from the ones that would maximize the tax return (the so-called Rawlsian tax scheme). This is even more true as we choose steeper or lower poverty line. This has to do with the way we define the well-being index which social preference tries to maximin. Indeed, it is typically not true that an agent with a larger wage ends up with a larger well-being level (whereas in the classical framework an agent with a larger wage always ends up with a higher utility). This comes from the responsibility axiom, which forces us to look at implicit budgets rather than consumption bundles.

This brings us to the evaluation of the US tax system. It is surprisingly close to the one that we rationalize with our axioms, with three major differences. First, we cannot rationalize negative marginal rates in the neighborhood of zero earning. The rates should be zero according to our formula. This is not a huge difference, but the current scheme is better explained in this interval by arguments in terms of extensive margin effects (people deciding to move from unemployment to employment), like in Saez (2002), or in terms of incentives to go beyond overestimated costs of finding jobs, such as in Lockwood (2016).

The second difference between our optimal tax scheme and the current one in the US is that it is hard to rationalize a brutal decrease in marginal tax rates after the phasing out of the IETC, followed by a plateau. Our results push towards a smoothing of this part of the tax system, where the constraint that the well-being of agents earning incomes in this bracket should not be lower than the well-being of the hardworking poor is likely to be binding.

The third and main difference is that the US tax system for households without children does not look at all like the one our axioms rationalize. It is therefore clear that responsibility and poverty reduction may be seen as the fairness norms guiding the tax system only for

what concerns households with children.

Regarding rates on high incomes, some parametrization of the social objective are more redistributive than others. For redistributive enough social preferences, marginal rates are entirely determined by efficiency considerations in the tail and only depend on the distribution of wages in the economy. Less redistributive social preferences (a steeper or lower poverty line for instance) lead to lower marginal tax rates in the tail than would be implied by maximizing tax revenue under efficiency considerations alone.

A recent development - and attractive feature - of modern optimal tax literature has been to derive formula that are somewhat robust to model specification and are expressed in terms of sufficient statistics. Though our results do not share this feature, we view it as an unavoidable consequence of the normative stance taken in this paper: we assume that society aims at correcting for certain sources of inequalities, the ones generating poverty, but remain neutral towards others, the ones emerging from different labor time choices. This requires a *model* of what the sources of inequality are and it is therefore not surprising that the corresponding tax formula are *model dependent*. This relates to a long tradition in ethics positing that normative judgments must be context dependent. Our different steps from the axioms to the optimal tax formula have forced us to add assumptions on the way. At the end, we have restricted our attention to an economy in which taxes are only distortive at the intensive margin (changes in the tax system only impact labor time choices marginally) and the elasticity of labor supply is constant. Even if we have not been able to generalize our results, we still note that studying incentives also at the extensive margin often leads to optimal tax schemes exhibiting lower marginal tax rates on very low incomes, whereas we have already obtained such a result.

CHAPTER 2

INTERGENERATIONAL ELASTICITY OF CONSUMPTION

(Joint work with Jung Sakong)

2.1 Introduction

The increase in inequality of earnings, consumption and wealth over the last decades is large and widely documented.¹ This dramatic increase has led to concerns that society is getting more and more stratified with few opportunities for mobility of individuals at the bottom of the distribution.² Another striking feature of the distribution of income in the United States is the persistence of the racial gap. Bayer and Charles (2018) show that in 2014 the median earnings gap between blacks and whites in the US was back to its 1950 level and, maybe more strikingly, that the position of the median black man in the overall income distribution relative to the median white man has lagged behind and remained roughly constant between 1940 and 2014. The racial wealth gap is of an even larger magnitude: the average net worth of a white household is 7 times that of a black household.³

Fears of a stratified society and the persistence of the racial gap, call for an understanding of long-run dynamics of income, wealth and consumption and in particular for an intergenerational perspective. A large literature on intergenerational mobility of income has emerged.⁴ Estimates of the intergenerational elasticity of earnings in the US range between 0.3 and 0.6

1. See for instance Autor (2014), Murphy and Topel (2016) for labor income; Saez and Zucman (2016) for wealth; Piketty, Saez and Zucman (2018) for labor and capital income. See Attanasio, Hurst and Pistaferri (2015) for consumption.

2. Another major concern has been about the top of the income distribution pulling apart. See for instance Piketty and Saez (2003) and its online update to 2015 data.

3. See Sakong (2018), Appendix B. The figure is based on the 2013 Survey of Consumer Finances.

4. For studies on mobility in earnings, see Chetty et al. (2014) as well as related work by the same team of researchers and collaborators, Solon (1999) for a survey, and Mazumder (2015). Regarding mobility of wealth, see for instance Charles and Hurst (2003).

suggesting a reasonable level of mobility. More recently, researchers have also studied race-specific intergenerational dynamics of earnings and their relation to the racial gap.⁵ These studies generally converge on the findings that blacks and whites experience relatively similar within race mobility but blacks have much lower absolute mobility. In particular, black Americans have lower rates of upward mobility and higher rates of downward mobility than whites. Chetty et al. (2018) show that these dynamics - when extrapolated to a steady-state economy - account for the persistence of the current racial gap. Another important finding from the literature on intergenerational mobility is the important heterogeneity across places within the US.

Though the literature on inequality has focused on earnings, wealth and consumption, most studies on intergenerational mobility have focused on earnings and, to a lesser extent, wealth with relatively little attention to consumption.⁶ Addressing Marx and Piketty's fears of a stratified society however requires measuring persistence through the transmission of both traits affecting earnings and physical capital. In Marx's view, *La bourgeoisie* has no labor income but is nevertheless pulling apart from the rest of society. In other words, intergenerational persistence in permanent income is affected not only by the transmission of human capital but also by the ability to pass physical capital onto the next generation as well as the correlation between wealth and returns to wealth. These channels have received a lot of attention in the debate on taxation of bequests. Another mechanism of insurance typically missed by income measures are government assistance programs: both cash - typically under-reported in survey income measures - and in-kind.⁷ For example, Meyer and Sullivan (2017) find lower poverty rates and inequality using consumption based measures. Finally,

5. See Bhattacharya and Mazumder (2011), Mazumder (2014), Mazumder and Davis (2018) and Chetty et al. (2018).

6. There are notable exceptions. Studies on intergenerational mobility of consumption include Mulligan (1997), Aughinbaugh (2000), Waldkirch, Ng and Cox (2004) and Charles et al. (2014).

7. See Meyer, Mok and Sullivan (2015) for evidence on the underreporting of government assistance in surveys.

families may also share risk through pecuniary and non-pecuniary *inter vivos* transfers. Consumption and its intergenerational dependence captures all these channels of transmission of economic well-being and maps more directly into welfare. Comparing these estimates to the much studied persistence in earnings is informative on the ability of families to smooth consumption across generations and insure against shocks. A simple version of the neoclassical model would predict an intergenerational correlation of consumption of 1. The previous literature has typically found numbers well below this benchmark.⁸ Studying the intergenerational elasticity of consumption and its heterogeneity across ethnicities and place therefore has the potential to complement studies on the persistence of earnings and bring insights into the transmission of welfare and long-run savings behavior of “dynasties”.

Another - but closely related - reason for focusing on consumption has to do with the permanent income hypothesis.⁹ Individuals’ consumption decision - and eventually welfare - depends on their permanent income. Directly measuring permanent income would require extensive knowledge of the lifetime income process of the individual - decomposed in transitory and permanent income shocks - along with a measure of credit constraints. Simply averaging household income over years does not quite measure permanent income. The value of average yearly income for the individual is a function of the credit constraints, the variance of income and its pro-cyclicality among other things. Focusing on consumption solves this issue by *letting the household do the averaging*. These considerations may be important in the context of measuring intergenerational mobility because the variance, pro-cyclicality of income and credit constraints are likely to systematically differ across the income distribution - and race - leading to biased estimates of mobility in permanent income.

In this paper, we show that focusing on consumption leads to slightly higher estimates

8. Mulligan’s (1997) preferred estimates are between 0.7 and 0.8. Charles et al. (2014) find a much lower correlation around 0.3.

9. See Friedman (1957)

of intergenerational mobility than the estimates in the previous literature that are based on average yearly income.¹⁰ We also find that the focus on consumption is important for inter-race comparisons: blacks have relatively lower within-race persistence in consumption than whites. A potential explanation for the larger racial gap in persistence of consumption relative to earnings is that limited access to credit and instruments of asset accumulation hinders the ability of blacks to insure future generations. Finally, some of the regional heterogeneity in consumption mobility that we uncover also contrasts with the patterns found for earnings mobility in the previous literature. In particular, while the Northeast appears more mobile than the South on the earnings dimension, it has more persistence in consumption. This again suggests a connection to credit markets which were generally more developed in the Northeast than in the South around 1940.

The lack of datasets with information on consumption have hindered research on the intergenerational elasticity of consumption. To our knowledge, all previous evidence was derived using the Panel Study of Income Dynamics (PSID). Mulligan (1997) uses a weighted sum of housing, spending on utilities and automobiles as a proxy for consumption and find an intergenerational elasticity of around 0.7 – 0.8. More recently, Charles et al. (2014) use the more comprehensive measures of expenditures collected in the PSID since 2005 and find a correlation of 0.28. We complement the previous literature by presenting long-run estimates for the period 1940-2012, by computing the first estimates of heterogeneity by race and geography and our evidence is the first that is not based on the PSID. Though the PSID is a natural starting point for studies of intergenerational persistence of consumption, these studies need to be complemented with evidence from other data sources. The PSID has a relatively small sample size which - among other things - restricts possibilities to study heterogeneity across demographic groups. Blundell, Pistaferri and Saporta-Eksten (2016)

10. Mazumder (2015) shows that averaging over more years of income (up to 15 years) lead to larger estimates of IGE. Focusing on consumption, we find slightly higher persistence.

compare PSID aggregates of nondurables and services expenditures to National Income and Product Accounts (NIPA). The PSID to NIPA ratio varies between 0.64 and 0.73 across years 1998 to 2008 which raises questions about the quality of its consumption data. Finally, the PSID starts in 1968 and only enriches its collection of consumption data in 1999 and then again in 2005. As a result, consumption measures for the previous generation must necessarily be imputed from categories such as housing, automobiles and spending on utilities.

We propose an alternative methodology to track consumption across generations. We measure housing consumption in the 1940 full-population Census and in current data including 94% of all homeowners. We use last-names to define cohorts and link the two datasets. Our method allows to go back further in time and estimate persistence over 1940-2012. The dataset in each period is rich and includes virtually the entire population.¹¹ This allows us to explore heterogeneity across races and geography. On the other hand, our data contains only housing - not total - consumption. In section 5, we assess the quality of housing as a proxy for overall consumption and explicitly derive the assumptions under which persistence in housing is equal to persistence in overall consumption. Another challenge is that last-name level correlations do not necessarily need to equal their individual-level counterparts. This is an occurrence of what is known in statistics and related fields as the *ecological inference problem*.¹² In section 4, we develop an econometric framework based on the theory of “linking independent samples” with common variables and use it to understand the assumptions under which surname- and individual-level estimates coincide.¹³ We devise placebo tests to evaluate these assumptions. The results lead us to adjust estimates using covariates - including at least race - observed in the 1940 data. We also estimate intergenerational correlations by race and region. We show that using last names instead of the missing individual

11. The 1940 Census is a full population Census while the DataQuick sample used for current data includes 94% of all homeowners.

12. On the ecological inference problem, see King (1997, 2004) and Spenkuch (2018) among others.

13. Our econometric framework is based on Ridder and Moffitt (2007).

linkages somewhat limit the extent to which we are able to break down correlations by finer and finer demographic groups. Our estimates by region and race nevertheless reveal patterns that are somewhat different from the existing literature on intergenerational correlation of earnings. Finally, because the Dataquick sample contains only the housing consumption of homeowners, we have to adjust our estimates for selection into home ownership. In section 6, we show how to adjust for selection. In a later version of this paper, we will replace the Dataquick sample with the InfoGroup data that includes both renters and homeowners. Section 8 concludes and present suggestions for future work on the topic.¹⁴

2.2 Data and Variables

Datasets: Our estimates of Intergenerational Mobility in Housing Consumption combine two data sets: the 1940 Census and DataQuick. We use the Consumer Expenditure Survey (CEX) for years 1980-2015 to impute overall consumption based on housing consumption in 1940 and 2012. Finally, we also use the 2010 Census list of frequently occurring surnames. This list contains all last names - 149,436 names - for which there are more than 100 members in the US that year.

2.2.1 1940 Census

Dataset Description: The 1940 Census surveyed a hundred percent of the US population in that year. It is the first Census to include questions about an individual's income and education. In addition, the original manuscripts - containing first and last names of individuals - were released in 2012 as they are no longer protected under the 72 year privacy period. We obtained a digitized version of the 1940 Census collected by Ancestry.com and made available to researchers affiliated with the NBER.

14. Including a few analyses that will be added to later versions of this study.

Sample Definition: We keep all male heads of households who did not live in group quarters in 1940. The rationale for keeping only the heads of household is that housing is a public good within the household. We drop women from the sample because surnames are traditionally inherited from male ancestors. In addition, we get rid of observations with missing housing consumption, race or Census region. We further drop all individuals with names that were not in the 2010 Census list of frequently occurring surnames. Our reason for using the most common surnames only is that it allows us to get an estimate of the proportion of individuals for each last name that are homeowners - because we see how many individuals appear in Dataquick and have the total count from the Census list. The proportion of homeowners for each last name is then used to correct for selection into homeownership. The next version of this paper will use the InfoGroup data which also include renters. With this new data, there will be no reason for us to limit ourselves to the most frequently occurring last names. We will then be able to measure persistence using rare surnames as well. It should be noted though that large last names may help with classical measurement error in the 1940 housing consumption variable and typos or digitization errors in the surnames. The original sample of all male heads of households with “common surnames” and not living in group quarters contains 24,765,202 observations. After dropping missing values, there remains 23,671,908 observations.

Variable Definitions: Housing consumption is defined as the monthly rent for renters. For owner occupied housing, we divide self-reported house value by 100. This is justified by 1) the original Census questionnaire which indicates that 1 percent of the total house value is a fair monthly rental if there is no other basis for estimating the rental value of the home and 2) it corresponds to existing estimates of the price-to-rent ratio in the US in 1940. Nominal values are converted into 2012 \$ using the CPI. We use the logarithm of housing consumption in our regressions. We define three race categories: non-hispanic whites, blacks and others. In terms of geography, we focus on the 4 Census regions: Northeast, South,

Midwest and West.¹⁵ Regarding surnames, we harmonize them in the following way: we drop all special characters and numbers, leading and trailing spaces as well as consecutive blank spaces. We then convert all letters to upper case.

2.2.2 *DataQuick*

Sample Definition: DataQuick is the basis for our 2012 sample.¹⁶ It provides public records on housing characteristics and transactions (including the price) collected from county assessor and register of deeds officers. DataQuick includes a cross-section, in 2012, of house values for about 94% of all residential real estate in the US. The coverage is slightly biased towards urban areas. For each house in the dataset, whenever the primary address of the owner matches the address of the house, we consider that it is the individual’s primary residence. Finally, we keep only observations with a “frequently occurring surname” according to the 2010 Census list.

Variable Definitions: We harmonize surnames in exactly the same way as in the 1940 Census. For each last-name (records contain the first and last-name of the owner), we average the log of the value of primary residences. We convert the house value into a monthly rent by dividing by 122.69 - the house value to monthly rent ratio in the US in 2012 according to Zillow.

15. **Northeast:** Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York and Pennsylvania. **South:** Delaware, DC, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma and Texas. **Midwest:** Illinois, Indiana, Michigan, Ohio and Wisconsin. **West:** Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon and Washington.

16. IMPORTANT: By design, DataQuick only links houses to their occupant if they are owner-occupied. This means that our measures of housing consumption in 2012 only include homeowners. For this reason, in the next version of this paper, we are going to replace the DataQuick sample with another data set (InfoGroup) that contains both renters and homeowners.

2.2.3 CEX

Dataset Description: The Consumer Expenditure Survey is collected by the Bureau of Labor Statistics and used mainly for revising the CPI. We use the interview portion of the CEX. It is a rotating panel with about 5,000 families interviewed each quarter until 1998 and about 7,500 thereafter.¹⁷ Each family is interviewed up to 5 times and provide information about quarterly expenditures, annual income and demographics.

Sample Definition: We pool all years from 1980-2015 except for years 1982-1983 which had no rural sample and are therefore not nationally representative.¹⁸ We keep only respondents that have been interviewed four times. The head of household is defined as the person or one of the persons who owns or rents the unit. Each household can potentially be interviewed in two different calendar years depending on the start date. We link each observation to the year for which there is at least a 6 months overlap with the interview period. The final sample contains 137,332 households.

Variable definitions: Demographics such as age, gender, race and education of the head of household, urban/rural status, SMSA status, region, state and home ownership are measured in the third interview (corresponding more or less to the midpoint of the interview period). For most years, income is reported in the second and fifth interview only and refers to the previous 12 months. For simplicity, we take the average of income in each quarter in which there is a non-missing value. We also tried to use income reported in the fifth interview only (since it matches the period over which expenditures are reported) and results are virtually unchanged. All expenditure variables are obtained by summing quarterly expenditures reported in each of the four interviews. We closely follow Meyer and Sullivan (2017) for the definition of total expenditures, total consumption and housing consumption. Expenditures are total expenditures reported in the interview minus miscellaneous expenditures and cash

17. 13,728 families are interviewed in 1960-1961 and 19,975 families in 1972-1973.

18. In the next version of this paper, we will add data for 1960-1961 and 1972-1973.

contributions which have not been collected in all interviews. Consumption is defined as expenditures minus educational expenses (which are better described as investment), out of pocket medical expenses (which may reflect needs rather than consumption value) and payments to retirement accounts, pension plans and social security. Again, following Meyer and Sullivan (2017), housing and vehicle expenses are converted to service flows. Appendix B2 describes how those service flows are calculated (and which exact expenditures are converted). Note that we define housing consumption as the rent paid by renters (or rental equivalent for homeowners) excluding utilities and other services. This is because we only observe rent (or house value) in the Census and DataQuick. Imputations must therefore be based on this narrower definition of housing consumption. All variables are converted into 2016 real \$ using the CPI.

2.3 Raw Correlation in Housing Consumption across Last Names

Figure 1 shows the raw data: the correlation - across last names - between average log housing consumption in 2012 and the same measure in 1940.¹⁹ The OLS coefficient is 0.42 between 1940 and 2012. In other words, if the *Smiths* have, on average, a 1% higher housing consumption in 1940 relative to the *Washingtons*, they tend to have a 0.42% higher housing consumption in 2012. Assuming that there are 2 – 2.5 generations separating these two samples, this number translates into an estimated coefficient of intergenerational elasticity between two successive generations of 0.65 – 0.7.

In table 2.1, we keep only last-names for which more than 80% of members in 1940 were white and black respectively. The fourth column contains the residual last-names. The OLS coefficient between largely white names is 0.42 while it is 0.21 for largely black names. Regarding the constant term in the regression, it is 4.7 for whites and 5.6 for blacks. Figure

19. In this version, because we work with the DataQuick data, the 2012 measure only includes homeowners. We are in the process of obtaining new data - from InfoGroup - that would also include renters.

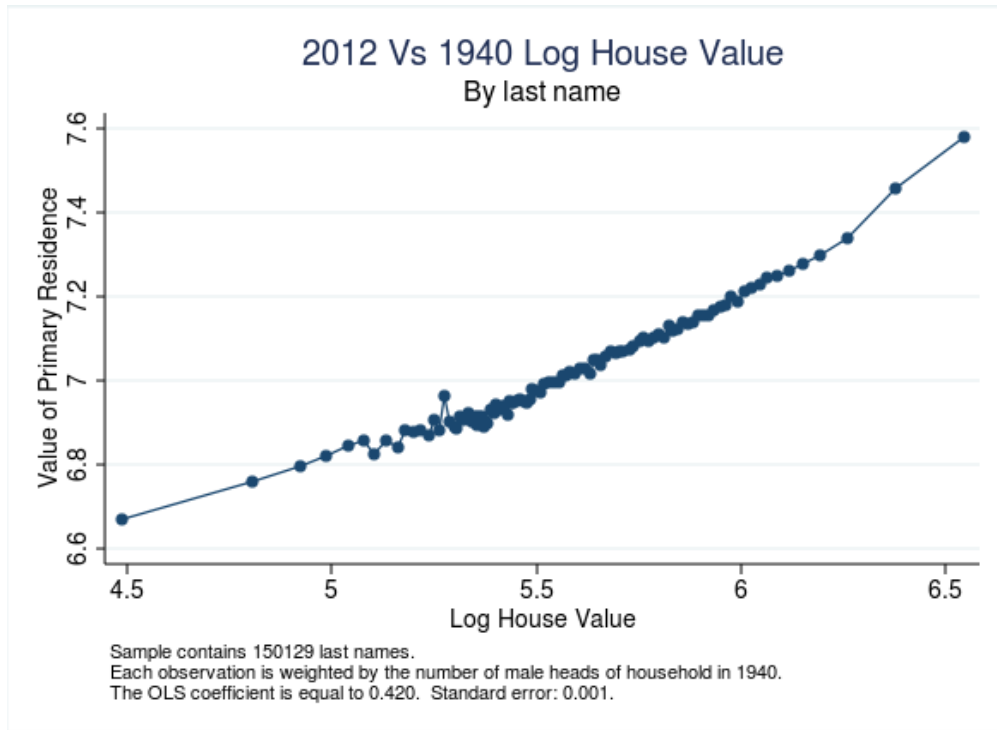


Figure 2.1: Last-name level intergenerational correlation in housing consumption.

Note: Data is divided in 100 percentiles of surname average 1940 housing consumption. The graph shows average 2012 log housing consumption per last name for each those percentiles.

Table 2.1: Surname-level correlations in Housing Consumption (By race)

	All	Mostly Whites	Mostly Blacks	Residual
β	0.420*** (0.00347)	0.419*** (0.00347)	0.208*** (0.0266)	0.350*** (0.0151)
α	4.677*** (0.0199)	4.687*** (0.0199)	5.634*** (0.125)	5.035*** (0.0777)
Observations	149905	139696	715	9494
R^2	0.535	0.521	0.126	0.194
F	14694.4	14566.9	60.98	541.1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Surname level regressions of 2012 log housing consumption (homeowners only) on 1940 log housing consumption. Surnames are weighted by the number of male heads of households in 1940. α is the intercept and β the slope coefficient. Columns 2 and 3 contain last names with at least 80% of members of a given race and column 4 contains all other last names.

B1 in the appendix is the binscatter plot corresponding to this table. We can see on figure B1 that the first percentile of housing consumption by last name is around 4.5 for whites. At that level of housing consumption in 1940, the expected consumption in 2012 would be around 6.6 for a mostly white name and 6.5 for a mostly black name. This suggests that at the last name level, upward mobility at the bottom of the distribution - conditioning on average consumption - was roughly similar for blacks and whites. Because of the large difference in slope coefficients however, richer white names have fared much better than richer black names that happened to have the same consumption in 1940. Of course, blacks were also much poorer on average which also translates into lower overall expected consumption in 2012. The difference in persistence of housing consumption suggests that richer black families in 1940 did not have access to the same insurance mechanisms - asset accumulation for instance - than white families. Another potential mechanism is that blacks have a higher discount rate than whites leading to higher consumption today and less bequests for future generations.²⁰

Table 2.2: Surname-level correlations in Housing Consumption (By region)

	Northeast	South	Midwest	West	Residual
β	0.561*** (0.0119)	0.280*** (0.0125)	0.0670*** (0.00602)	0.236*** (0.0186)	0.429*** (0.00361)
α	3.954*** (0.0716)	5.464*** (0.0631)	6.743*** (0.0353)	5.779*** (0.0957)	4.627*** (0.0206)
Observations	11028	3638	12258	1454	111318
R^2	0.342	0.294	0.015	0.299	0.533
F	2231.0	504.6	123.8	160.5	14088.4

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Surname level regressions of 2012 log housing consumption (homeowners only) on 1940 log housing consumption. Surnames are weighted by the number of male heads of households in 1940. α is the intercept and β the slope coefficient. Columns 1 – 4 contain last names with at least 80% of members of a given Census region and column 5 contains all other last names.

20. A way of testing this mechanism would be to compare the persistence for blacks that lived in places where “redlining” practices were more or less prevalent. This will be done in a later version of this paper.

Table 2.2 breaks down the estimates for mostly white names across Census regions and figure B2 in the appendix corresponds to those regressions. Mostly white names experienced more persistence in the Northeast (0.56) compared to the South (0.28), West (0.23) and especially Midwest (0.07). These patterns are noteworthy because they contrast with the findings of the previous literature on earnings persistence. For the period 1980-2012, Chetty et al. (2014) find more persistence in the South than other regions. This reversal between the Northeast and South - when looking at consumption instead of earnings - may also be driven by the financial sector being more developed in the Northeast in 1940.

Finally, it is worth noting that there is quite a bit of variance - across last-names - in log housing consumption in 1940. This raises the possibility of using last-names as a vehicle to study intergenerational mobility. In terms of the framework developed in section 4, the large between surname variance in housing consumption forms the basis of a strong *first-stage*. Between-last name correlations however do not need to be equal to their individual-level counterpart - that is the correlation that would be measured if we were able to link directly fathers to their male descendants. This is an occurrence of what is known in statistics and related fields as the *ecological inference problem*.²¹ A large part of this study is concerned with deriving the conditions under which last-name comparisons are informative of the Intergenerational Elasticity between fathers and sons and reconstructing estimates of the individual-level correlations.

Other studies have used last-names to study intergenerational mobility.²² Clark (2014) compiles evidence from a large range of countries and periods of time. He finds remarkably consistent measures of between surname intergenerational correlations across time and countries. The implied elasticity is on the order of 0.7–0.8. Clark argues that any single measure

21. On the ecological inference problem, see King (1997, 2004) and among others.

22. See among others Collado et al. (2012), Clarke and Cummins (2015), Clarke (2014), Olivetti and Paserman (2015) and Guell et al. (2015).

- for instance income, education or wealth - usually considered in the literature is merely a noisy signal of some underlying social status. By looking at surnames, Clark argues, one encompasses other dimensions of class and social status which explains why long-run surname regressions tend to show less intergenerational regression to the mean. Chetty et al. (2014) - in appendix B - compare their individual estimates to surname based estimates in their tax data. They compute both measures for samples of more and less common last names. The surname based estimates tend to be slightly larger than individual-level estimates when excluding the most common last names. The estimates tend to blow up when restricting the sample to the very few most common last names and then attain a magnitude roughly comparable to the between race correlation between the two periods considered. Chetty and coauthors conclude that last names are informative about ethnicity and that surname based estimates place more weight on the between race component of the covariance. We make a similar point and adjust our estimates for race in section 4. Also note that the last name correlations by race presented in table 2.1 (and by region in table 2.2 since it includes only mostly white names) are within-race estimates. Olivetti and Paserman (2015) is an ingenious study using the informativeness of received first names regarding one's parents social status to estimate social mobility in the late nineteenth and early twentieth century. It is, as pointed by the authors themselves a "Two Sample Two Stage Least Square" estimator and is similar in that regard to our methodology. Our methodology differs from theirs in that we use surnames - instead of first names - as instruments in the estimation. We also allow for heterogeneity across demographics and provide placebo tests aimed at testing the exclusion restrictions associated with the use of last names as instruments. Guell et al. (2015) use a calibrated model of the joint evolution of the distribution of surnames and economic advantage. Using cross-sectional data and the calibrated structural model, they recover intergenerational elasticity measures.

2.4 From Surnames to Individual-level Correlations

We have previously noted that last-name level correlations need not be equal to their individual level counterparts. In this section, we develop an econometric framework to clarify under what assumptions they coincide and how to evaluate the plausibility of the underlying assumptions. We are interested in estimating - at the individual level - the population Best Linear Predictor (BLP) of 2012 (log-)consumption using the (log-)consumption of one's ancestor in 1940. We will refer to this statistics as the Intergenerational Elasticity of Consumption (IGC) by analogy to the literature on persistence of earnings across generations. To be clear, we are interested in parameter β in the following population regression:

$$c_{i,t+1} = \alpha + \beta c_{i,t} + \varepsilon_{i,t+1} \quad (2.1)$$

In practice, we do not observe directly c but use housing consumption, h or imputed consumption c^* as a proxy. We discuss these considerations in the next section. By definition of β being the population best linear predictor (or linear projection) and ε being the prediction error in that population regression, the following moment conditions hold:

$$E [c_{it}\varepsilon_{it+1}] = 0$$

$$E [\varepsilon_{it+1}] = 0$$

Of course, β in the regression above is not the causal effect of consumption in t on consumption in $t + 1$. In order to fix ideas, it might be helpful to consider the following very general data generating process (DGP) and derive β as a function of these structural parameters:²³

23. Of course, this data generating process is too general in the sense that - with two periods - one could never hope to estimate such a model even if in possession of a dataset containing the entire population and intergenerational links. This DGP will be useful to discuss exclusion restrictions and their relation to group-level heterogeneity in our GMM estimation.

$$c_{i,t+1} = \delta_{0i} + \delta_{1i}c_{i,t} \tag{2.2}$$

The latter model allows for individual-specific heterogeneity in the intercept and causal effect of time t consumption. Under such a DGP, β , our parameter of interest, is:

$$\beta = E[\delta_{1i}] + \frac{Cov(\delta_{0i}, c_{it})}{Var(c_{it})} + \frac{Cov((\delta_{1i} - E[\delta_{1i}])c_{it}, c_{it})}{Var(c_{it})}$$

The formula highlights that β is larger 1) if the average causal effect of time t consumption is large and 2) if individuals with large initial c_{it} draw large intercept values and/or causal effect. Though β does not recover the causal transmission of consumption across generations, it is a parameter of interest for describing mobility (and insurance) across generations.

The issue that we aim to solve in this paper is one of missing data: though we observe a proxy of consumption for the entire US population in 1940 and for the universe of homeowners in 2012, we do not observe individual linkages between observations in the two samples. In other words, we are unable to directly link individuals in the 2012 sample to their ancestor in 1940. Both samples do however include each individual's last-name. Consider a set of dummy variables, $\{z_i^\ell\}$, one for each last-name indexed by ℓ . We use last-names as an instrument to solve the missing data issue. The estimator developed below is a Two Sample Instrumental Variable (2SIV) estimator as presented in Ridder and Moffitt (2007) and first introduced by Klevmarken (1982). We start from the following moment conditions:

$$E[(c_{t+1} - \beta c_t - \alpha) z^\ell] = 0 \quad \forall \ell$$

Written in terms of the error term in the linear projection equation :

$$E \left[\varepsilon_{t+1} z^\ell \right] = 0 \quad \forall \ell$$

The identifying assumption is that last names are uncorrelated with the error term in the linear projection equation. We will evaluate this assumption later in the section.

Denote by Y the $N_2 \times 1$ matrix where the j^{th} element is equal to $c_{j,t+1}$, X the $N_1 \times 2$ matrix of regressors where the i^{th} row is equal to $[1 \ c_{it}]$, Z_2 and Z_1 the $N_2 \times L$ and $N_1 \times L$ matrices of last name dummies in 2012 and 1940 respectively and $B = [\alpha \ \beta]^T$. The sample analog of these moment conditions can be computed from our two independent samples.

$$m_N(B) = \frac{1}{N_2} Z_2^T Y - \frac{1}{N_1} Z_1^T X B = 0$$

The 2SIV estimator minimizes $m_N(B)^T W_N m_N(B)$ with a weighting matrix $W_N \rightarrow_p W$ and W is positive definite. Solving the minimization problem, the solution is:

$$\hat{B}_N = \left(\frac{1}{N_1} X^T Z_1 W_N \frac{1}{N_1} Z_1^T X \right)^{-1} \left(\frac{1}{N_1} X^T Z_1^T W_N \frac{1}{N_2} Z_2^T Y \right)$$

The expression looks familiar as it is analogous to a GMM estimator. The probability limit as $N_1 \rightarrow \infty$ and $N_2 \rightarrow \infty$ is

$$\hat{B}_N \rightarrow_p B + \left(\frac{1}{N_1} X^T Z_1 W_N \frac{1}{N_1} Z_1^T X \right)^{-1} \left(\frac{1}{N_1} X^T Z_1 W_N \frac{1}{N_2} Z_2^T \Xi \right)$$

where Ξ is the $N_2 \times 1$ matrix of projection errors $\varepsilon_{j,t+1}$. Under the assumed exclusion restriction that $E \left[\varepsilon_{t+1} z^\ell \right] = 0$, B is a consistent estimator of α and β . Also notice that matrix $\left(\frac{1}{N_1} X^T Z_1 W_N \frac{1}{N_1} Z_1^T X \right)$ must be invertible - in other words there must be enough variation in consumption across last names. It is worth examining the exclusion restriction and its plausibility. Concerns for the exclusion restriction are more transparently expressed using the causal model in Eq. (2.2). We show in Appendix B3 that a sufficient condition for $E \left[\varepsilon_{i,t+1} z_i^\ell \right] = 0$ is:

$$\frac{Cov(\delta_{0i}, z^\ell)}{var(z^\ell)} = \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} \frac{Cov(c_{it}, z^\ell)}{var(z^\ell)}$$

and

$$\frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell)}{Var(z_i^\ell)} = \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})} \frac{Cov(c_{it}, z^\ell)}{var(z^\ell)}$$

Intuitively, it compares (focusing on the first of the two conditions) the coefficient of a regression of δ_{0i} on z_i^ℓ to the product of the coefficients resulting from regressing δ_{0i} on c_{it} and c_{it} on z_i^ℓ respectively. A concern arises if last names are very informative - through a large between surname variance for instance - about an omitted variable that correlates strongly with δ_{0i} or δ_{1i} :

$$\frac{Cov(u_i, z^\ell)}{var(z^\ell)} >> \frac{Cov(u_i, c_{it})}{var(c_{it})} \frac{Cov(c_{it}, z^\ell)}{var(z^\ell)}$$

and $Cov(u_i, \delta_{0i}) \neq 0$ or $Cov(u_i, \delta_{1i}) \neq 0$. This would lead to surname level estimates being different from their individual level counterpart. A concern is that last names contain information about other dimensions of social classes such as ethnicity, geography or social status and that those correlate with $\{\delta_{0i}, \delta_{1i}\}$. Finally, we can re-arrange the above sufficient conditions:

$$\frac{Cov(\delta_{0i}, z^\ell)}{Cov(c_{it}, z_i^\ell)} = \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} \tag{2.3}$$

and

$$\frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell)}{Cov(c_{it}, z_i^\ell)} = \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})} \tag{2.4}$$

	Variable	Between OLS	Std. Error	Overall OLS	Std. Error
Immigration	Foreign	0.223	(0.005)	0.031	(0.000)
Geography	Northeast	0.347	(0.003)	0.081	(0.000)
	Midwest	0.148	(0.004)	0.022	(0.000)
	South	-0.473	(0.005)	-0.110	(0.000)
	West	-0.022	(0.004)	0.006	(0.000)
	Urban	0.288	(0.003)	0.142	(0.000)
	County Wage	0.448	(0.003)	0.166	(0.000)
Race	white	0.171	(0.007)	0.049	(0.000)
	Black	-0.165	(0.007)	-0.047	(0.000)
	Hispanic	-0.070	(0.007)	-0.005	(0.000)
	Asian	-0.001	(0.000)	-0.000	(0.000)

Table 2.3: Placebo test: Immigration, Geography and race

Note: Column 2 (“Between OLS” represents the coefficient of a regression - at the surname level - of the variable in the first column on log housing consumption in 1940 while column 4 (“Overall OLS”) represents the coefficient from the same regression ran at the individual level.

The re-arranged conditions suggest a placebo test. Let us focus on Eq. (2.3) for instance. The term on the left-hand side is the estimate resulting from a regression of δ_{0i} on c_{it} using z_i^ℓ as an instrument while the right-hand side is the corresponding OLS coefficient. We do not observe δ_{0i} directly but we suspect that it is correlated to observables such as race, geography, education or social status. These variables are all included in our 1940 sample and we can thus regress such observables on 1940 consumption using both OLS (by analogy to the right-hand side) and IV (by analogy to the left-hand side). In practice, the 2SIV estimator uses a weighted sum of all L moment conditions rather than one instrument z_i^ℓ . We discuss the choice of weighting matrix in the appendix. Our estimates use an easily interpretable weighting across last-names such that our estimator is a “Two-Sample Two Stage Least Square” (2S2SLS).²⁴ The left-hand side becomes equivalent to running OLS on last-name averages of δ_0 and c_t . As a result, the placebo tests will coincide with evaluating whether the between surname and individual OLS coefficients are equal.

Results of this Placebo test can be seen on tables 2.3 and 2.4. The difference between the

24. See Inoue and Solon (2010).

	Variable	Between OLS	Std. Error	Overall OLS	Std. Error
Education	Education	1.357	(0.030)	0.971	(0.001)
Work	Occ score	4.519	(0.029)	2.800	(0.002)
	Occ. prestige	3.384	(0.064)	2.280	(0.002)
	Ind. wage	0.268	(0.001)	0.147	(0.000)
	Farm	-0.200	(0.002)	-0.100	(0.000)
	Employer	0.013	(0.000)	0.007	(0.000)
	Self-employed	-0.062	(0.002)	-0.039	(0.000)
	Private Sector	0.068	(0.002)	0.045	(0.000)
	Public Sector	-0.017	(0.001)	-0.005	(0.000)
Family	Married	0.001	(0.001)	0.018	(0.000)
	N. children	-0.219	(0.008)	-0.116	(0.000)

Table 2.4: Placebo test: Education, Work and Family

Note: Column 2 (“Between OLS” represents the coefficient of a regression - at the surname level - of the variable in the first column on log housing consumption in 1940 while column 4 (“Overall OLS”) represents the coefficient from the same regression ran at the individual level.

individual level and last name level OLS indicate that direct correlations between last-names and ethnicity, geography and immigration status are a threat to identification. To a somewhat lesser extent, direct correlations between last-names and other variables - education, work and family - that may be related to social status are also a concern. To be clear, these correlations only matter to the extent that these dimensions are also correlated with δ_{0i} and δ_{1i} . The literature on intergenerational elasticity of earnings has already documented wide heterogeneity in IGE between places and races which suggests such correlations.²⁵ As a result, the exclusion restrictions presented above appear too strong. We therefore propose another estimator that adjusts for dimensions of heterogeneity that are likely to correlate with surnames and is based on the weaker exclusion restrictions that

$$\frac{Cov\left(\delta_{0i}, z_i^\ell \mid \{D_i^g\}\right)}{Cov\left(c_{it}, z_i^\ell \mid \{D_i^g\}\right)} = \frac{Cov\left(\delta_{0i}, c_{it} \mid \{D_i^g\}\right)}{var\left(c_{it} \mid \{D_i^g\}\right)}$$

and

25. See Chetty, et al (2014) for geography and Chetty et al. (2018) for race.

$$\frac{\text{Cov} \left((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell \mid \{D_i^g\} \right)}{\text{Cov} \left(c_{it}, z_i^\ell \mid \{D_i^g\} \right)} = \frac{\text{Cov} \left((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it} \mid \{D_i^g\} \right)}{\text{Var} \left(c_{it} \mid \{D_i^g\} \right)}$$

where $\{D_i^g\}$ is a set of dummy variables indicating that an observation belongs to group g . An obvious case in which those conditions are satisfied is if $\delta_{0i} = \delta_0^g$ and $\delta_{1i} = \delta_1^g \forall i$ s.t. $D_i^g = 1$. In practice, we define the groups as the intersection of geography and ethnicity. How to define the proper groups is an important question - the partition needs to be fine enough that the above exclusion restriction plausibly holds but coarse enough that the “first-stage” remains strong - and we discuss it below. For now, let us consider that we have a partition into G mutually exclusive groups. We are interested in estimating:

$$c_{i,t+1} = \alpha^g + \beta^g c_{i,t} + \tilde{\varepsilon}_{i,t+1} \quad (2.5)$$

The estimator is similar to the one presented above. It is now based on the following exclusion restriction:

$$E \left[\left(c_{it+1} - \sum_{g=1}^G (\alpha^g + \beta^g c_{it}) D_i^g \right) z_i^\ell \right] = 0 \quad \forall \ell \quad (2.6)$$

We modify the X matrix such that it is $N_1 \times 2G$ with i^{th} row equal to $\left[D_i^{g1} \ D_i^{g2} \ \dots \ D_i^G \ c_{it} D_i^{g1} \ c_{it} D_i^{g2} \ \dots \ c_{it} D_i^G \right]$. We also modify the B matrix so that it is $\left[\alpha_{g1} \ \alpha_{g2} \ \dots \ \alpha^G \ \beta^{g1} \ \beta^{g2} \ \dots \ \beta^G \right]^T$. The (two) sample(s) analog of these moment conditions is:

$$m_N(B) = \frac{1}{N_2} Z_2^T Y - \frac{1}{N_1} Z_1^T X B = 0$$

Minimizing the quadratic $m_N(B)^T W_N m_N(B)$ leads to the following estimator:

$$\hat{B}_N = \left(\frac{1}{N_1} X^T Z_1 W_N \frac{1}{N_1} Z_1^T X \right)^{-1} \left(\frac{1}{N_1} X^T Z_1^T W_N \frac{1}{N_2} Z_2^T Y \right)$$

We discuss the choice of the weighting matrix, W_N and the derivation of asymptotic standard errors in the appendix. In practice, we choose the weighting matrix corresponding to a “Two Sample Two Stage Least Square” estimator. Each (α^g, β^g) parameter is interesting in its own right and we will comment on them in the results section. We are also interested however in recovering the overall Intergenerational Elasticity of Consumption. The overall IGC, labeled $\hat{\beta}$, can be re-constructed from the following formula:

$$\hat{\beta} = \frac{Var(\bar{c}_{gt})}{Var(c_{it})} \hat{\beta}^G + \sum_g \frac{\frac{N_g}{N} Var_g(c_{it})}{Var(c_{it})} \hat{\beta}^g \quad (2.7)$$

with

$$\hat{\beta}^G = \frac{Cov(\bar{c}_{gt}, \hat{\beta}^g \bar{c}_{gt} + \hat{\alpha}^g)}{Var(\bar{c}_{gt})}$$

$\hat{\beta}^G$ is the “between-goup” intergenerational elasticity, \bar{c}_{gt} denotes the mean and $Var_g(c_{it})$ the variance of 1940 consumption within group g .

In order to test whether adjusting for covariates such as race and geography is likely to deliver the correct estimates, we reproduce the placebo tests from table 2.4 but we add the covariate adjustment presented above. These placebo tests can be interpreted in two ways. First, it is a test of conditions (2.3) and (2.4) - this time, after adjusting for race and/or geography - to the extent that the variables included in the placebo test (education, work and family) are correlated with δ_{0i} or δ_{1i} . The second (related) interpretation is that the placebo tests act as a “laboratory” for our adjustment: they compare the coefficients of an individual-level regression with the corresponding last name level regression (after performing our adjustment). By the analogy principle, if surname- and individual- level estimates are close in the placebos, we also expect them to be close in the intergenerational regression.

Table 2.5: Placebo Test: Comparison of true OLS with (covariates adjusted - race only) between last-names OLS

	(1)	(2)	(3)
Education			
Education	1.061	1.334	1.046
Work			
Occ. Score	3.212	5.017	4.756
Occ. Prestige	2.495	3.368	2.336
Farmer	-0.110	-0.197	-0.229
Employer	0.008	0.013	0.010
Self-employed	-0.042	-0.063	-0.086
Private Employee	0.049	0.065	0.086
Public Employee	-0.005	-0.018	-0.013
Family			
Married	0.021	0.002	-0.010
N. Children	-0.125	-0.214	-0.228

Columns:

(1) True OLS

(2) Between Last Names OLS - Unadjusted

(3) Between Last Names OLS - Adjusted for covariates (race only)

Note: Coefficients from regression of the variable in the first column on log housing consumption in 1940 at the individual- and surname level. Column 3 contains coefficients corresponding to formula (7) after estimating race specific parameters. All last names are weighted by their number of male heads in 1940.

We present two covariates adjustments. The first one adjusts for race only (we divide the sample into three races: non-hispanic whites, blacks and others). The second adjustment divides the sample into 12 mutually exclusive groups defined by the intersection of the four Census regions (Northeast, Midwest, South and West) with the three race categories defined above.

We can see in table 2.5 that the adjustment for race works well and helps bring the between surnames OLS coefficient closer to its true value. The difference is striking especially for education and variables related to occupational prestige. This is noteworthy because these variables are most likely to be correlated with future consumption. Table 2.6 serves as a cautionary tale: adjusting for more covariates does not necessarily always help the estimation. When adjusting for both regions and race, the between last names OLS can be

Table 2.6: Placebo Test: Comparison of true OLS with (covariates adjusted -) between last-names OLS

	(1)	(2)	(3)
Education			
Education	1.061	1.334	1.971
Work			
Occ. Score	3.210	5.017	5.337
Occ. Prestige	2.503	3.368	4.806
Farmer	-0.110	-0.197	-0.168
Employer	0.008	0.013	0.019
Self-employed	-0.042	-0.063	-0.044
Private Employee	0.049	0.065	0.045
Public Employee	-0.005	-0.018	-0.016
Family			
Married	0.021	0.002	0.004
N. Children	-0.125	-0.214	-0.393

Columns:

(1) True OLS

(2) Between Last Names OLS - Unadjusted

(3) Between Last Names OLS - Adjusted for covariates

Note: Coefficients from regression of the variable in the first column on log housing consumption in 1940 at the individual- and surname level. Column 3 contains coefficients corresponding to formula (7) after estimating race and region specific parameters. All last names are weighted by their number of male heads in 1940.

Table 2.7: Long-run Mobility Estimates by Race

	α	β	1940 Cons.	2012 Cons.
Race				
Whites	4.886 (0.02)	0.387 (0.003)	5.668 (0.004)	7.078 (0.002)
Blacks	6.138 (0.083)	0.071 (0.019)	4.631 (0.003)	6.466 (0.012)
Other	4.896 (0.09)	0.394 (0.019)	4.84 (0.019)	6.8 (0.015)
All	4.996 (0.019)	0.372 (0.003)	5.564 (0.006)	7.02 (0.004)

Note : Estimates of mobility by race: regression of 2012 log housing consumption on 1940 log housing consumption - using surnames as instruments. α is the intercept coefficient and β the slope. 2012 average consumption is an estimate using α , β and mean housing consumption in 1940.

inflated away from the individual-level correlation. It is most striking again for education and variables related to occupational prestige. In appendix B5, we present more evidence on the performance of the estimator when estimating heterogeneity by race and region including simulations and evaluations of the sensitivity of the placebo test to the size of last names.

Table 2.7 presents race-specific and overall intergenerational elasticities after adjusting for race. Table 2.5 suggests that this adjustment may perform well and resulting estimates should be close to individual-level correlations. The Intergenerational Elasticity of Housing Consumption between 1940 and 2012 is 0.37 for the US population as a whole. Assuming that 2 – 2.5 generations separate the two samples, this corresponds to a 0.6 – 0.67 elasticity between two successive generations. Bootstrapped standard errors are in parenthesis.²⁶ We also present heterogeneity in consumption persistence across races. Consistent with table 2.1, persistence is lower among blacks - but roughly similar for non-hispanic whites and other races.

26. We “block bootstrap” at the last-name level. Asymptotic standard errors are also easy to compute for race specific parameters but a bootstrap is needed to make inference about the overall IGC given its nonlinear formula.

The tables also present estimates of the race specific intercept parameter (α^g) and 2012 average expected housing consumption. In DataQuick, we do not observe the ethnicity of individuals in 2012.²⁷ The average expected consumption in 2012 is therefore not raw data - it is an estimate based on average consumption in 1940 and the estimates of α^g and β^g . Overall, the racial gap in housing consumption has decreased but is still large in 2012: on average whites consumed twice as much housing as blacks in 1940, their consumption is still 60% higher in 2012. The intercept α^g represents the expected 2012 log consumption of an individual with zero log consumption in 1940. Since nobody has zero consumption, it is better for the sake of interpretation to sum α^g and $3.48\beta^g$ to obtain the race-specific expected log consumption of an individual at the 5th percentile of the consumption distribution. Expected consumption in 2012, for the bottom of the 1940 distribution is roughly comparable across races: 6.23 (whites), 6.39 (blacks) and 6.27 (other). Of course, given the much lower β for blacks, they fare considerably worse than other races at higher ranks in the distribution. These patterns are very similar to the heterogeneity uncovered in the raw correlations from section 3.

2.4.1 Regional Heterogeneity

In order to better understand the dynamics of consumption mobility and the stark differences between races, it is informative to also explore heterogeneity across regions in the US. We define D_i^g as dummies interacting race and Census region and apply the methodology outlined above. Estimated parameters α^g and β^g are then combined with the shares of each group g and measures of between- and within-group variances to reconstruct race-specific, region-specific and overall correlations. The placebo test of table 2.6 suggests to be cautious with those results - when adjusting for many covariates the resulting estimates may be different from individual-level correlations. In appendix B5, we explore the sources of

27. In the next version of this study, using InfoGroup data, we will observe individuals race in 2012 and will be able to compute average housing consumption directly from the data.

Table 2.8: Mobility Estimates (β)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	0.67 (0.009)	0.2 (0.005)	0.245 (0.007)	0.096 (0.011)	0.311 (0.004)
Black	0.071 (0.099)	-0.093 (0.093)	0.206 (0.024)	0.024 (0.193)	0.08 (0.019)
Other	-0.119 (0.079)	0.209 (0.034)	0.254 (0.036)	0.497 (0.023)	0.368 (0.015)
All	0.654 (0.009)	0.202 (0.006)	0.247 (0.007)	0.129 (0.01)	0.308 (0.003)

Note : Estimates of mobility by race and region: regression of 2012 log housing consumption on 1940 log housing consumption - using surnames as instruments. β is the slope coefficient. Observations are weighted by number of male heads of households.

variation in surname averages of housing consumption across races and regions and attempt to diagnose why the placebo test performs poorly when adjusting for both Census regions and races. First, simulations from a model under which the exclusion restrictions of section 4 are satisfied - and drawing from the actual data for generating the right-hand side variables - suggest that there is enough variation in the data to estimate the α^g and β^g parameters. This is true when using either the full sample, either only “rare last names” (less than 50 male heads of households in 1940) or using only “common last names”. We also check whether varying the set of surnames used in the estimation (iteratively dropping more and more of the most common surnames) helps with estimating the parameters. In general, placebos perform better when using smaller last names but the increase in performance is slow relative to size of last names. Keeping the poor performance of the placebo tests in mind, we analyze the region specific results presented in table 2.8 and 2.9.

When adjusting for Census regions - in addition to race - we estimate a slightly lower overall intergenerational persistence (around 0.31 which translates to 0.56 – 0.63 over one generation) and similarly a slightly lower persistence for whites (0.31). The heterogeneity

Table 2.9: Mobility Estimates (α)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	3.248 (0.055)	5.889 (0.031)	5.575 (0.034)	6.583 (0.065)	5.181 (0.024)
Black	6.44 (0.614)	6.037 (0.501)	5.679 (0.101)	5.618 (1.061)	5.789 (0.108)
Other	7.942 (0.501)	5.47 (0.179)	5.174 (0.166)	4.713 (0.108)	5.335 (0.092)
All	3.39 (0.056)	5.891 (0.035)	5.59 (0.034)	6.456 (0.062)	5.237 (0.022)

Note : Estimates of mobility by race and region: regression of 2012 log housing consumption on 1940 log housing consumption - using surnames as instruments. α is the intercept. Observations are weighted by number of male heads of households.

across races is preserved: blacks have lower insurance than whites. Looking at regional patterns, we see that persistence for whites is higher in the Northeast compared to other regions. This may reflect the existence of better financial markets in the Northeast allowing individuals to accumulate assets and insure future generations. The white-black gap is also larger in the Northeast suggesting - again - that blacks did not enjoy the same access to credit markets. A somewhat surprising result is that persistence is roughly similar for blacks and whites in the South. The finding of higher persistence (lower mobility) in the Northeast compared to the South is in stark contrast with estimates of mobility in earnings between 1986 - 2015 obtained in Chetty et al. (2014). They find that the South is the least mobile region while the Northeast has high levels of upward and relative mobility. The west appears very mobile in both earnings and consumption. While institutions in the Northeast may favor higher mobility in earnings and human capital - the existence of strong financial markets seem to also generated a higher persistence in consumption. The role of financial institutions in providing consumption insurance for future generations needs to be explored further. In a later version of this study, we plan to study whether places that were “hit”

by a large number of bank runs in the 30s have lower levels of intergenerational insurance between 1940 and 2012.

2.5 Housing Consumption as a Proxy for Total Consumption

Though intergenerational correlations in housing consumption are interesting in their own right, we would like to translate these numbers into intergenerational elasticities of overall consumption. We use the CEX, a nationally representative sample, to estimate the following log-linear housing demand equations:

$$h_{it} = \gamma_{0t} + \gamma_{1t}c_{it} + \nu_{it} \tag{2.8}$$

$$h_{it+1} = \gamma_{0t+1} + \gamma_{1t+1}c_{it+1} + \nu_{it+1} \tag{2.9}$$

where h is log housing consumption, c is log total consumption and t and $t + 1$ denote 1940 and 2012 respectively.

Figure 2.2 shows strong support for the assumption of log linearity. It plots the mean log housing consumption for each percentile of total consumption. The relationship is remarkably linear - including in the tails - and has a slope of 0.81. The R-squared is .57. The measure of housing consumption used in figure 2.2 includes only the rent (or rent equivalent for homeowners) because it is ultimately what we are able to measure in the 1940 Census and DataQuick. The same figure for a more comprehensive measure of housing consumption that includes utilities and home services has a higher slope (0.92) and higher R-squared (0.81).²⁸ This more comprehensive measure would be an even better proxy and preferences for housing as a whole seem very close to homothetic. To the extent that utilities and other home services (appliances, maids,...) are luxuries, the rent equivalent itself (“shelter”) is a

28. These plots are available on request to the authors.

necessity and has a slope slightly lower than 1.

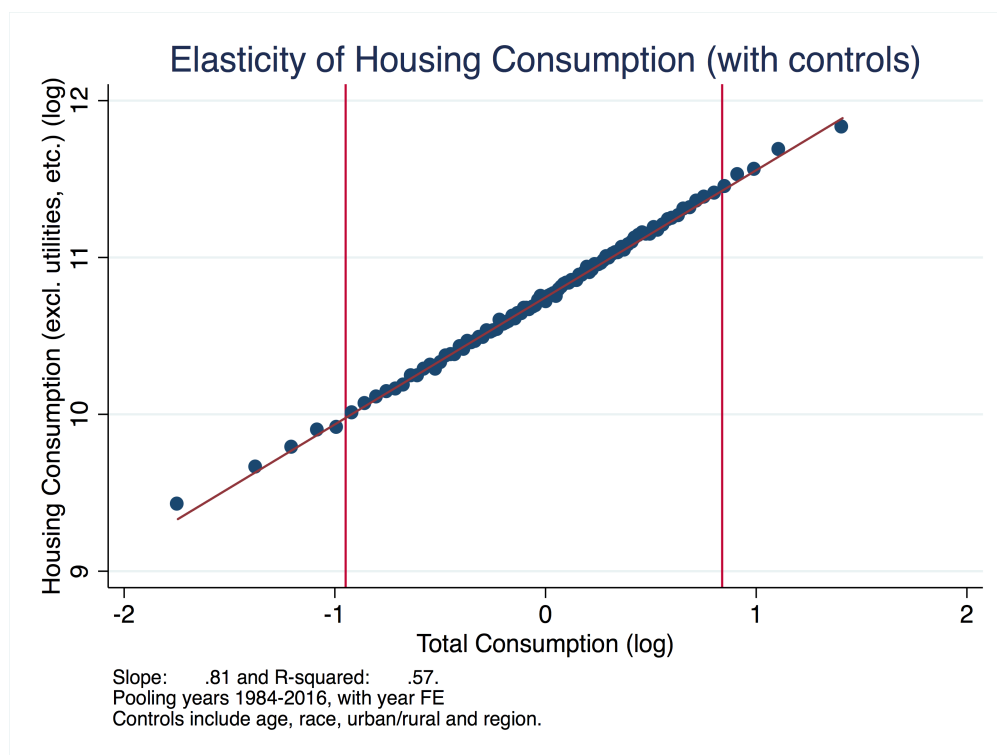


Figure 2.2: Engel Curve for Housing Consumption

There are obvious endogeneity issues with running OLS for estimation of Engel Curves. Housing consumption and total consumption are endogeneously determined by a utility maximizing agent. Moreover, the total consumption measures are self-reported and sometimes include imputed service flows from durables (see appendix B2). As a result, the independent variable, total log consumption, may be plagued with (hopefully) classical measurement error. In order to get around this endogeneity issue, we also run a 2SLS estimation in which total consumption is instrumented by the average salary, in the current year, of individuals with the same education and birth year as the head of household. Results of the IV estimation are shown in table 2.10.

There are two ways to impute overall consumption in the 1940 and 2012 samples based on these Engel curve estimations. First - our preferred solution - one can estimate Engel curves in the CEX and allow γ_{0t} and γ_{1t} to vary across observable demographics and year. A proxy for overall consumption can then be formed in each sample equal to:

Table 2.10: Demand estimations

	OLS	OLS	IV	IV
Tot. Cons. (log)	0.808*** (0.00301)	0.810*** (0.00327)	0.681*** (0.00751)	0.874*** (0.0108)
Observations	118697	114021	114920	110488
R^2	0.517	0.566	0.506	0.563
F	72250.1	61287.8	8241.5	6546.0
Controls		x		x

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

IV specification instrument total consumption with cohort-year-education average salary.

Controls include year, age, race, urban and region FE.

$$c_{it}^* = -\frac{\gamma_{0t}(x_{it})}{\gamma_{1t}(x_{it})} + \frac{1}{\gamma_{1t}(x_{it})}h_{it}$$

where x_{it} are demographics that are observable in both the CEX and housing consumption sample at time t . We can then run the surname level regressions on this proxy instead of housing consumption. This is clearly our preferred method but it requires observing demographics x_{it} in all samples at time t . Because Dataquick is short on demographics other than location we cannot implement these imputations. The next version of this paper using the InfoGroup data will contain those imputations because more details are available in the latter dataset.

In appendix B6, we discuss under what conditions using housing consumption as a proxy for total consumption leads to consistent estimates of the intergenerational elasticity of *total* consumption. Consistency requires last names to be a “valid” instrument that is uncorrelated with housing preferences.²⁹

In case last names are correlated with housing preferences, the estimate is a weighted

29. A sufficient condition would be that $\nu_{it} \perp z_i^\ell \forall \ell$. We show in appendix B6 that we actually need a weaker condition: $\sum_\ell \omega_\ell cov(\nu_{it}, z_i^\ell) = 0$. This is similar, in spirit, to Kolesár et al. (2015) in that multiple invalid instruments can recover the parameter of interest if the violation of the exclusion restriction is quasi-random across the many instruments.

average between the true elasticity of housing consumption and the intergenerational persistence of housing preferences. This leads to a downward bias if preferences are less persistent across generations than consumption (and vice-versa).³⁰ As the formula makes clear in appendix B6, a larger R-squared of the Engle curve regressions lead to a smaller bias (in the sense that more weight is then placed on the true parameter of interest β). Intuitively, if the R-squared is close to 1, measuring housing consumption is “almost as good” as observing total consumption. In that regard, it is encouraging - but far from perfect - that the Engle curve regressions have a reasonably large R-squared of 0.57. Finally, notice that if we had run OLS - which is infeasible in our dataset - using housing consumption as a proxy for total consumption, the OLS coefficient would be biased downward due to the fact that the proxy is a (classical) error-ridden measure of the true regressor.³¹ The use of last names solves the measurement error problem if last names are uncorrelated with housing preferences. Another way to think about this is that - under that assumption - housing preferences ν_{it} will be averaged out when we form last-name averages of our proxy c_{it}^* . This argument tends to favor the use of more common last names to average over larger numbers of individuals.

To the extent that there is little heterogeneity in Engel curves across observables, we can also simply run the intergenerational regression on housing directly and scale the resulting coefficient by $\frac{\gamma_{1t}}{\gamma_{1t+1}}$ instead of doing the full imputation. Under the assumption that last names are uncorrelated with housing preferences and that heterogeneity in γ_1 is negligible, we would recover the true β .

Figure 2.3 and 2.4 show that there is little heterogeneity in Engel Curves across regions and races. This suggests that scaling the raw housing estimate by $\frac{\gamma_{1t}}{\gamma_{1t+1}}$ may work relatively well and yield estimates that are close to the full imputations. Figure 2.5 shows the Engel

30. This is true if the weight Ω is smaller than 1.

31. We show the complete formula in appendix B6. The attenuation bias may be canceled by another source of (upward) bias coming from potential intergenerational persistence in housing preferences.

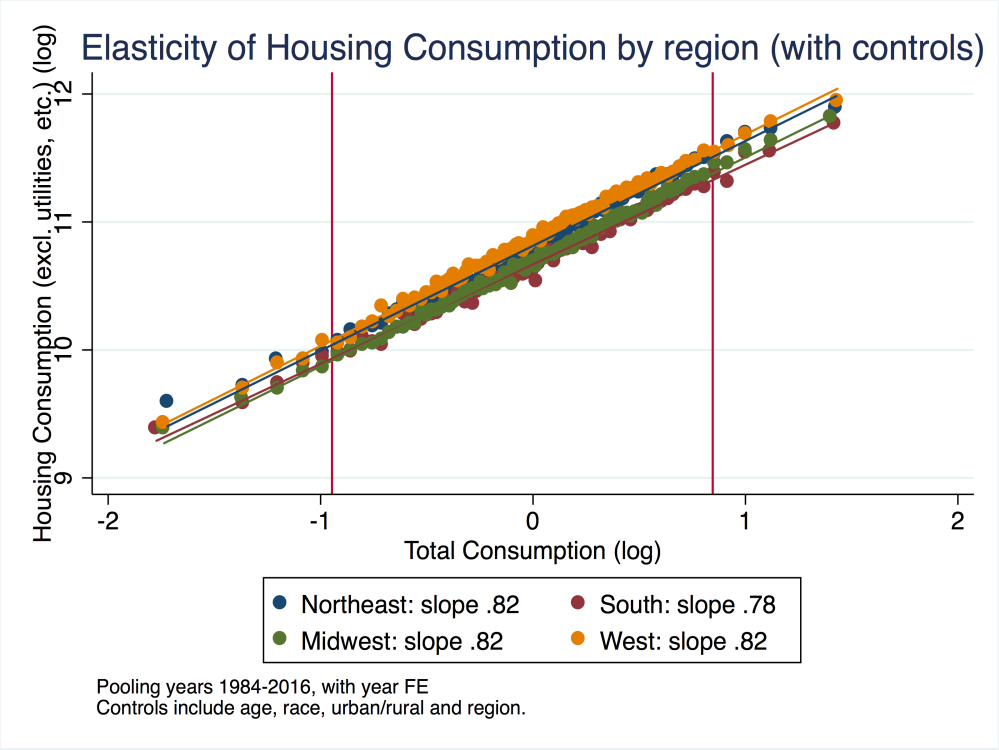


Figure 2.3: Engel Curve for Housing Consumption (Separately by Region)

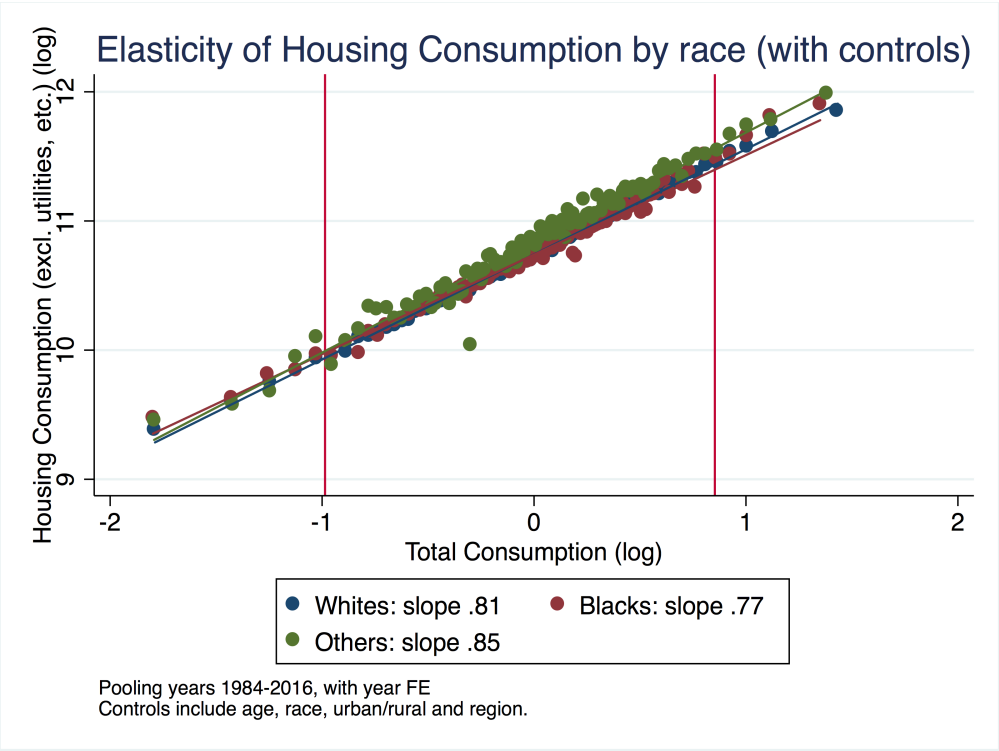


Figure 2.4: Engel Curve for Housing Consumption (Separately by Race)

curve estimation by decade. Again, we uncover relatively little changes in γ_{1t} over time. The raw housing correlations almost need no re-scaling. Assuming that the 1940 Engel curve is closest to its shape in 1980, raw housing consumption correlations can be rescaled by a factor of $\frac{\gamma_{1t}}{\gamma_{1t+1}} = \frac{0.77}{0.82} = 0.94$.

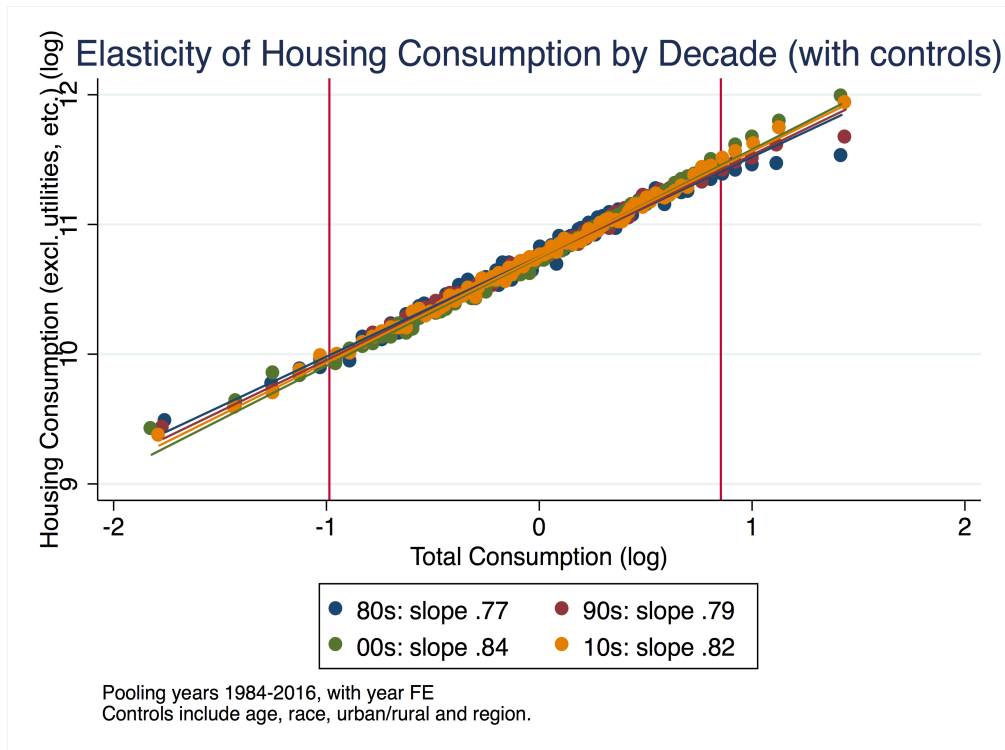


Figure 2.5: Engel Curve for Housing Consumption (Separately by Decade)

There are precedents in the literature for imputing total consumption based on narrower categories. One such early example is Skinner (1987) who estimates a linear prediction of total consumption in the CEX based on housing consumption, utilities, number of automobiles and food at home and away from home. The coefficients from this regression are then used to impute total consumption in the PSID based on the same items. More recently, Blundell, Pistaferri and Preston (2008) estimate a demand function for food expenditure in the CEX. They invert the estimated Engel curve to impute nondurable consumption in the PSID. They study inequality in consumption and derive conditions under which their proxy accurately measure trends in cross-sectional inequality. Our focus is on the covariance of

consumption across generations.

2.6 Selection into Home Ownership in 2012

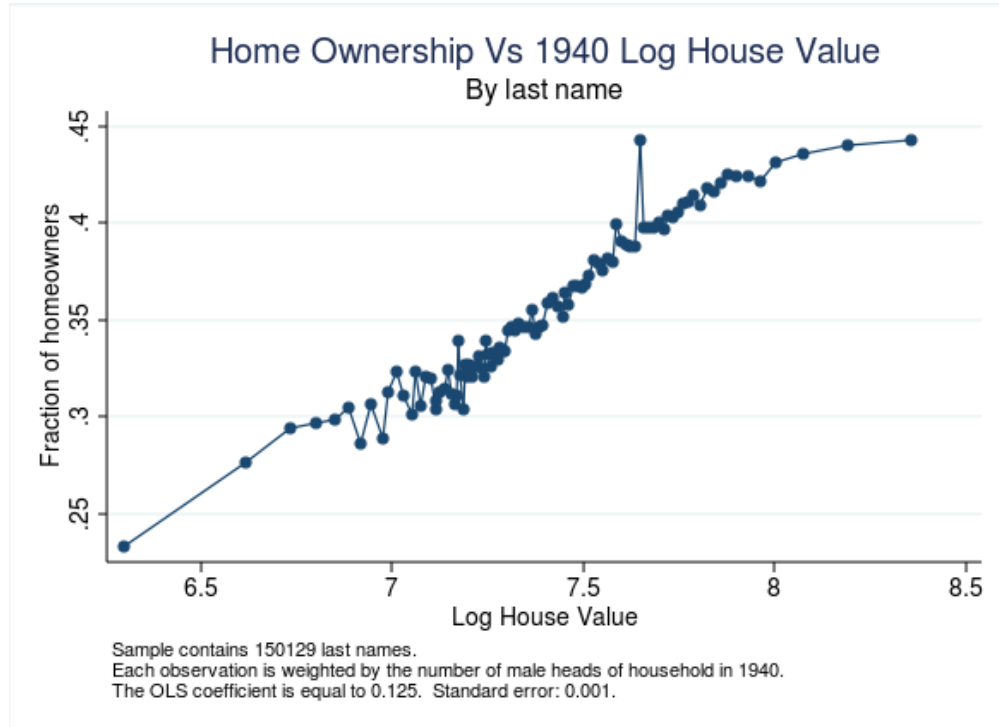


Figure 2.6: Surname level correlation between 1940 housing consumption and fraction of homeowners in 2012.

The discussion of consistency in section 4 - and of the econometric framework - assumes that both samples, in 1940 and 2012, are independently and identically distributed. They are assumed to represent the same population of pairs of ancestor and their descendant. Our 2012 sample however currently includes only homeowners. Renters are omitted.³² As a result, we need to re-write the moment conditions accounting for sample selection. The adjustment can be thought of as tackling two difficulties raised by selection into homeownership. First, the two samples need to represent the same population. The selection adjustment will therefore re-weight the 1940 sample in a way that makes it representative of the ancestors of the 2012

32. As already mentioned, we are in the process of acquiring the InfoGroup data which includes renters. This will solve the selection problem and make this section irrelevant.

population of homeowners. Second, it is not reasonable to assume that the error term in the population represented by our two (re-weighted) samples is mean zero.³³ Figure 2.6, shows that richer last names in 1940 are more likely to be homeowners in 2012. Since Heckman (1979), there are well-known procedures to deal with a selection induced non-zero error term. We now present the moment conditions that form the basis of our estimator.

We start with the selection equation. Let binary variable $B_{i,t+1} = \{0, 1\}$ denote owning a home in 2012. We assume that:

$$B_{i,t+1} = 1 \Leftrightarrow B_{i,t+1}^* = \tilde{X}_{it}\Gamma + \tilde{\eta}_{i,t+1} \geq 0$$

where \tilde{X}_{it} is a vector of variables, in the 1940 sample, that includes at least c_{it} and the constant 1 - both potentially interacted with D_i^g .³⁴ We further assume joint normality of the error terms:

$$\begin{pmatrix} \tilde{\varepsilon}_{i,t+1} \\ \tilde{\eta}_{i,t+1} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\tilde{\varepsilon}}^2 & \sigma_{\tilde{\varepsilon}\tilde{\eta}} \\ \sigma_{\tilde{\varepsilon}\tilde{\eta}} & \sigma_{\tilde{\eta}}^2 \end{pmatrix} \right)$$

Under those assumptions, the exclusion restrictions become:

$$E \left[\left(\tilde{\varepsilon}_{i,t+1} - \sigma_{\tilde{\varepsilon}\tilde{\eta}} \lambda \left(\tilde{X}_i \Gamma \right) \right) z_i^\ell \mid B_{i,t+1} = 1 \right] = 0 \quad \forall \ell$$

where $\lambda \left(\tilde{X}_i \Gamma \right)$ is the inverse Mills ratio. In order to derive the 2SIV estimator, we need

33. Again, the error term here is the prediction error - corresponding to each observation - from the population best linear predictor. In other words, it is $\varepsilon_{i,t+1}$ in equation Eq. (2.5) and $\varepsilon_{i,t+1}$ in equation Eq. (2.1).

34. In the estimation, we also include many other covariates in vector \tilde{X}_i such as education, occupational status, homeownership in 1940,... The inclusion of covariates in the selection equation that are not included in the “second-stage” is unusual. Not including these covariates leads to very little variation in propensity scores and inverse Mills ratio across individuals and even less across last names. The lack of variation in inverse Mills ratio makes identification of $E[\varepsilon_{i,t+1} \mid B_{i,t+1} = 1]$ problematic in the second-stage. As a result, we have decided to include these other covariates. The inclusion of covariates in the selection equation that are not in the “second-stage” is rendered less problematic by the fact that identification of $E[\varepsilon_{i,t+1} \mid B_{i,t+1} = 1]$ relies on the joint normality assumption imposed on the error terms. As a result, variation in the “excluded covariates” only affect estimation through their effect on the inverse Mills ratio.

a sample analog to the above moment conditions. Let us define:

$$m_i(B, \Gamma) = \left(\sum_{g=1}^G (\alpha^g + \beta^g c_{it}) D_i^g + \sigma_{\tilde{\eta}} \lambda (\tilde{X}_i \Gamma) \right) z_i^\ell$$

The moment condition (for each ℓ) can be decomposed into:

$$E \left[c_{i,t+1} z_i^\ell \mid B_{i,t+1} = 1 \right] - E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1 \right] = 0$$

A sample analog to $E \left[c_{i,t+1} z_i^\ell \mid B_{i,t+1} = 1 \right]$ is $\frac{1}{N_2} \sum_{j=1}^{N_2} c_{j,t+1} z_j^\ell$ since the 2012 sample only contains homeowners. In order to obtain a sample analog for $E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1 \right]$, we write:

$$\begin{aligned} E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1 \right] &= \int E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1, \tilde{x}_i \right] dPr(\tilde{x}_i \mid B_{i,t+1} = 1) \\ &= \int E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1, \tilde{x}_i \right] \frac{Pr(B_{i,t+1} = 1 \mid \tilde{x}_i)}{Pr(B_{i,t+1})} dPr(\tilde{x}_i) \end{aligned}$$

We can obtain a consistent estimate of $Pr(B_{i,t+1} = 1 \mid \tilde{x}_i)$ under the exclusion restriction that $E \left[\tilde{\eta}_{i,t+1} z_i^\ell \right] = 0 \forall \ell$. The estimation procedure is detailed in the appendix. The resulting estimate is $\Phi(\tilde{X}_i \hat{\Gamma})$. The sample analog of $E \left[m_i(B, \Gamma) \mid B_{i,t+1} = 1 \right]$ is:

$$\frac{1}{N_1} \sum_{i=1}^{N_1} m_i(\hat{B}, \hat{\Gamma}) \frac{\Phi(\tilde{X}_i \hat{\Gamma})}{\bar{B}}$$

where \bar{B} is an estimate of the average probability that descendants of individuals in 1940 are homeowners in 2012. We obtain this estimate by dividing the number of individuals with a “common last name” in our Dataquick sample in 2012 by the number of individuals with a “common last name” in 2010 according to the Census.³⁵

35. We give more details about this in the Appendix.

Tables B.1 and B.2 in appendix B1 present estimates of β and α after adjusting for selection using the method described in this section and the propensity score estimation explained in appendix B6. The intercepts α are implausibly large which leads to concerns about the correction for selection. The issue may stem - at least partly - from the joint normality assumption as well as from the fact that we do not allow the coefficient on the inverse Mills ratio to vary across demographics. As a result, the “ecological inference problem” may be severe in estimating this coefficient and the error may propagate to the other coefficients. Despite these important reservations regarding the estimated intercepts, the patterns of heterogeneity in β across race and regions are largely preserved. The intergenerational elasticity for whites and the overall population are considerably lower after adjusting for selection. The overall population coefficient of 0.176 translates into a one generation elasticity of 0.42 – 0.5 compared to our previous estimates that range between 0.56 – 0.67. It is worth noting that the “group specific” (by region times race) parameters do not change substantially after adjusting for selection even though the “overall coefficients” (white specific and overall β) decrease markedly. The decrease must therefore come from the “between group” component of the intergenerational elasticity which is especially sensitive to the α parameters. Given the difficulties in estimating the coefficient on the inverse Mills ratio, the potential for error propagation and the strong and untested assumption of joint normality of the error terms, we do not place much weight on these new estimates. Running the regressions in the InfoGroup data - which include renters in 2012 - will settle the issue.

2.7 Conclusion

In this paper, we have estimated the intergenerational correlation of housing consumption across surnames between 1940 and 2012. We map these estimates into an estimated intergenerational elasticity of total consumption at the individual-level. We write down the assumptions under which surname-level correlations correspond to their individual counterpart and attempt to test those assumptions through a battery of placebo tests. From this

exercise, we conclude that it is important to adjust for race when calculating the individual level elasticity using between last name variation. We also estimate heterogeneous elasticities by races and regions. Both the raw surname level estimates and our “adjusted estimates” show similar patterns across races and regions. Black americans have much lower elasticities than whites. Though predicted 2012 consumption at the very bottom of the distribution - for a given level of 1940 consumption - is similar for blacks and whites, the lower elasticity imply that richer blacks have much lower expected 2012 consumption than whites with similar level of 1940 consumption. The Northeast has much larger persistence - controlling for race - than the other three regions including the South. This is a sharp reversal compared to the regional patterns found in the previous literature on intergenerational persistence of earnings. A current shortcoming of this study is that we are not able to separate how much of this racial gap is caused by institutions - like lack of access to credit and asset accumulation for blacks for example - versus heterogeneity in taste and rate of discount. In a later version of this paper, we will make progress on this issue by comparing 1) individuals living in counties that were hit by a large number of bank runs in the 1930s to individuals in counties with fewer bank runs and 2) compare black Americans in places where exclusionary practices of “redlining” were more or less prevalent.³⁶

We present estimates of heterogeneous patterns in consumption insurance across races and regions. We also show that accounting for heterogeneity across observables in our estimation is important to obtain consistent estimates when using surnames to link generations. Some of the placebo tests suggest that there are limitations on the dimensions of heterogeneity that we are able to estimate using this method. An interesting avenue for future research would be to find a way to divide the parameter space across groups in a more “efficient” way by using - for example - machine learning techniques such as trees or random forests in

³⁶. Research suggests that financial shocks such as bank runs have long term effects on the local banking system.

order to define the relevant dimensions of heterogeneity.

REFERENCES

- [1] Athey, S. and G. Imbens (2016). “Recursive partitioning of heterogeneous causal effects.”, *Proceedings of the National Academy of Sciences*, 113(27):7353-7360, 2016.
- [2] Athey, S. and S. Wager (2017), “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests”, *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2017.1319839
- [3] Attanasio, O., Hurst, E. and L. Pistaferri (2015), “The Evolution of Income, Consumption, and Leisure Inequality in the United States, 1980-2010” In *Improving the Measurement of Consumer Expenditures.*”, edited by C.D. Carroll, T.F. Crossley and J. Sabelhaus, *Studies in Income and Wealth*, Vol.74, pp.100-140, Chicago: University of Chicago Press.
- [4] Aughinbaugh (2000). “Reapplication and Extension: Intergeneration Mobility in the United States.”, *Labour Economics*, 7(6):785-96.
- [5] Autor, D.H. (2014), “Skills, education, and the rise of earnings inequality among the “other 99 percent””, *Science*, 23 May 2014: Vol. 344, Issue 6186, pp.843-851
- [6] Bhattacharya, D. and B. Mazumder (2011), “A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States.”, *Quantitative Economics*, 2(3):335-379
- [7] Bayer, P., and K.K. Charles (2018), “Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men since 1940”, *Quarterly Journal of Economics* 133(3): 1459-1501
- [8] Blundell, R., Pistaferri, L. and I. Preston (2008), “Consumption Inequality and Partial Insurance.”, *American Economic Review*, 98 (5):1887-1921

- [9] Blundell, R., Pistaferri, L. and I. Saporta-Eksten (2016), “Consumption Inequality and Family Labor Supply”, *American Economic Review*, 106(2):387-435
- [10] Blundell, Richard and Shephard, Andrew. (2012). “Employment, hours of work and the optimal taxation of low-income families,” *The Review of Economic Studies* 79(2), 481–510.
- [11] Boadway, R. and L. Jacquet 2008, “Optimal marginal and average income taxation under maximin.” *Journal of Economic Theory* 143, 425–441.
- [12] Boadway, Robin, Marchand, Maurice, Pestieau, Pierre and Racionero, Maria del Mar. (2002). “Optimal redistribution with heterogenous preferences for leisure,” *Journal of Public Economic Theory*, 4(4) 475–498.
- [13] Brett, Craig, and John A. Weymark. 2003. “Financing Education Using Optimal Redistributive Taxation.” *Journal of Public Economics*, 87: 2549–2569.
- [14] Charles, K.K., and E. Hurst (2003), “The Correlation of Wealth across Generations.”, *Journal of Political Economy*, 111(6)155-82.
- [15] Charles, K.K., Danziger, S., Li, G., and R. Schoeni (2014), “The Intergenerational Correlation of Consumption Expenditures”, *American Economic Review: Papers & Proceedings*, 104(5):136-140.
- [16] Chetty, Raj. 2012. “Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, *Econometrica*, *Econometric Society*, vol. 80(3), pages 969-1018.
- [17] Chetty, R., Hendren, N., Kline, P., and E. Saez (2014), “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”, *Quarterly Journal of Economics* 129(4):1553-1623

- [18] Chetty, R., Hendren, N., Jones M.R. and S.R. Porter (2018), “Race and Economic Opportunity in the United States: An Intergenerational Perspective.”, Working Paper (March version)
- [19] Choné, Philippe, and Guy Laroque. 2010. “Negative marginal tax rates and heterogeneity.” *American Economic Review* 100(5): 2532–2547.
- [20] Clark, Gregory (2014), ”The Son Also Rises: Surnames and the History of Social Mobility”. Princeton: Princeton University Press.
- [21] Clarke G. and N. Cummins (2015), ”Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility”, *The Economic Journal* 125(February), 61-85.
- [22] Collado, M.D., Ortuno Ortìn, I. and A. Romeu (2012). “Long-run Intergenerational Social Mobility and the Distribution of Surnames”, manuscript, Universidad de Alicante
- [23] DeNavas-Walt, Carmen and Bernadette D. Proctor, U.S. Census Bureau, Current Population Reports, P60-252, *Income and Poverty in the United States: 2014*, U.S. Government Printing Office, Washington, DC, 2015.
- [24] Diamond, Peter. (1980). “Income taxation with fixed hours of work,” *Journal of Public Economics* 13(1), 101–110.
- [25] Diamond A. 1998, “Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. ” *The American Economic Review*, 88(1):83-95.
- [26] Feenberg Daniel and Elisabeth Coutts 1993, “An Introduction to the TAXSIM Model”, *Journal of Policy Analysis and Management*, Vol. 12, No.1, pp.189–194
- [27] Fleurbaey M. 2008, *Fairness, responsibility, and welfare*, Oxford: Oxford University Press.

- [28] Fleurbaey M. and F. Maniquet 2005, “Fair social orderings with unequal production skills”, *Social Choice and Welfare* 24: 1–35.
- [29] Fleurbaey, Marc, and François Maniquet. 2006. “Fair Income Tax.” *Review of Economic Studies*, 73: 55–83.
- [30] Fleurbaey, Marc, and François Maniquet. 2007. “Help the Low-Skilled, or Let the Hard-working Thrive? A Study of Fairness in Optimal Income Taxation.” *Journal of Public Economic Theory*, 9: 467–500.
- [31] Fleurbaey M. and F. Maniquet 2011, *A Theory of Fairness and Social Welfare*, Cambridge: Cambridge University Press.
- [32] Fleurbaey M. and F. Maniquet (2018), Optimal income taxation theory and principles of fairness, *Journal of Economic Literature*, forthcoming.
- [33] Fleurbaey M. and A. Trannoy 2003, “The impossibility of a Paretian Egalitarian”, *Social Choice and Welfare* 21: 243–464.
- [34] Friedman, Milton (1957). “The Permanent Income Hypothesis” in “A Theory of the Consumption Function”, Chapter 3, *Princeton University Press*. ISBN 0-691-04182-2
- [35] Guell, M., Rodriguez Mora J.V. and C.I. Telmer (2015), “The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating”, *Review of Economic Studies* 82:693-735
- [36] Guesnerie, R. and J.-J. 1984, “A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm.,” *Journal of Public Economics* 25, 329-369.
- [37] Hajargasht Gholamreza and William E. Griffiths 2013, “Pareto-lognormal distributions: Inequality, poverty and estimation from grouped income data,” *Economic Modelling* 33, 593–604.

- [38] Heckman, J. (1979), “Sample Selection Bias as a Specification Error”, *Econometrica*, Vol. 47, No. 1 (January, 1979), 153-161.
- [39] Heckman, James and Richard Robb. (1985). “Alternative methods for evaluating the impact of interventions: An overview”, *Journal of Econometrics* 30, 239–267.
- [40] Inoue, A. and G. Solon (2010). “Two-sample Instrumental Variables Estimators”, *Review of Economics and Statistics*, 92(3):557-561
- [41] IPUMS-CPS, University of Minnesota, www.ipums.org
- [42] Kanbur, R. M. Keen and M. Tuomala 1994a, “Optimal non-linear income taxation for the alleviation of income-poverty,” *European Economic Review* 38: 1613-1632.
- [43] Kanbur, R. M. Keen and M. Tuomala 1994b, “Labor supply and targeting in poverty alleviation programs,” *The World Bank Economic Review* 8, 191–211.
- [44] King, G. (1997). “A Solution to the Ecological Inference Problem.” Princeton, NJ: Princeton University Press.
- [45] King, G., Rosen, O. and M. Tanner (eds.) (2004). *Ecological Inference: New Methodological Strategies*. Cambridge, UK: Cambridge University Press.
- [46] Klevmarcken, W.A. (1982). “Missing variables and two-stage least-squares estimation from more than one data set”. In: 1981 Proceedings of the American Statistical Association, Business and Economic Statistics Section. pp. 156-161.
- [47] Kolesàr, M., Chetty, R., Friedman, J., Glaeser, E. and G.W. Imbens (2015). “Identification and Inference with Many Invalid Instruments.”, *Journal of Business & Economics Statistics* 33(4):474-484.
- [48] Lockwood, Benjamin B. 2016, “Optimal income taxation with present bias.”, Mimeo

- [49] Lockwood, Benjamin B., and Matthew Weinzierl. 2015. “De Gustibus Non Est Taxandum: Heterogeneity in Preferences and Optimal Redistribution.” *Journal of Public Economics*, 124: 74–80.
- [50] Maniquet, Francois and Neumann, Dirk. (2016). “Well-Being, Poverty and Labor Income Taxation: Theory and Application to Europe and the U.S.” CORE Discussion Paper 2016/29.
- [51] Mazumder, B. (2014). “Black-White Differences in Intergenerational Economic Mobility in the United States.”, *Economic Perspectives* XXXVIII(1).
- [52] Mazumder, B. (2015). “Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data”, Federal Reserve Bank of Chicago, Working paper (September 2015 Version).
- [53] Mazumder, B. and J. Davis (2018). “Racial and Ethnic Differences in the Geography of Intergenerational Mobility”. SSRN.
- [54] Meyer Bruce D., Mok Wallace K.C. and James X. Sullivan 2015, ”The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences,” Mimeo.
- [55] Meyer Bruce D., Mok Wallace K.C. and James X. Sullivan 2015, ”Household Surveys in Crisis,” *Journal of Economics Perspectives*, Fall 2015, 29(4), pp.199-226.
- [56] Meyer, B. and J.X. Sullivan (2017). “Consumption and Income Inequality in the US since the 1960s”, *NBER Working Paper*, No.23655
- [57] Mirrlees, James A. 1971. “An exploration in the theory of optimum income taxation.” *The Review of Economic Studies* 38(2) 175–208.
- [58] Mirrlees, James A. 1976. “Optimal Tax Theory: A Synthesis.” *Journal of Public Economics*, 6: 327–358.

- [59] Robert Moffitt. (1993). “Identification and estimation of dynamic models with a time series of repeated cross-sections”, *Journal of Econometrics* 59, 99-123.
- [60] Mulligan, C. (1997), *Parental Priorities and Economic Inequality*.. Chicago: University of Chicago Press.
- [61] Murphy, K.M. and R.H.Topel (2016), “Human Capital Investment, Inequality and Economic Growth”, *NBER Working Paper*, No.21841
- [62] Olivetti, C. and M. Daniele Paserman (2015), “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940”, *American Economic Review*, 105(8):2695-2724
- [63] Piketty, T. and E. Saez (2003), “Income Inequality in the United States, 1913-1998”, *Quarterly Journal of Economics*, 118(1):1-39
- [64] Piketty, T., Saez, E. and G. Zucman (2018), “Distributional National Accounts: Methods and Estimates for the United States”, *Quarterly Journal of Economics* 133(2):553-609
- [65] Reed, William J. 2003, “The Pareto law of incomes - an explanation and an extension,” *Physica A* 319, 469–486
- [66] Ridder, G. and R. Moffitt (2007), “The Econometrics of Data Combination”, In *Handbook of Econometrics.*, Vol. 6, JJ. Heckman & E.E. Leamer (eds), chapter 75. Amsterdam: Elsevier Science, North-Holland.
- [67] Roemer, John E. 1998, *Equality of Opportunity*, Cambridge: Harvard University Press.
- [68] Saez, Emmanuel. (2002). Optimal income transfer programs: intensive versus extensive labor supply responses?, *The Quarterly Journal of Economics* 117(3), 1039–1073.
- [69] Saez, E. and S. Stantcheva 2016, “Generalized social marginal welfare weights for optimal tax theory,” *American Economic Review* 106, 24–45.

- [70] Saez, E. and G. Zucman (2016), “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”, *Quarterly Journal of Economics* 131(2):519–578
- [71] Jung Sakong (2018), “Cyclical housing transactions and wealth inequality”, Working Paper, October 2018.
- [72] Samuelson, Paul A. 1974 “Complementarity – An Essay on the 40th anniversary of the Hicks-Allen revolution in demand theory,” *Journal of Economic Literature* 12, 1255–1289.
- [73] Solon, G. (1999), “Intergenerational Mobility in the Labor Market.”, In *Handbook of Labor Economics.*, Vol. 3A, edited by Orley Ashenfelter and David Card, 1761-1800. Amsterdam: Elsevier Science, North-Holland.
- [74] Skinner, J. (1987), “A Superior Measure of Consumption from the Panel Study of Income Dynamics.”, *Economics Letters* 23(2):213-6.
- [75] Spenkuch, J.L. (2018), “Causal Ecological Inferences”, Working Paper, February 2018.
- [76] Waldkirch, A., Ng, S., and D. Cox (2004), “Intergenerational Linkages in Consumption Behavior.”, *Journal of Human Resources*, 39(2):355-81.
- [77] Wane, W. 2001, “The optimal income tax when poverty is a public bad,” *Journal of Public Economics* 82: 271-299.
- [78] Weinzierl, Matthew. 2014. “The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice.” *Journal of Public Economics*, 118: 128–142.

APPENDIX A

APPENDIX TO CHAPTER 1

A.1 Proofs

A.1.1 Proof of Proposition 2

We need to introduce the following terminology, which corresponds to the implicit budget terminology of Samuelson (1974). The equivalent income is the intercept of an implicit budget. Formally, for (z, w, R) , we let $m(z, w, R)$ denote the intercept of $IB(z, w, R)$, that is,

$$m(z, w, R_i) = \min \{t \in \mathbb{R} \mid \exists (\ell, c) \in X, c = t + w\ell, (\ell, c) R z\}.$$

With the same abuse of notation as above, we define

$$m(w, PL) = \min \{t \in \mathbb{R} \mid \exists (\ell, c) \in PL, c = t + w\ell\}.$$

Proof: We prove the lemma in three steps. The first step is a preliminary result about the existence of specific preference relations. Take any two $w, w' \in [\underline{w}, \infty)$, and any $m, m', m'', m''' \in \mathbb{R}$ such that

- $m < m' < m(w, PL)$,
- $m(w', PL) < m'' < m'''$.

There necessarily exists $R^\circ \in \mathcal{R}$ and $z = (\ell, c), z' = (\ell', c'), z'' = (\ell'', c''), z''' = (\ell''', c''') \in X$ such that

- $\ell = \ell' = \ell'' = \ell'''$,
- $c + c''' = c' + c''$,
- $m(z, w, R^\circ) = m$,

- $m(z', w', R^\circ) = m'$,
- $m(z'', w', R^\circ) = m''$,
- $m(z''', w', R^\circ) = m'''$.

Instead of proving the claim algebraically, we illustrate its simple intuition in the following figure. The distance between the two budgets of slope \tilde{w}' is much larger than that between the budgets of slope \tilde{w} . in spite of that, it is possible to construct indifference curves that are tangent to the former budgets and yet become arbitrarily close to each other.

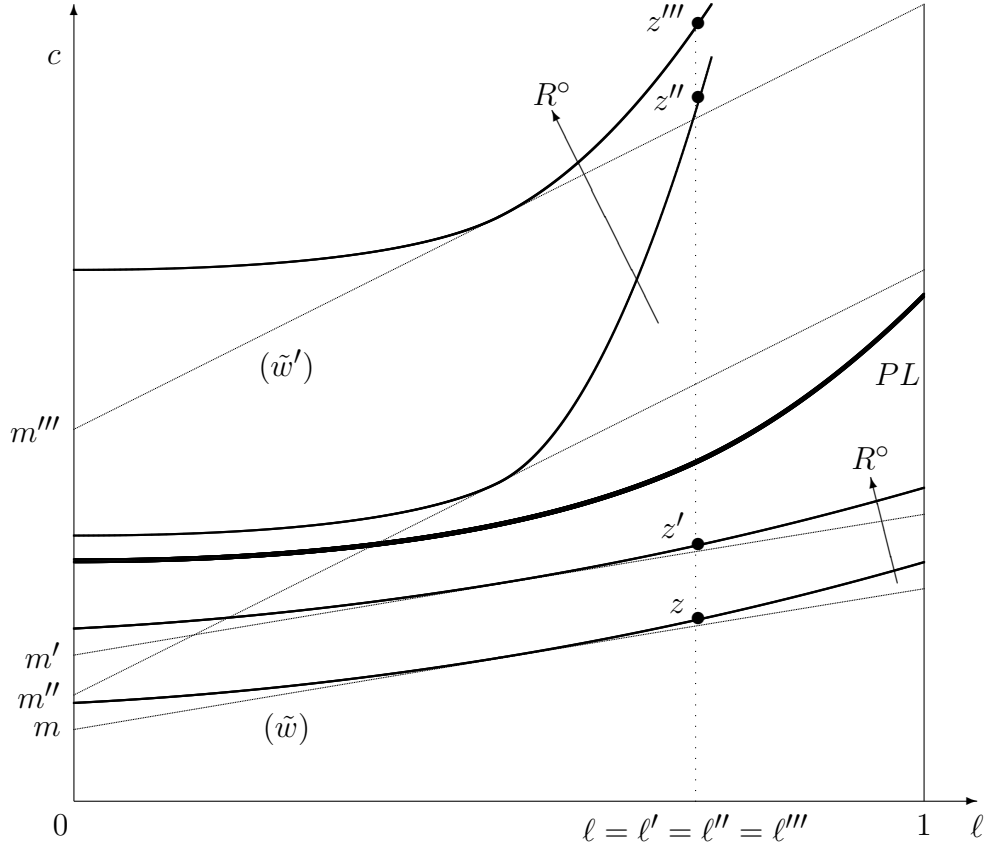


Figure A.1: Illustration of the existence of R° .

We now turn to the second step. We prove the lemma in a specific case. Let $E = (w_N, R_N) \in \mathcal{E}$, $z_N, z'_N \in X^N$, and $j, k \in N$ satisfy the conditions of the statement. Assume,

moreover, that there exist $m_j > m(w_j, PL)$ and $m_k < m(w_k, PL)$ such that

$$\begin{aligned} m(z'_j, w_j, R_j) - m_j &= m(z_j, w_j, R_j) - m(z'_j, w_j, R_j), \\ m_k - m(z'_k, w_k, R_k) &= m(z'_k, w_k, R_k) - m(z_k, w_k, R_k). \end{aligned}$$

As $|N| \geq 4$, we assume $|N| = 4$ to save on notation. Let $N = \{j, k, a, b\}$. We need to prove that $(z'_j, z'_k, z_a, z_b) P(E) (z_j, z_k, z_a, z_b)$. By *Pareto indifference*, we can do as if z_j and z'_j (resp. z_k and z'_k) were the best bundles of j (resp. k) in the corresponding implicit budget. If it is not the case, indeed, then we can identify those best bundles and replace z_j, z'_j, z_k and z'_k by those bundles. Let $R^\circ \in \mathcal{R}$ and $z = (\ell, c), z' = (\ell', c'), z'' = (\ell'', c''), z''' = (\ell''', c''') \in X$ be such that

- $\ell = \ell' = \ell'' = \ell'''$,
- $c + c''' = c' + c''$,
- $m(z, w_k, R^\circ) = m(z'_k, w_k, R_k)$,
- $m(z', w_k, R^\circ) = m_k$,
- $m(z'', w_j, R^\circ) = m_j$,
- $m(z''', w_j, R^\circ) = m(z'_j, w_j, R_j)$.

Let $E' \in \mathcal{E}$ be defined by

$$E' = ((w_j, w_k, w_j, w_k), (R_j, R_k, R^\circ, R^\circ)).$$

By *equal-skill transfer* applied twice, first between agents j and a and then between agents k and b ,

$$(z'_j, z'_k, z''', z) P(E') (z_j, z_k, z'', z').$$

By *poverty-reduction* applied between agents a and b ,

$$(z'_j, z'_k, z'', z') P(E') (z'_j, z'_k, z''', z).$$

By transitivity,

$$(z'_j, z'_k, z'', z') P(E') (z_j, z_k, z'', z').$$

By *separability*,

$$(z'_j, z'_k, z_a, z_b) P(E) (z_j, z_k, z_a, z_b),$$

the desired outcome. Of course, it is unlikely that such m_j and m_k exist, as $m(z'_j, w_j, R_j)$ can be arbitrarily close to $m(w_j, PL)$ and $m(z'_k, w_k, R_k)$ arbitrarily close to $m(w_k, PL)$.

As a third and final step, we deal with the general case. We can find $m_j > m(w_j, PL)$ and $m_k < m(w_k, PL)$ and an integer n such that

$$\begin{aligned} m(z'_j, w_j, R_j) - m_j &= \frac{1}{n} \left(m(z_j, w_j, R_j) - m(z'_j, w_j, R_j) \right), \\ m_k - m(z'_k, w_k, R_k) &= \frac{1}{n} \left(m(z'_k, w_k, R_k) - m(z_k, w_k, R_k) \right). \end{aligned}$$

Then, we can define two series of $n - 1$ bundles z_j^1, \dots, z_j^{n-1} and z_k^1, \dots, z_k^{n-1} such that for all $p \in \{1, \dots, n - 1\}$,

$$m(z_j^p, w_j, R_j) = m(z_j, w_j, R_j) - p \left(m(z'_j, w_j, R_j) - m_j \right), \quad (\text{A.1})$$

$$m(z_k^p, w_k, R_k) = m(z_k, w_k, R_k) + p \left(m_k - m(z'_k, w_k, R_k) \right). \quad (\text{A.2})$$

For all $p \in \{1, \dots, n\}$, we can apply step 2 and prove that

$$(z_j^p, z_k^p, z_a, z_b) P(E) (z_j^{p-1}, z_k^{p-1}, z_a, z_b),$$

with the obvious convention that $z_j^0 = z_j, z_j^n = z'_j$, and $z_k^0 = z_k, z_k^n = z'_k$. The proof then

goes on by transitivity applied to those n steps, to reach the final outcome that

$$(z'_j, z'_k, z_a, z_b) P(E) (z_j, z_k, z_a, z_b).$$

□

A.1.2 Proof of Corollary 1

Proof. Let $E = (w_N, R_N) \in \mathcal{E}$, $z_N, z'_N \in X^N$, and $j \in N$ satisfy the conditions of the corollary. Let us assume, contrary to the claim, that $z_N R(E) z'_N$. Let $M \subset N$ be defined by

$$M = \{i \in N \mid i \neq j, z_i P_i z'_i\}.$$

If $M = \emptyset$, then, by *strong Pareto*, we have reached a contradiction. Let us then assume that $M \neq \emptyset$. Let $z''_N \in X^N$ be such that

$$\begin{aligned} \forall i \in M, \quad & IB(w_i, z_i, R_i), IB(w_i, z'_i, R_i) \supset IB(w_i, z''_i, R_i) \supset IB(w_i, PL), \\ \forall i \notin M, i \neq j \quad & z''_i = z_i, \\ & = IB(w_j, PL) \supset IB(w_j, z''_j, R_j) \supset IB(w_j, z_j, R_j). \end{aligned}$$

Existence of such a z''_N is guaranteed by the conditions imposed on z_N and z'_N in the statement of the corollary.

Let $r > 0$ and $z_j^0, z_j^1, \dots, z_j^m \in X$, $m = |M|$ be such that

$$\begin{aligned} IB(w_j, z_j^0, R_j) &= IB(w_j, z_j, R_j), \\ IB(w_j, z_j^k, R_j) &= IB(w_j, z_j^{k-1}, R_j) + (0, r), \forall k \in \{1, \dots, m\}, \\ IB(w_j, z_j^m, R_j) &= IB(w_j, z''_j, R_j). \end{aligned}$$

By applying Proposition 2 m times, we can prove that $z''_N P(E) z_N$. Indeed, at each step,

z_j^{k-1} is replaced by z_j^k , which means that an agent whose implicit budget is strictly below the poverty line is made better-off, and z_i is replaced with z_i'' , which means that an agent strictly above the poverty line is made worse-off, which implies a strict social preference. By transitivity, $z_N'' P(E) z_N'$, which contradicts *strong Pareto*, because $z_i' R_i z_i''$ for all $i \in N$ and the preference is strict for some agents, including j . \square

A.1.3 Proof of Proposition 3

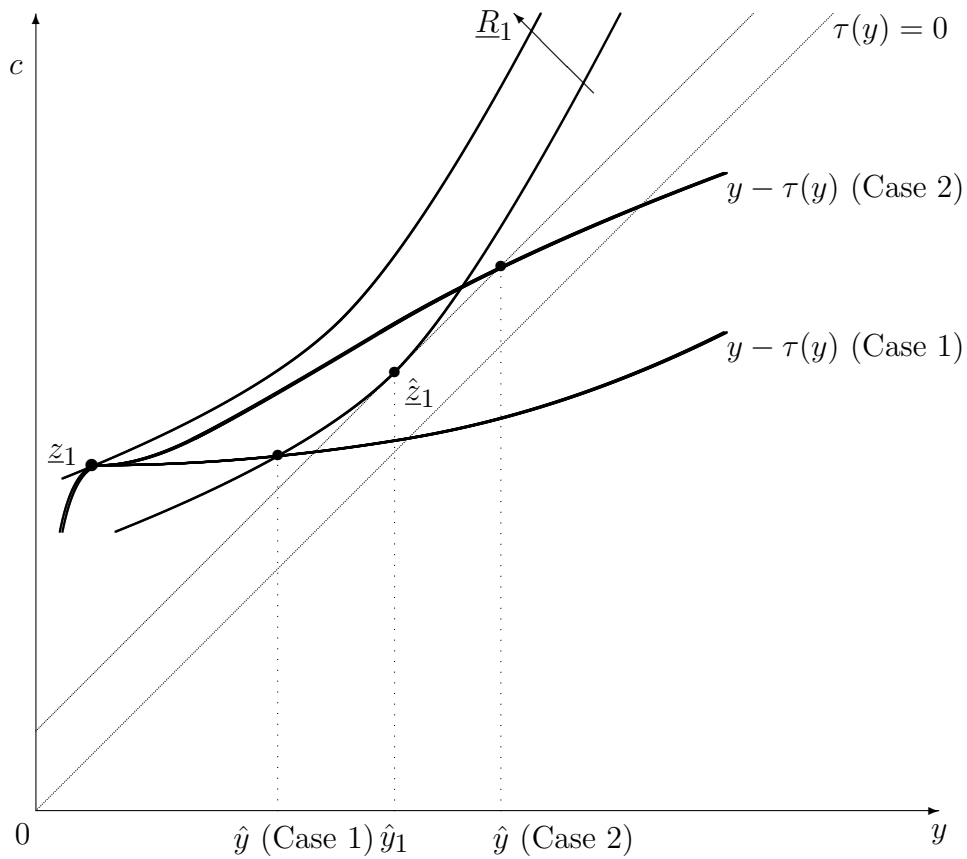


Figure A.2: Illustration of the proof of Proposition 3.

Proof. Fig. A.2 illustrates the reasoning. Let $E = (w_N, R_N)$. Let τ be optimal in E for $R^{\tilde{R}lex}$. Let \underline{z}_N be generated by τ . Let z_N corresponds to \underline{z}_N in the (ℓ, c) space. Let us

assume, contrary to the claim, that

$$b(z_1, w_1, R_1) > b(z_j, w_j, R_j) = b$$

for some $j \in N$. Let $\hat{z}_1 = (\hat{y}_1, \hat{c}_1) \in \underline{X}$ be defined by:

$$b\left(\left(\frac{\hat{y}_1}{w_1}, \hat{c}_1\right), w_1, R_1\right) = b$$

and

$$\hat{z}_1 \in \underline{IB}(\hat{z}_1, R_1).$$

Note that by assumption, $z_1 \underline{P}_1 \hat{z}_1$. Later on in the proof, we will use the following fact:

$$\forall y \leq \hat{y}_1, \tau(y) \leq \hat{y}_1 - \hat{c}_1. \quad (\text{A.3})$$

Let us prove this fact. Assume it is not true. Then there exists $y^* < \hat{y}_1$, such that $\tau(y^*) = \max_{y < \hat{y}_1} \tau(y)$ and $\tau(y^*) > \tau(\hat{y}_1)$. By Assumption 1, there exists $i \in N$ such that $\underline{z}_i \underline{I}_i(y^*, y^* - \tau(y^*))$. Note that among all the feasible bundles that leave i indifferent to $(y^*, y^* - \tau(y^*))$, this bundle itself maximizes the paid tax (or minimizes the received transfer), so that, if τ is supposed to be optimal, it is i 's bundle: $\underline{z}_i = (y^*, y^* - \tau(y^*))$. Indeed, if it were not the case, by replacing \underline{z}_i with $(y^*, y^* - \tau(y^*))$, the planner would not change anything in the well-being of the agents, all the incentive constraints would remain satisfied and there would be a budget surplus, which could be redistributed. As a consequence, by Assumption 3, there exists $j \in N$ such that $w_j = w_1$ and $\underline{z}_j = (y^*, y^* - \tau(y^*))$. As a result,

$$\underline{IB}(\underline{z}_j, \underline{R}_j) = \{z = (y, c) \in \underline{X} \mid c = y - \tau(y^*)\}.$$

That implies

$$\underline{IB}(\underline{z}_j, \underline{R}_j) \subset \underline{IB}(\hat{z}_1, \underline{R}_1),$$

with the consequence that

$$b\left(\left(\frac{y^*}{w_j}, y^* - \tau(y^*)\right), w_j, R_j\right) < b,$$

the desired contradiction, proving the fact stated above.

We now define an income threshold \hat{y} by distinguishing two cases.

1. $\hat{c}_1 > \hat{y}_1 - \tau(\hat{y}_1)$: \hat{y} is defined by

$$(\hat{y}, \hat{y} - \tau(\hat{y})) \underline{I}_1 \hat{z}_1,$$

and in case several such \hat{y} exist, we take the largest;

2. $\hat{c}_1 > \hat{y}_1 - \tau(\hat{y}_1)$: \hat{y} is defined by

$$\tau(\hat{y}) = \hat{y}_1 - \hat{c}_1,$$

and in case several such \hat{y} exist, we take the largest.

Let τ' be defined by: for all $y \geq 0$

$$\tau'(y) = \max\{\tau(y), \tau(\hat{y})\}. \tag{A.4}$$

and let \underline{z}_N be the allocation generated by τ .

Claim 1: τ' collects a budget surplus: $\sum_{i \in N} \tau'(y'_i) > 0$. By construction, $\underline{B}(\tau') \subset \underline{B}(\tau)$.

Therefore, if $\tau'(y_i) = \tau(y_i)$, that is, τ' and τ coincide at y_i , then $\underline{z}'_i = \underline{z}_i$, that is agent i does not change her choice of bundle. If $\tau'(y_i) \neq \tau(y_i)$, because $\tau(y_i) < \tau(\hat{y})$, then $\underline{z}'_i \neq \underline{z}_i$ but by construction of τ' , $\tau'(y'_i) \geq \tau(y_i)$. In any case, $\tau'(y'_1) > \tau(y_1)$, which proves the claim.

Claim 2: $\min_{i \in N} b(z'_i, w_i, R_i) = b$. Let $k \in N$ be such that

$$b(z'_k, w_k, R_k) = \min_{i \in N} b(z'_i, w_i, R_i).$$

We consider two cases in turn. Case 1: $w_k = w_1$. Because $R_k \leq R_1$, by Assumption 2, $y'_k \leq y'_1$. Gathering Eqs. A.4, and the fact that $\tau(\hat{y}) \leq \hat{y}_1 - \hat{c}_1$, we obtain

$$\forall y \leq \hat{y}_1, \tau'(y) \leq \hat{y}_1 - \hat{c}_1.$$

This implies

$$\underline{IB}(z_k, R_k) \subseteq \underline{IB}(\hat{z}_1, R_1),$$

with the immediate consequence that $b(z'_k, w_k, R_k) \geq b$, the desired outcome. Case 2: $w_k > w_1$. If $y'_k \geq \hat{y}$, then $y'_k = y_k$, $z'_k = z_k$, nothing changes in the well-being levels and the claim is proven. If $y'_k < \hat{y}$, then

$$\underline{IB}(z_k, R_k) \supseteq \underline{IB}(\hat{z}_1, R_1),$$

so that, because $w_k > w_1$,

$$IB(z'_k, R_k, w_k) \supset IB(z'_1, R_1, w_1),$$

so that

$$b(z'_k, R_k, w_k) \geq b(z'_1, R_1, w_1),$$

the desired outcome that completes the proof of claim 2.

Gathering both claims, τ' does not decrease the minimal well-being level whereas it collects a budget surplus. Redistributing that surplus among all agents would strictly increase the minimal well-being level, contradicting the assumption that τ is optimal. \square

A.1.4 Proof of Corollary 2

Proof. Let $E = (w_N, R_N)$. Let τ be optimal for $R^{\tilde{R}lex}$. Let \underline{z}_N be generated by τ with $\underline{z}_1 = (y_1, c_1)$. A) By Proposition 3, $b(z_1, w_1, R_1) \leq b(z_i, w_i, R_i)$, for all $i \in N$. Let $\underline{IB}(\underline{z}_1, R_1) = \{(y, c) \in \underline{X} \mid c \leq y - t\}$. By replicating the proof of the fact appearing at the beginning of

the proof of Proposition 3, we show that $\tau(y) \leq t$ for all $y \leq y_1$. Assume that, contrary to the claim of the Corollary, $\tau(y_1) > \tau(y^*)$ for some $y^* \leq y_1$. Then, we can replicate the proof of Claim 2 of the proof of Proposition 3, define τ' as $\tau'(y) = \max\{\tau(y), \tau(y^*)\}$, and show that τ' is better than τ according to $R^{\tilde{R}lex}$, the desired contradiction. B) Let $j \in N$ be such that $w_j = w_1$, $\underline{IB}(z_1, \underline{R}_1) = \underline{IB}(z_j, \underline{R}_j)$ and $\underline{R}_j < \underline{R}_1$. Let $z_j^* = (y_j^*, c_j^*) \in \underline{X}$ be such that $z_j^* \underline{I}_j z_j$ and $z_j^* \in \underline{IB}(z_j, \underline{R}_j)$. Note, in particular, that $\tau(y_j^*) = t$. We first claim that there is no loss of generality in assuming that $z_j^* = z_j$. Assume, on the contrary, that $z_j^* \neq z_j$. By construction, $\tau(y_j) < \tau(y_j^*)$. Therefore, z'_N , defined by $z'_i = z_i$ for all $i \in N \setminus \{j\}$ and $z'_j = z_j^*$ is budget balanced. Moreover, it is incentive compatible. Indeed, assume not. There exists $k \in N$ such that $z_j^* \underline{P}_k z_k$. By Assumption 3, there exists $\ell \in N$ such that $z_j^* \underline{P}_\ell z_\ell$ and $w_\ell = w_j = w_1$. Consequently,

$$\underline{IB}(z_\ell, \underline{R}_\ell) \subset \underline{IB}(z_j^*, \underline{R}_j) = \underline{IB}(z_j, \underline{R}_j),$$

so that

$$b(z_\ell, w_\ell, \underline{R}_\ell) < b(z_1, w_1, \underline{R}_1),$$

contradicting Proposition 3. This proves the claim. Now, it follows that for all $y \leq y_j$,

$$\tau(y) = t.$$

Indeed, if it were not the case, then, as in the proof of part A) above, it would be possible to define τ' as follows: for all $y \geq y_j$, $\tau'(y) = \tau(y)$ and for all $y \leq y_j$, $\tau'(y) = \tau(y_j)$, which leads to the same social welfare but collects a budget surplus, contradicting the optimality of τ . □

A.1.5 Derivation of Optimal Income Tax Formulas under Additive

Separability and proof of Proposition 4

Proof. Both $V^L(y)$ $V^U(y)$ can be solved independently for a given y_1 using Hamiltonians. By definition of $y^*(n_1)$ being the optimal income of agents of type n_1 , it is a solution to the following maximization problem:

$$V = \max_y \left\{ V^L(y) + V^U(y) \right\}$$

The first-order conditions allow to solve for $y^*(n_1)$:

$$V_y^L(y^*(n_1)) + V_y^U(y^*(n_1)) = 0 \tag{A.5}$$

where the derivatives of $V^L(y)$ and $V^U(y)$ can be found using the envelope theorem.

We define the Hamiltonian for the upper subproblem in the main text. The Hamiltonian for the lower subproblem is defined similarly and left for the appendix.

$$\begin{aligned} \mathcal{H}(u, y, n, \lambda, \mu, \nu^U) \equiv & \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) g(n) \\ & + \lambda(n) h' \left(\frac{y(n)}{n} \right) \left(\frac{y(n)}{n^2} \right) \\ & + \mu(n) \left[u(n) - \underline{u(n)} \right] \\ & + \nu^U(n) [y(n) - y_1] \end{aligned}$$

The second line is the law of motion of utility given by the first-order conditions of the incentive compatibility constraints. The corresponding co-state variable is $\lambda(n)$. The third line is the lower-bound on utilities and $\mu(n)$ is the lagrange multiplier associated with this constraint. The fourth line is the constraint that all incomes be greater or equal to some income y_1 (which is equal to $y^*(n_1)$ at the solution).¹ Writing the first-order conditions for

1. Notice that we ignore the constraint that $y(n)$ be increasing for all n other than n_1 . In our calibrations,

the Hamiltonians corresponding to each subproblem and rearranging, we obtain the following formula for the optimal marginal tax rate:

$$\frac{T'(y(n))}{1 - T'(y(n))} = A(n)D(n)C(n) \quad \forall n \in [\underline{n}, \underline{n}_b] \quad (\text{A.6})$$

$$\frac{T'(y(n))}{1 - T'(y(n))} = A(n)B(n)C(n) \quad \forall n \in [\underline{n}_b, \infty) \quad (\text{A.7})$$

with

$$A(n) = \left[\frac{1 + \epsilon^u}{\epsilon^c} \right] \quad (\text{A.8})$$

$$B(n) = v'(c(n)) \int_n^\infty \left[\frac{g(t)}{v'(c(t))} - \mu(t) \right] dt \quad (\text{A.9})$$

$$C(n) = \frac{1}{ng(n)} \quad (\text{A.10})$$

$$D(n) = -v'(c(n)) \int_{\underline{n}}^n \left[\frac{g(t)}{v'(c(t))} - \mu(t) \right] dt \quad (\text{A.11})$$

where ϵ^c and ϵ^u , the compensated and uncompensated elasticity of labor earnings, are defined by

$$\begin{aligned} \epsilon^c &= \frac{1 - T'(y)}{y} \frac{\partial y}{\partial(1 - T'(y))} \Big|_{\bar{u}} = \frac{h'(\frac{y}{n}) \left(\frac{1}{n}\right)^2}{y \left[h''(\frac{y}{n}) \left(\frac{1}{n}\right) - (1 - T'(y))^2 v''(c) \right]} \\ \epsilon^u &= \frac{1 - T'(y)}{y} \frac{\partial y}{\partial(1 - T'(y))} = \frac{h'(\frac{y}{n}) \left(\frac{1}{n}\right) + (1 - T'(y))^2 y v''(c)}{y \left[h''(\frac{y}{n}) \left(\frac{1}{n}\right)^2 - (1 - T'(y))^2 v''(c) \right]} \end{aligned}$$

Notice that ϵ^c and ϵ^u are, in general, functions of the marginal tax rate $T'(y)$ and the individual's type n .

The terms $A(n)$ and $C(n)$ are standard. The interesting terms are $B(n)$ and $D(n)$,

this constraint is only binding in n_1 . But in case it was binding at some n for some distribution of types, then the solution would consist in "ironing" the income schedule. See Guesnerie and Laffont (1984).

describing how the social planner's preferences influence the optimal tax. The key normative determinants of the tax scheme are the interval of earnings over which the constraint $u(n) \geq \underline{u}(n)$ is binding.

The third claim made in Proposition 4 is rather obvious. By incentive compatibility, $y^*(n)$ is increasing. Therefore, the constraint that $y^*(n) \leq y_1 = y^*(n_1)$ can only bind (if at all) on an interval $[n_{\underline{b}}, n_1]$ for some $n_{\underline{b}}$. If the constraint is binding on this interval, then $y^*(n) = y^*(n_1) \forall n \in [n_{\underline{b}}, n_1]$. Notice that we do not rule out the case where $n_{\underline{b}} = n_1$.

We will prove the rest of Proposition 4 in three steps. The first step is intuitive and is described in the text above. Formally, we want to show that if $\exists a, b \in [\underline{n}, n_1]$, $a < b$, such that $\mu(n) > 0 \forall n \in [a, b]$, then

$$T(y(n)) = \bar{T}, \quad \forall n \in [a, b]$$

where \bar{T} is such that $\max_{y \in [0, \infty)} (v(y - \bar{T}) - h(\frac{y}{b})) = \underline{u}(b)$. The statement also implies that $T'(y(n)) = T''(y(n)) = 0 \forall n \in [a, b]$.

Clearly, letting agents with $n \in [a, b]$ choose over a segment where $T(y(n)) = \bar{T}$ equalizes the implicit budget of all individuals with $n \in [a, b]$ and $w_i = \underline{w}$. As a result, all individuals in this interval, with $w_i = \underline{w}$, are against the lower-bound on their utilities: $u(n) = \underline{u}(n) \forall n \in [a, b]$. Also, for any other $(y'(n), c'(n))$ such that $u'(n) \geq \underline{u}(n)$, we would have

$$\int_a^b (y'(n) - c'(n))g(n)dn \leq \int_a^b (y(n) - c(n))g(n)dn = \bar{T}(G(b) - G(a)).$$

The above reasoning implies that $\mu(n) > 0 \forall n \in [a, b] \Rightarrow T'(y(n)) = 0$ and $T''(y(n)) = 0 \forall n \in [a, b]$.

For the second step, we show that $T'(y(n)) \leq 0 \forall n \in [\underline{n}, n_{\underline{b}}]$. From equation (A.6), we have

$$\frac{T'(y(n))}{1 - T'(y(n))} = A(n)D(n)C(n)$$

This equation holds for all $n \in [\underline{n}, n_b]$. We want to show that $A(n)D(n)C(n) \leq 0$. We have $A(n) = \frac{1+\epsilon^u}{\epsilon^c} = 1 + \frac{h''\left(\frac{y(n)}{n}\right)y(n)}{h'\left(\frac{y(n)}{n}\right)} > 0$ and $C(n) = \frac{1}{ng(n)} > 0$. The proof boils down to showing that $\int_{\underline{n}}^n \frac{g(t)}{v'(c(t))} dt \geq \int_{\underline{n}}^n \mu(t) dt$ in which case $D(n) \leq 0$. Rearranging (A.6), we have

$$\int_{\underline{n}}^n \mu(t) dt = \frac{1 + \epsilon^u}{\epsilon^c} \frac{ng(n)}{v'(c(n))} \frac{T'(y(n))}{1 - T'(y(n))} + \int_{\underline{n}}^n \frac{g(t)}{v'(c(t))} dt$$

As a result, we can write

$$\mu(n) = \frac{d}{dn} \left[\frac{1 + \epsilon^u}{\epsilon^c} \frac{ng(n)}{v'(c(n))} \frac{T'(y(n))}{1 - T'(y(n))} \right] + \frac{g(n)}{v'(c(n))}$$

Whenever $\mu(n) > 0$, we have $T'(y(n)) = 0$ and $T''(y(n)) = 0$ because the marginal tax rate along the lower-bound is constant at zero. As a result, either $\mu(n) = 0$ or $\mu(n) = \frac{g(n)}{v'(c(n))}$. This implies that $\int_{\underline{n}}^n \frac{g(t)}{v'(c(t))} dt \geq \int_{\underline{n}}^n \mu(t) dt$.

The third step of the proof consist in noticing that if marginal tax rates are everywhere zero or negative, the lower-bound (along which the marginal tax rate is zero) can be binding on at most one interval from \underline{n} to some n_u .

Finally, because the lower-bound is only binding on interval $[\underline{n}, n_u]$, we have that $\int_{\underline{n}}^n \mu(t) dt = \int_{\underline{n}}^{n_u} \mu(t) dt = \int_{\underline{n}}^{n_u} \frac{g(t)}{v'(c(t))} dt \forall n \in [n_u, n_1]$. Claim 4 follows directly from this and equation (A.6). \square

A.1.6 Solving the lower-subproblem

The lower subproblem consists in

$$V^L(y_1) \equiv \max_{\{u(n), y(n)\}} \int_{\underline{n}}^{n_1} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n)$$

subject to (C1),(C2),(C4), (C5) and

$$y(n) \leq y_1 \quad \forall n$$

We solve the problem by defining the following Hamiltonian:

$$\begin{aligned}
\mathcal{H} \left(u, y, n, \lambda, \mu, \nu^L \right) \equiv & \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) g(n) \\
& + \lambda(n) h' \left(\frac{y(n)}{n} \right) \left(\frac{y(n)}{n^2} \right) \\
& + \mu(n) \left[u(n) - \underline{u(n)} \right] \\
& + \nu^L(n) [y_1 - y(n)]
\end{aligned}$$

Let us first deal with the constraint that $y(n) \leq y_1 \forall n$. Because of (C2), $y(n)$ must be increasing. As a result, this constraint can only be binding on an interval $[n_b, n_1]$. Over this interval, $\nu^L(n) \geq 0$ and $y(n) = y_1$. Solving for ν^L turns out to be important in determining $V_{y_1}^L(y_1)$ which in turns need to be known in order to solve for $y^*(n_1)$. From the first-order conditions (evaluated at the solution $y(n) = y_1$ and corresponding $c(n) = c_1$): $^2 \frac{\partial}{\partial y} \mathcal{H} = 0 \Rightarrow \nu^L(n) =$

$$\left[1 - \frac{1}{v'(c_1)} h' \left(\frac{y_1}{n} \right) \frac{1}{n} \right] g(n) + \lambda(n) \left[h'' \left(\frac{y_1}{n} \right) \left(\frac{y_1}{n^3} \right) + h' \left(\frac{y_1}{n} \right) \left(\frac{1}{n^2} \right) \right]$$

Now, we focus on the solution over $[n, n_b]$. By assumption, the constraint that $y(n) \leq y_1$ is not binding and $\nu^L(n) = 0 \forall n \in [n, n_b]$. Under this assumption, the first-order conditions for $y(n)$ and $u(n)$ respectively yield:

$$\frac{\partial}{\partial y} \mathcal{H} = 0 \Rightarrow$$

$$\left[1 - \frac{1}{v'(c(n))} h' \left(\frac{y(n)}{n} \right) \frac{1}{n} \right] g(n) = -\lambda(n) \left[h'' \left(\frac{y(n)}{n} \right) \left(\frac{y(n)}{n^3} \right) + h' \left(\frac{y(n)}{n} \right) \left(\frac{1}{n^2} \right) \right] \quad (\text{A1})$$

$$\text{and } -\frac{\partial}{\partial u} \mathcal{H} = \lambda'(n) \Rightarrow$$

2. Because $u^*(n_1) = \underline{u_1}$ and y_1 are known, we can infer $c(n_1) = c_1$

$$\lambda'(n) = \frac{1}{v'(c(n))}g(n) - \mu(n)$$

Using the boundary (transversality) condition, $\lambda(\underline{n}) = 0$, we can write:

$$\lambda(n) = \int_{\underline{n}}^n \left\{ \frac{g(t)}{v'(c(t))} - \mu(t) \right\} dt \quad (\text{A2})$$

Using the fact that $T'(y(n)) = 1 - \frac{h'(\frac{y}{n})(\frac{1}{n})}{v'(c)}$, we can re-arrange (A1) into:

$$T'(y(n))g(n) = -\lambda(n) [1 - T'(y(n))] \left[\frac{v'(c(n))}{h'(\frac{y(n)}{n})\frac{1}{n}} h''\left(\frac{y(n)}{n}\right) \left(\frac{y(n)}{n^2}\right) + v'(c(n)) \left(\frac{1}{n}\right) \right]$$

The latter can be re-arranged into

$$\frac{T'(y(n))}{1 - T'(y(n))} = -\frac{\lambda(n)}{ng(n)} v'(c(n)) \left[\frac{h''\left(\frac{y(n)}{n}\right) y(n)}{h'\left(\frac{y(n)}{n}\right)} + 1 \right]$$

Plugging in (A2), we have

$$\frac{T'(y(n))}{1 - T'(y(n))} = - \left[\frac{h''\left(\frac{y(n)}{n}\right) y(n)}{h'\left(\frac{y(n)}{n}\right)} + 1 \right] \frac{v'(c(n))}{ng(n)} \int_{\underline{n}}^n \left[\frac{g(t)}{v'(c(t))} - \mu(t) \right] dt$$

In order to obtain (A.6), we need to show that $\left[\frac{h''\left(\frac{y(n)}{n}\right) y(n)}{h'\left(\frac{y(n)}{n}\right)} + 1 \right]$

$$= \frac{h''\left(\frac{y(n)}{n}\right) y(n) \left(\frac{1}{n}\right) + h'\left(\frac{y(n)}{n}\right) \left(\frac{1}{n}\right)}{h'\left(\frac{y(n)}{n}\right) \left(\frac{1}{n}\right)}$$

$$\begin{aligned}
&= \frac{h''\left(\frac{y(n)}{n}\right)y(n)\left(\frac{1}{n}\right) + h'\left(\frac{y(n)}{n}\right)\left(\frac{1}{n}\right) + (1 - T'(y(n)))^2 y(n)v''(c(n)) - (1 - T'(y(n)))^2 y(n)v''(c(n))}{h'\left(\frac{y(n)}{n}\right)\left(\frac{1}{n}\right)} \\
&= \frac{\frac{h''\left(\frac{y(n)}{n}\right)y(n)\left(\frac{1}{n}\right) - (1 - T'(y(n)))^2 y(n)v''(c(n))}{y\left[h''\left(\frac{y}{n}\right)\left(\frac{1}{n}\right) - (1 - T'(y))^2 v''(c)\right]} + \frac{h'\left(\frac{y(n)}{n}\right)\left(\frac{1}{n}\right) + (1 - T'(y(n)))^2 y(n)v''(c(n))}{y\left[h''\left(\frac{y}{n}\right)\left(\frac{1}{n}\right) - (1 - T'(y))^2 v''(c)\right]}}{\frac{h'\left(\frac{y(n)}{n}\right)\left(\frac{1}{n}\right)}{y\left[h''\left(\frac{y}{n}\right)\left(\frac{1}{n}\right) - (1 - T'(y))^2 v''(c)\right]}} \\
&= \left[\frac{1 + \epsilon^u}{\epsilon^c} \right]
\end{aligned}$$

The final expression is derived using the definition of ϵ^u and ϵ^c given in the main text. Equation (A.6) holds for all $n \in [\underline{n}, \underline{n}_b]$. There are two possible cases. Either the lower-bound on utilities is not binding in n in which case $\mu(n) = 0$ and the marginal tax rate can be obtained from equation (A.6). Or the lower-bound is binding in n in which case $T'(y(n)) = 0$ and $\mu(n)$ can be solved using equation (A.6). In the proof of Proposition 4, we show that the lower-bound on utilities can only be binding on at most one interval $[\underline{n}, n_u]$ for some $\underline{n} \leq n_u \leq n_1$.

Notice that using the envelope theorem, we can find:

$$V_{y_1}^L(y_1) = \int_{n_b}^{n_1} \nu^L(t) dt$$

Finally, we need to solve for n_u and n_b . Recall that we defined utility of type n along the lower-bound as $\underline{u}(n)$. Utility of agents of type 1 is known and fixed at \underline{u}_1 . n_u is the largest type for which the lower-bound is binding:

$$\underline{u}_1 - \int_{n_u}^{n_1} u'(t)dt = \underline{u}(n_u) \quad (\text{A3})$$

$n_{\underline{b}}$ is the lowest type for which the constraint that $y(n) \geq y_1$ is binding. Let $\tilde{y}(n)$ be income implicitly defined by

$\frac{T'(y(n))}{1-T'(y(n))} = - \left[\frac{1+\epsilon^u}{\epsilon^c} \right] \frac{v'(c(n))}{ng(n)} \left[\int_{\underline{n}}^n \frac{g(t)}{v'(c(t))} dt - \int_{\underline{n}}^{n_u} \frac{g(t)}{v'(c(t))} dt \right]$. $n_{\underline{b}}$ is defined by the following equation:

$$\tilde{y}(n_{\underline{b}}) = y_1 \quad (\text{A4})$$

Equation (A3) and (A4) is the system of two equations in two unknowns used to solve for n_u and $n_{\underline{b}}$.

A.1.7 Solving the upper-subproblem

The upper-subproblem is solved similarly. It consists in

$$V^U(y_1) \equiv \max_{\{u(n), y(n)\}} \int_{n_1}^{\infty} \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) dG(n)$$

subject to (C1),(C2),(C4), (C5) and

$$y(n) \geq y_1 \quad \forall n$$

We define the following Hamiltonian:

$$\begin{aligned}
\mathcal{H}(u, y, n, \lambda, \mu, \nu^U) \equiv & \left(y(n) - v^{-1} \left(u(n) + h \left(\frac{y(n)}{n} \right) \right) \right) g(n) \\
& + \lambda(n) h' \left(\frac{y(n)}{n} \right) \left(\frac{y(n)}{n^2} \right) \\
& + \mu(n) \left[u(n) - \underline{u}(n) \right] \\
& + \nu^U(n) [y(n) - y_1]
\end{aligned}$$

From constraint (C2), we know that $y(n)$ is increasing. As a result, the constraint that $y(n) \geq y_1$ can only be binding (if at all) on an interval $[n_1, n_{\bar{b}}]$. On this interval, $y(n) = y_1$ and $\nu^U(n) \geq 0$. From the first-order conditions (evaluated at the solution $y(n) = y_1$ and corresponding $c(n) = c_1$): $\frac{\partial}{\partial y} \mathcal{H} = 0 \Rightarrow$
 $\nu^U(n) =$

$$- \left[1 - \frac{1}{v'(c_1)} h' \left(\frac{y_1}{n} \right) \frac{1}{n} \right] g(n) - \lambda(n) \left[h'' \left(\frac{y_1}{n} \right) \left(\frac{y_1}{n^3} \right) + h' \left(\frac{y_1}{n} \right) \left(\frac{1}{n^2} \right) \right]$$

We now focus on the solution over interval $[n_{\bar{b}}, \infty)$. By definition, $\nu^U(n) = 0$ over this interval. Under this assumption, the first-order conditions for $y(n)$ and $u(n)$ respectively yield:

$$\frac{\partial}{\partial y} \mathcal{H} = 0 \Rightarrow$$

$$\left[1 - \frac{1}{v'(c(n))} h' \left(\frac{y(n)}{n} \right) \frac{1}{n} \right] g(n) = -\lambda(n) \left[h'' \left(\frac{y(n)}{n} \right) \left(\frac{y(n)}{n^3} \right) + h' \left(\frac{y(n)}{n} \right) \left(\frac{1}{n^2} \right) \right] \quad (\text{A5})$$

$$\text{and } -\frac{\partial}{\partial u} \mathcal{H} = \lambda'(n) \Rightarrow$$

$$\lambda'(n) = \frac{1}{v'(c(n))} g(n) - \mu(n)$$

Using the transversality condition, $\lim_{n \rightarrow \infty} \lambda(n) = 0$, we can write:

$$\lambda(n) = - \int_n^\infty \left\{ \frac{g(t)}{v'(c(t))} - \mu(t) \right\} dt \quad (\text{A6})$$

Using the same reasoning as for the lower-subproblem, we can re-arrange (A5) into

$$\frac{T'(y(n))}{1 - T'(y(n))} = - \frac{\lambda(n)}{ng(n)} v'(c(n)) \left[\frac{h''\left(\frac{y(n)}{n}\right) y(n)}{h'\left(\frac{y(n)}{n}\right)} + 1 \right]$$

Plugging in (A6), we have

$$\frac{T'(y(n))}{1 - T'(y(n))} = \left[\frac{h''\left(\frac{y(n)}{n}\right) y(n)}{h'\left(\frac{y(n)}{n}\right)} + 1 \right] \frac{v'(c(n))}{ng(n)} \int_n^\infty \left[\frac{g(t)}{v'(c(t))} - \mu(t) \right] dt$$

Using a result derived in appendix A1.4.1, we obtain equation (A8):

$$\frac{T'(y(n))}{1 - T'(y(n))} = \left[\frac{1 + \epsilon^u}{\epsilon^c} \right] \frac{v'(c(n))}{ng(n)} \int_n^\infty \left[\frac{g(t)}{v'(c(t))} - \mu(t) \right] dt \quad (\text{A8})$$

Finally, using the envelope theorem, we can write

$$V_{y_1}^U(y_1) = \int_{n_1}^{n_b} \nu^U(t) dt$$

A.1.8 Deriving the Optimal Income Tax Formulas under Quasilinear and Iso-elastic Preferences and Proof of Proposition 5

Proof. Equalizing well-being in an interval of types, $\frac{\partial}{\partial n} b(u(n)) = 0$, implies:

$$u'(n) = n^\epsilon \quad \forall n \in [\underline{n}, n_1] \quad (\text{A.12})$$

$$u'(n) = \left[1 - \left(\frac{\bar{\theta}}{\theta} \right)^{1+\epsilon} \right] n^\epsilon \quad \forall n \in [n_1, \infty) \quad (\text{A.13})$$

The change in the slope of the lower-bound at n_1 comes from the fact that we are

equalizing the well-being index between individuals with the same skill (\underline{w}) before n_1 and between individuals with the same preference parameter ($\bar{\theta}$) but different skill after n_1 .

Combining Eq. A.12 and A.13 with the first-order conditions of the incentive compatibility constraint, adapted from constraint C1, $\left(u'(n) = \left(\frac{y(n)}{n}\right)^{\frac{1+\epsilon}{\epsilon}} \left(\frac{1}{n}\right)\right)$, we find income as a function of n at this allocation:

$$y(n) = n^{1+\epsilon} \quad \forall n \in [\underline{n}, n_1) \quad (\text{A.14})$$

$$y(n) = \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{\epsilon}{1+\epsilon}} n^{1+\epsilon} \quad \forall n \in [n_1, \infty) \quad (\text{A.15})$$

Finally, using the fact, adapted from Eq. 1.5, that $1 - T'(y(n)) = \left(\frac{y(n)}{n}\right)^{\frac{1}{\epsilon}} \left(\frac{1}{n}\right)$, we can derive the marginal tax rate along the lower-bound on utilities:

$$T'(y(n)) = 0 \quad \forall n \in [\underline{n}, n_1) \quad (\text{A.16})$$

$$T'(y(n)) = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \quad \forall n \in [n_1, \infty) \quad (\text{A.17})$$

Without restrictions on $G(n)$, we cannot rule out cases where the lower-bound on utilities is binding over multiple intervals. In our calibrations (for which we use a double Pareto log-normal), there is always only one such interval which we call $[n_l, n^*]$ with $n_l \leq n^* \in (n_{\bar{b}}, \infty)$.

Regarding the optimal tax rate in the upper tail, notice that when $\int_n^\infty \mu(t) dt = 0$, Eq. 1.8 simply characterizes the marginal tax rates that maximize tax revenue. It is well known that this tax rate converges to $\tau'(y(n)) \rightarrow \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ when $n \rightarrow \infty$.³ Two scenarii are possible depending on whether the marginal tax rates along the lower-bound of utilities are

3. See Boadway and Jacquet (2008) for a discussion of the Rawlsian marginal tax rates under quasilinear preferences and various distributions of skills.

lower or larger than revenue maximizing marginal tax rates in the upper tail. If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \leq 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$, there exist some n^* such that $T'(n^*) = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$ and $T'(n) \leq 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \forall n \in [n^*, \infty)$. Assuming that the lower-bound is binding over interval $[n_l, n^*]$ only, we have:

$$\int_n^\infty \mu(t)dt = \int_{n_l}^{n^*} \mu(t)dt = 1 - G(n_l) - \frac{1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right) \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}} n_l g(n_l)$$

$$\forall n \in [n_1, n_l].^4$$

Combining the latter with Eq. 1.8 yields an expression for marginal tax rates over $[n_{\bar{b}}, n_l]$.

If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \geq 1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$, the lower-bound on utilities remain binding in the tail. In that case, marginal tax rates over $[n_{\bar{b}}, n_l]$ are the same as in the previous case but tax rates over interval $[n_l, \infty)$ are given by Eq. A.17.⁵ \square

A.2 Other calibrated optimal tax rates and comparative statics

In this section, we show the optimal tax rates under different calibrations of the slope (figure A.3) and level (figure A.4) of the poverty line.

4. The derivation of this formula uses the definition of n^* which implies $1 - G(n^*) = \frac{1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right) \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}} n^* g(n^*)$.

5. Because $\lim_{n \rightarrow \infty} G(n) = 1$ and $\lim_{n \rightarrow \infty} \lambda(n) = 0$ by the transversality condition, we can derive $\int_n^\infty \mu(t)dt = 1 - G(n_l) - \frac{1 - \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right) \left[1 - \left(\frac{\tilde{\theta}}{\theta}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}} n_l g(n_l)$. This expression combined with Eq. 1.8 yields the desired expression for marginal tax rates over $[n_{\bar{b}}, n_l]$.

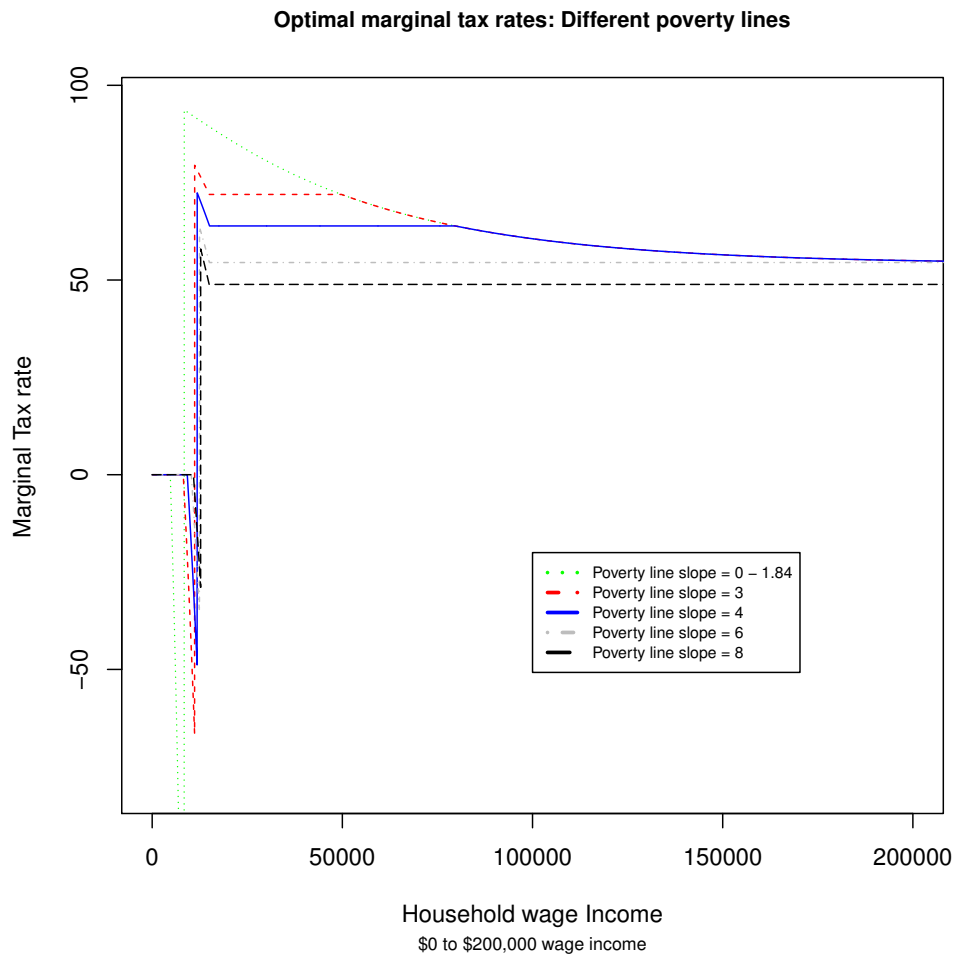


Figure A.3: Calibration of the optimal marginal tax rates according to different slopes of the poverty line.

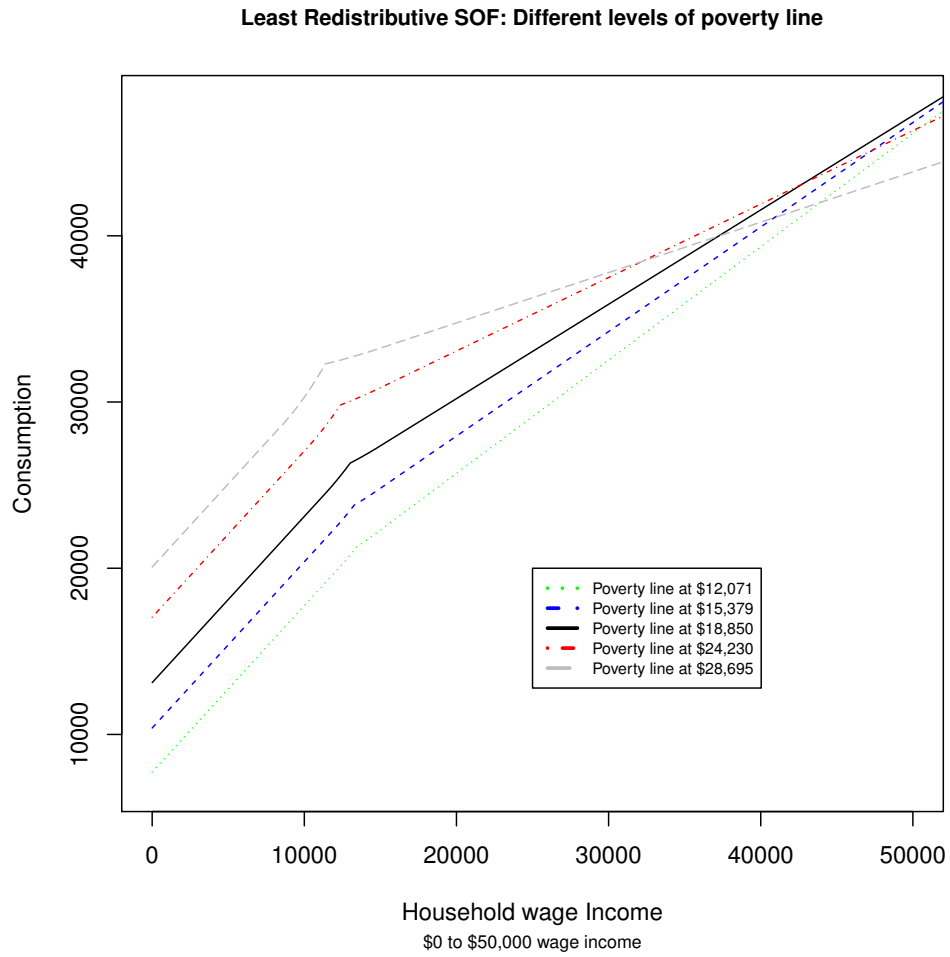


Figure A.4: Calibration of the optimal consumption schedule according to different levels of the poverty line.

A.3 Calibration details and description of the current US income tax schedule

The same sample is used and the same sample restrictions are applied to estimate the distribution of types, the net revenue raised by the government at the statu-quo and the current total marginal tax rate faced by households of different structure along the income distribution.

We pool CPS data from 2013-2015 and drop all households in group quarters and households with negative or zero income. We also drop households with any self-employment or farming income because of measurement error concerns. We drop households with any retirement income or whose head is above 60. Finally, we also drop married couples that are not of legal age to be married (because it might create issues in the TAXSIM simulator).

Marginal federal tax rate and welfare receipt We consider that married couples (with spouse present) are joint filers while other households are single filers. We feed this information along with the age, wage and salary income of each spouse, state of residence and year into NBER's TAXSIM. We also infer the number of dependent exemptions claimed by the head of households from the family relationship and household structure. We obtain estimates, for each households, of the federal tax liability, federal marginal tax rate, payroll tax liability (fica) and marginal payroll tax rate. The federal marginal tax rate is defined as the sum of these two tax rates and obviously include the EITC and Child Tax Credit. The tax schedule is then "smoothed" using a kernel regression.

In order to obtain a more accurate description of the labor supply incentives that households face, we also include a measure of average cash welfare receipt at each point in the wage distribution. We sum the supplemental security income (SSI), Temporary Assistance for Needy Families (TANF) and value of food stamps (SNAP) reported by each household in the CPS. We adjust for underreporting of welfare receipt in the CPS using the dollar reporting rate for each program (and specific to the CPS) documented in Meyer, Mok and

Sullivan (2015). The reporting rates can be found in appendix tables 3,4 and 8 of their study. We average the values for the four most recent years available (2009-2012). As a result, each respondent's reported welfare receipt is divided by 0.8758 for SSI, 0.5706 for SNAP and 0.573 for TANF/welfare income. Finally, we use a local kernel regression to estimate the average relationship between welfare receipt and wage income. We take the derivative to estimate phaseout of these programs and obtain an additional implicit marginal tax that we add to the marginal federal tax rate in our descriptions of the current tax schedule.

Below are graphs depicting the resulting estimates of marginal tax rates along the income distribution. The marginal tax rates for joint filers by number of children are in the main text.

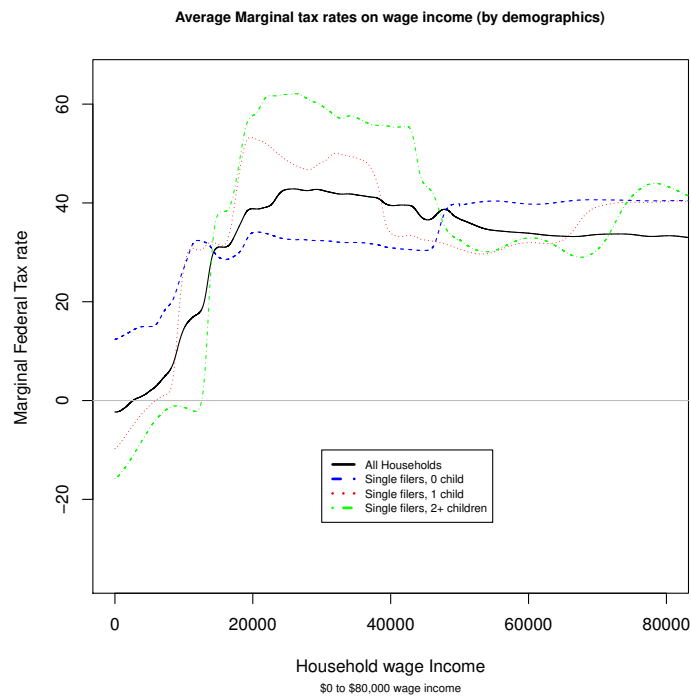


Figure A.5: Average Marginal Tax Rates - Single Filers - \$0-\$80,000

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

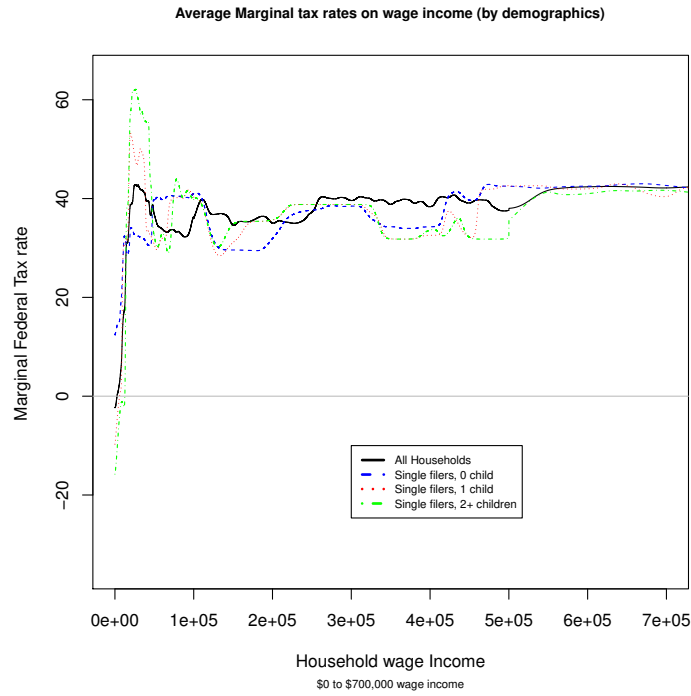


Figure A.6: Average Marginal Tax Rates - Single Filers - \$0-\$700,000

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

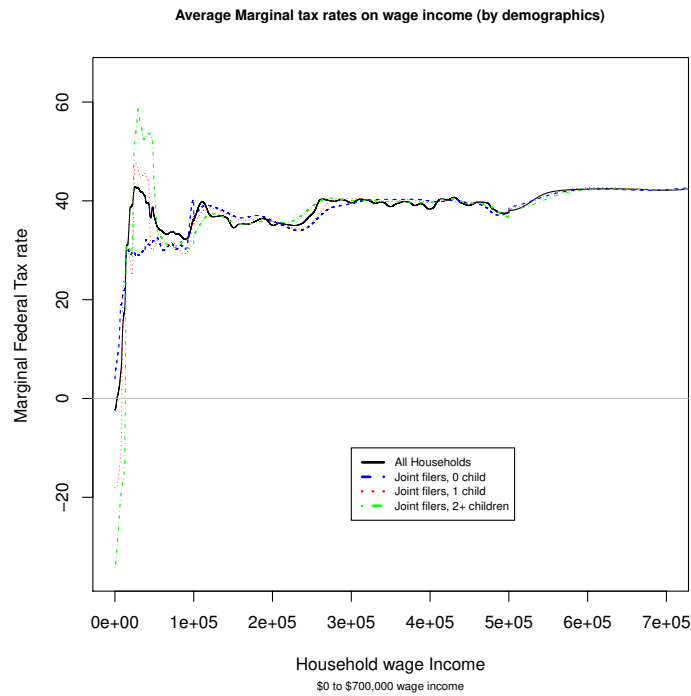


Figure A.7: Average Marginal Tax Rates - Joint Filers - \$0-\$700,000

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

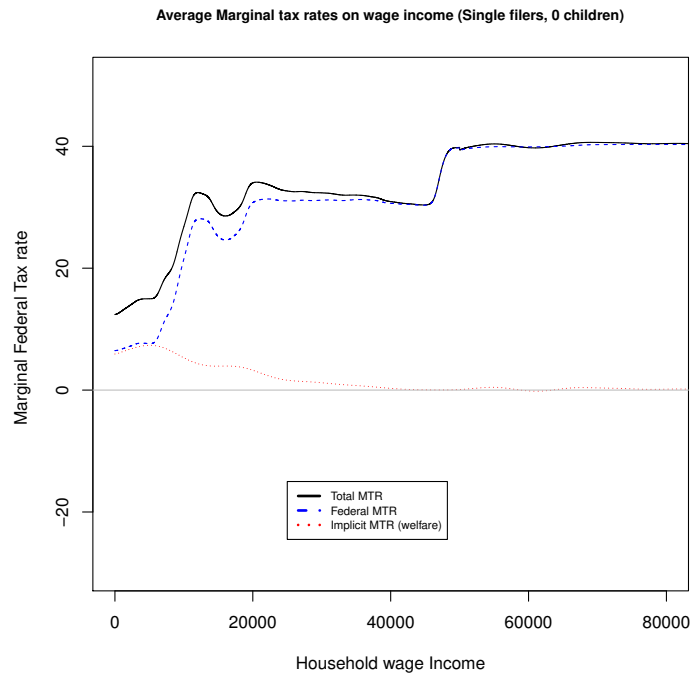


Figure A.8: Average Marginal Tax Rates (Zoom) - Single Filers without Children

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

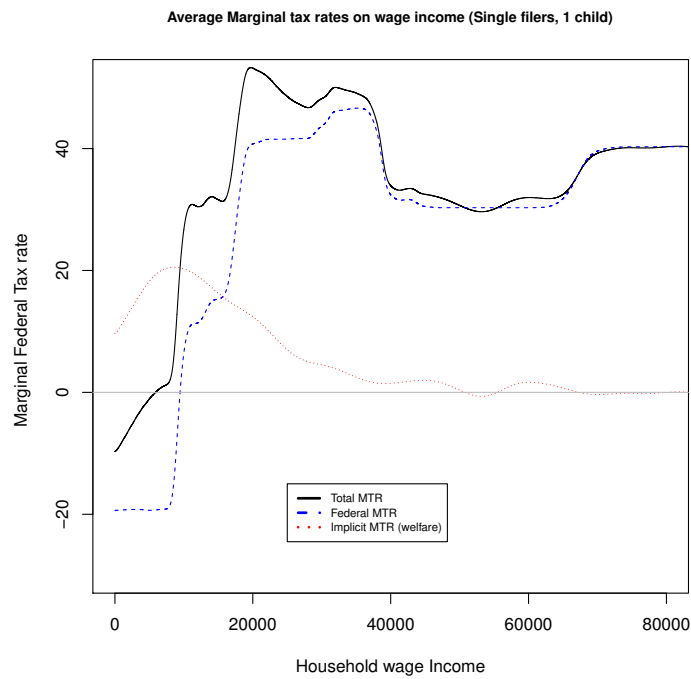


Figure A.9: Average Marginal Tax Rates (Zoom) - Single Filers with 1 Child

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

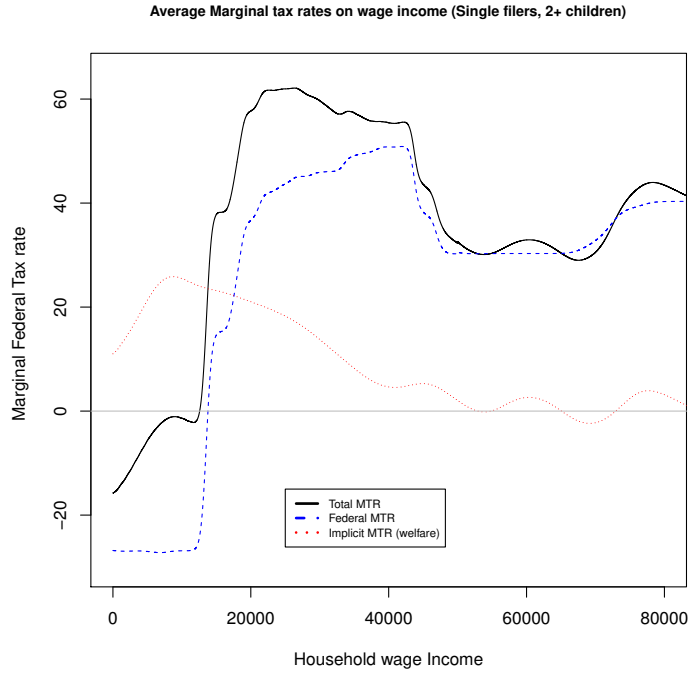


Figure A.10: Average Marginal Tax Rates (Zoom) - Single Filers with 2 Children

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

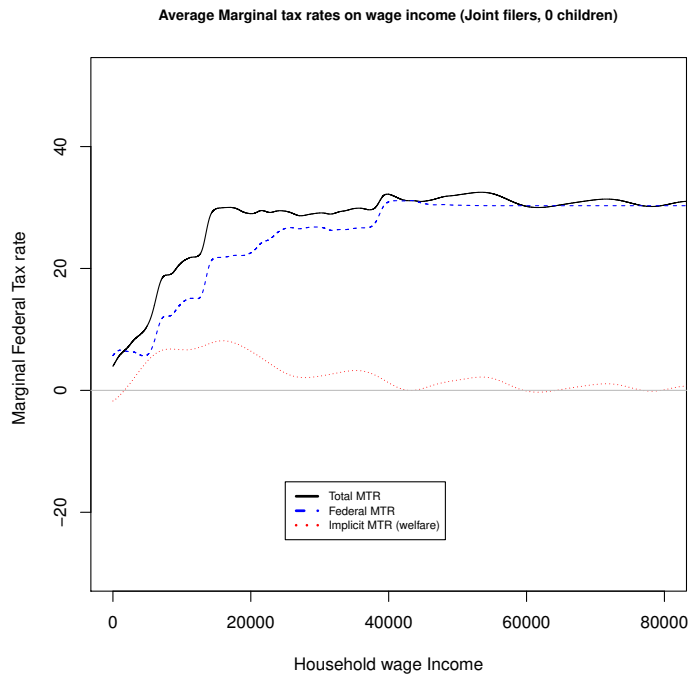


Figure A.11: Average Marginal Tax Rates (Zoom) - Joint Filers without Children

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

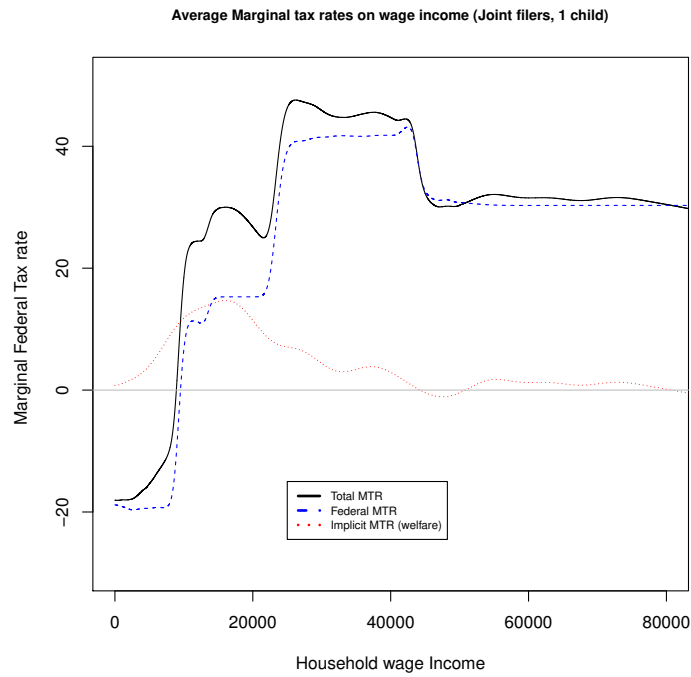


Figure A.12: Average Marginal Tax Rates (Zoom) - Joint Filers with 1 Child

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

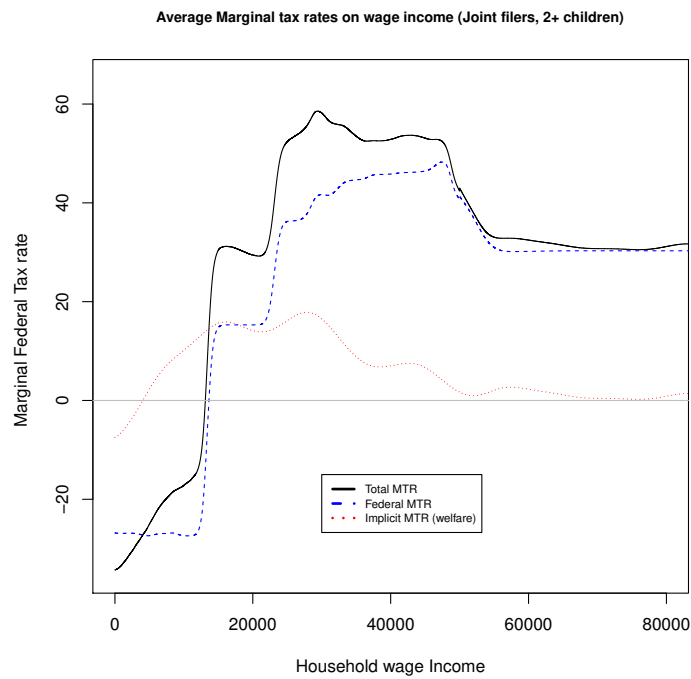


Figure A.13: Average Marginal Tax Rates (Zoom) - Joint Filers with 2 Children

Note: Source of data: CPS 2013-2015 and NBER TAXSIM.

Estimation of distribution of types The distribution of types is estimated via Maximum Likelihood using the same sample as for the description of the current tax schedule. This guarantees consistency in our comparisons between optimal and current tax rates. We estimate a four parameter double Pareto Lognormal distribution. For each household in the data, we simulate the total marginal federal tax rate faced by households using the procedure described above. Conditional on a labor supply elasticity parameter (we choose $\epsilon = 0.33$), the marginal tax rate and wage income of the households, we can recover type n by inverting the first-order condition:

$$n_i = (1 - \tau_i)^{-\frac{\epsilon}{1+\epsilon}} y_i^{\frac{1}{1+\epsilon}}$$

The distribution is then estimated by maximizing the log-likelihood of this sample of types n . Using the notation of Hajargasht and Griffiths (2013), the estimated parameters are: $m = 8.56$, $\sigma = 0.34$, $\alpha = 3.42$ and $\beta = 1.55$. α governs behavior in the upper tail. Because the marginal rate in the upper tail is approximately constant and given the first-order conditions of individual behavior, this translates into a Pareto tail of about $\alpha_{income} = \frac{\alpha}{1+\epsilon} = 2.57$ which is well within the range of previous estimates in the literature. The fit of the estimated distribution to the empirical distribution is depicted in the graphs below.

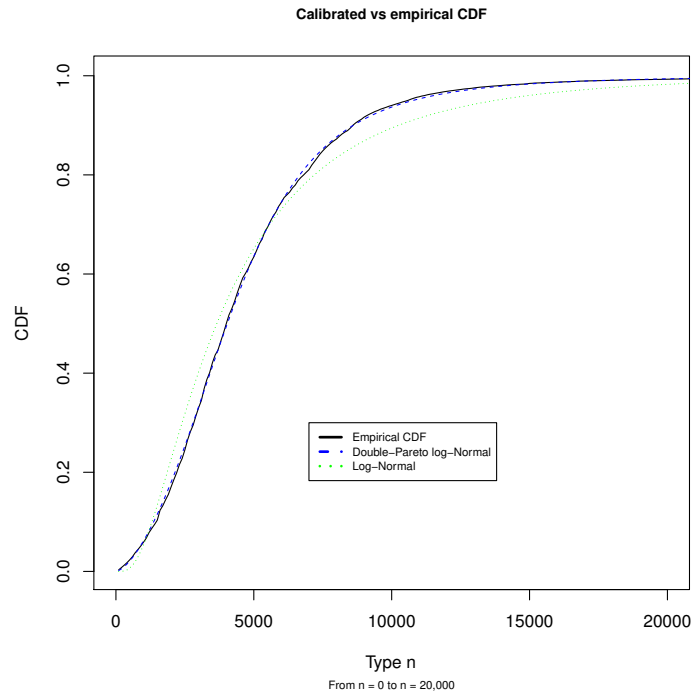


Figure A.14: Double Pareto Lognormal estimated on CPS data (2013-2015) versus the empirical CDF.

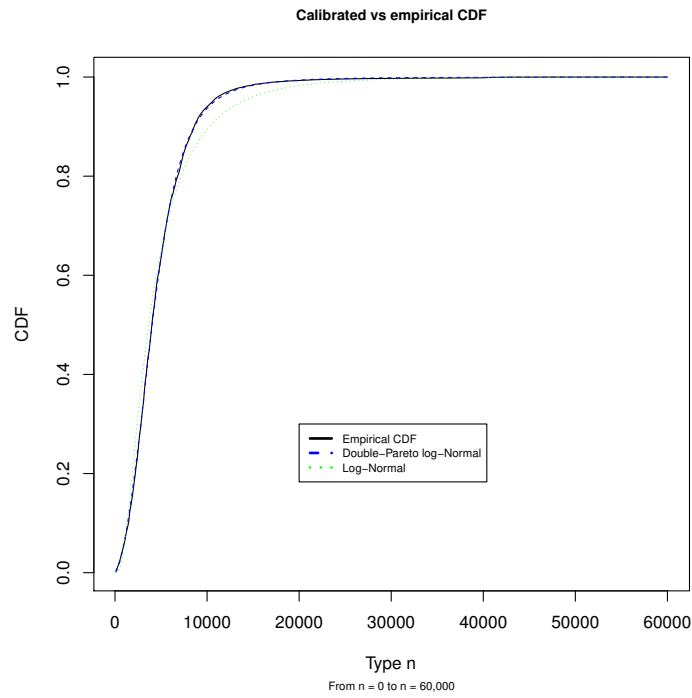


Figure A.15: Double Pareto Lognormal estimated on CPS data (2013-2015) versus the empirical CDF.

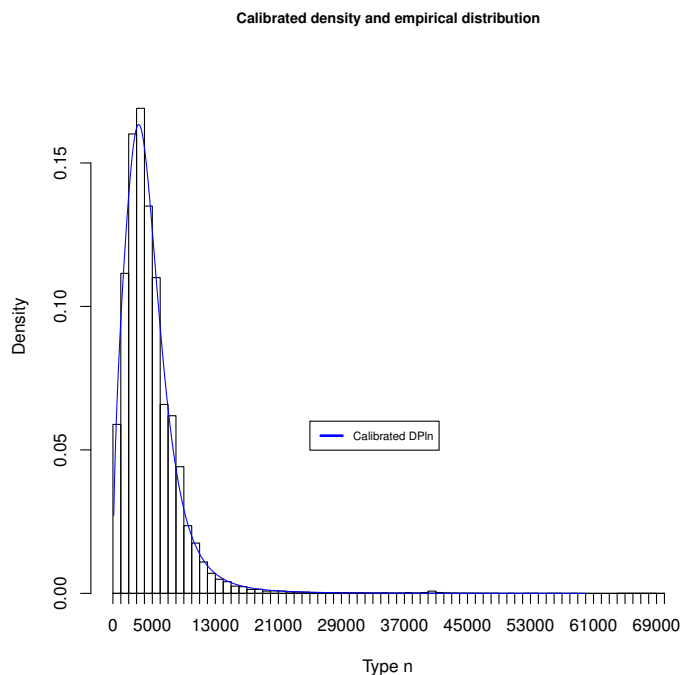


Figure A.16: Double Pareto Lognormal estimated on CPS data (2013-2015) versus empirical histogram.

Calibrating the other parameters We pick a labor supply elasticity $\epsilon = 0.33$ which is the preferred estimate in Chetty (2012). The value of the minimum skill in the economy and of the most hardworking preference parameter at that skill level show up in the optimal tax formula. We assume that the minimum skill is equal to the federal minimum wage $\underline{w} = \$7.25$. We choose $\bar{\theta}$ by assuming that the most hardworking individuals facing a wage of \$7.25 and marginal tax rate of %10 work 40 hours a week and 48 weeks per year. We vary this parameter in our comparative statics. In particular, we also consider the case in which the same most hardworking individuals work 35 and 44 hours a week. We explore the sensitivity of the tax formula to this parameter.

We assume that the government must raise \$12,152 on average per household to finance public goods. This was estimated again in our sample of CPS data. This is the average net simulated federal tax (Federal tax + payroll tax - welfare receipt) paid in the sample.

Regarding the poverty line in our calibration, $\tilde{\theta}$ governs the slope. We pick $\tilde{\theta}$ by choosing an average slope for the poverty line between a labor time of 0 and "full-time" $1920 = 40 * 48$.

The level of the poverty line also matters for the calibration of the "least redistributive" SOF. We consider five different values to anchor the level of the poverty line (all evaluated at a labor time of zero): \$12,071, \$15,379, \$18,850, \$24,230 and \$28,695. These are the weighted average poverty line thresholds in 2014 in the US for families of one to five people respectively⁶.

A.4 Least Redistributive SOF: Parametrizing the reference preferences

We define more formally the reference preferences used to calibrate the "least redistributive" social ordering function from section 1.9.2. The reference preferences can be represented by the following utility function for all indifference sets below the poverty line:

$$\text{If } c - \frac{\epsilon}{1+\epsilon} \left(\frac{\ell}{\bar{\theta}}\right)^{\frac{1+\epsilon}{\epsilon}} \leq \bar{P}_0:$$

$$\tilde{u}_L(c, \ell) = c - \frac{\epsilon}{1+\epsilon} \left(\frac{\ell}{\bar{\theta}}\right)^{\frac{1+\epsilon}{\epsilon}}$$

For all indifference sets above the poverty line, preferences can be represented by the following utility function:

$$\text{If } c - \frac{\epsilon}{1+\epsilon} \left(\frac{\ell}{\bar{\theta}}\right)^{\frac{1+\epsilon}{\epsilon}} > \bar{P}_0:$$

$$\tilde{u}_U(c, \ell) = -\ell \left[\frac{1+\epsilon}{\epsilon} (c - \bar{P}_0) \right]^{-\frac{\epsilon}{1+\epsilon}}$$

The utility function for bundles above the poverty line comes from the fact that $\ell \left[\frac{1+\epsilon}{\epsilon} (c - \bar{P}_0) \right]^{-\frac{\epsilon}{1+\epsilon}}$ is the value of θ such that $\bar{P}_0 = c - \frac{\epsilon}{1+\epsilon} \left(\frac{\ell}{\theta}\right)^{\frac{1+\epsilon}{\epsilon}}$. We add a minus sign so that \tilde{u}_U is increasing in c and decreasing in ℓ .⁷ As mentioned in the main text, calibration

6. Source: DeNavas-Walt and Proctor (2015).

7. More labor averse indifference curves are represented by a smaller θ . In this case, we want higher c or smaller ℓ to be associated with a smaller θ but higher utility.

using these reference preferences amount to finding the slope parameter $\tilde{\theta}$ for the poverty line such that $b(u(n_1)) = \bar{P}_0$. In other words, we solve for $\tilde{\theta}$ in the following equation:

$$\bar{P}_0 = u_1(n_1) + \frac{1}{1+\epsilon} \left[(\underline{w}\tilde{\theta})^{1+\epsilon} - (\underline{w}\bar{\theta})^{1+\epsilon} \right]$$

APPENDIX B

APPENDIX TO CHAPTER 2

B.1 Additional Figures and Tables

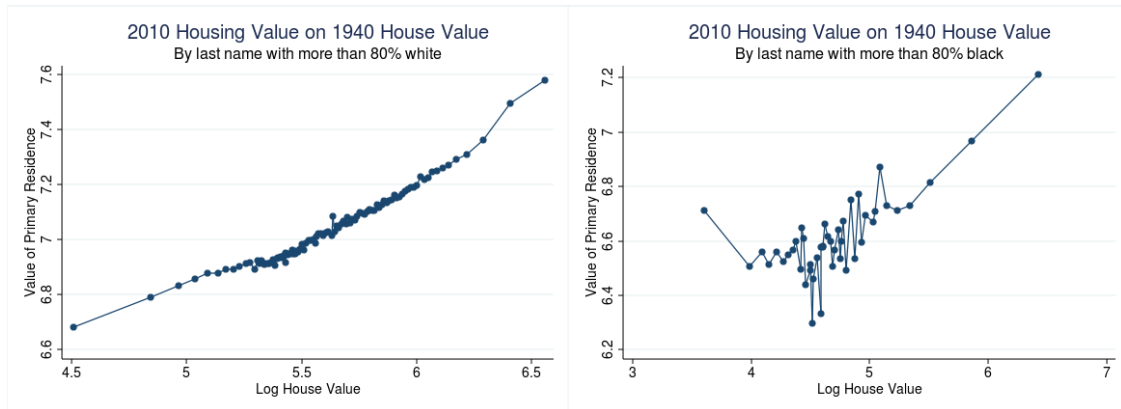


Figure B.1: Intergenerational correlation in housing consumption - mostly white versus mostly black last-names.

Note: Binscatter plot of last-names with more than 80% whites and more than 80% blacks respectively in 1940. Observations are weighted by size of last name. Regression coefficients are 0.42 and 0.21 respectively.

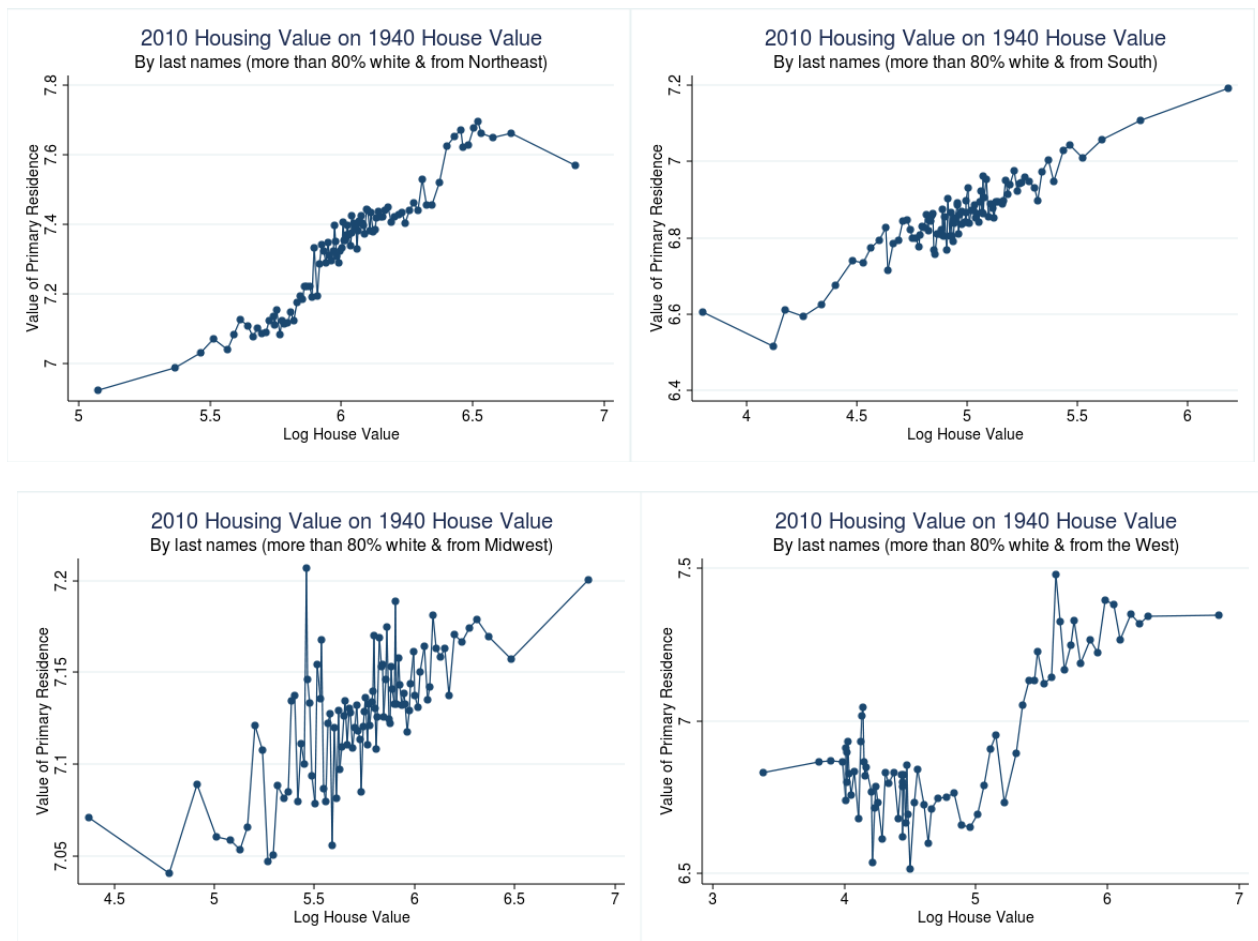


Figure B.2: Intergenerational correlation in housing consumption - Regional Heterogeneity for mostly white last names

Notes: Binscatter plot of last-names with more than 80% whites and in the Northeast, 80% whites and in the South, 80% whites and in Midwest and 80% whites and in the West in 1940 respectively. Observations are weighted by size of last name. Regression coefficients are 0.42 and 0.21 respectively.

Table B.1: Mobility Estimates adjusted for selection (β)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	0.531 (0.007)	0.105 (0.005)	0.241 (0.006)	0.106 (0.01)	0.204 (0.003)
Black	0.111 (0.089)	-0.144 (0.095)	0.164 (0.025)	-0.019 (0.196)	0.012 (0.019)
Other	-0.159 (0.077)	0.19 (0.039)	0.276 (0.048)	0.304 (0.024)	0.239 (0.018)
All	0.516 (0.008)	0.104 (0.005)	0.213 (0.007)	0.079 (0.01)	0.176 (0.003)

Note : Estimates of mobility by race and region adjusted for selection into 2012 home ownership: regression of 2012 log housing consumption on 1940 log housing consumption - using surnames as instruments. β is the slope coefficient. The selection adjustment assumes joint normality of the error terms in the selection and outcome equations. More details are provided in section 6. Observations are weighted by number of male heads of households.

Table B.2: Mobility Estimates adjusted for selection (α)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	9.5 (0.071)	12.04 (0.072)	11.575 (0.078)	12.359 (0.091)	11.272 (0.068)
Black	11.977 (0.572)	12.522 (0.527)	12.078 (0.123)	12.009 (1.1)	12.117 (0.133)
Other	14.581 (0.51)	12.159 (0.22)	11.997 (0.243)	12.464 (0.143)	12.549 (0.139)
All	9.622 (0.077)	12.056 (0.075)	11.69 (0.077)	12.362 (0.091)	11.367 (0.07)

Note : Estimates of mobility by race and region adjusted for selection into 2012 home ownership: regression of 2012 log housing consumption on 1940 log housing consumption - using surnames as instruments. α is the intercept. The selection adjustment assumes joint normality of the error terms in the selection and outcome equations. More details are provided in section 6. Observations are weighted by number of male heads of households.

Table B.3: Average 1940 Housing Consumption (Per Demography)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	6.117 (0.002)	5.711 (0.002)	5.152 (0.003)	5.717 (0.002)	5.668 (0.004)
Black	5.856 (0.003)	5.476 (0.002)	4.348 (0.002)	5.462 (0.007)	4.631 (0.003)
Other	6.039 (0.007)	5.233 (0.015)	4.376 (0.015)	4.77 (0.02)	4.84 (0.019)
All	6.108 (0.002)	5.702 (0.002)	4.966 (0.006)	5.655 (0.005)	5.564 (0.006)

Note : Average 1940 (log-) Housing Consumption per demographic group. Data are from the 1940 full Census. Housing consumption is defined as the monthly rent (for renters) or house value divided by 100 expressed in 2012 \$.

Table B.4: Estimates of Average 2012 Housing Consumption (Per Demographics)

	Region				
	Northeast	Midwest	South	West	All
Race					
White	7.346 (0.004)	7.032 (0.004)	6.839 (0.005)	7.134 (0.011)	7.076 (0.002)
Black	6.855 (0.057)	5.526 (0.068)	6.574 (0.015)	5.75 (0.185)	6.471 (0.012)
Other	7.225 (0.051)	6.562 (0.048)	6.285 (0.03)	7.085 (0.023)	6.794 (0.014)
All	7.329 (0.005)	6.983 (0.004)	6.773 (0.004)	7.116 (0.01)	7.019 (0.004)

Note : Estimates of average 2012 (log-) Housing Consumption per demographic group: $\bar{h}_{2012}^g = \hat{\alpha}^g + \hat{\beta}^g \bar{h}_{1940}^g$. Parameters $\hat{\alpha}^g$ and $\hat{\beta}^g$ are the group-specific estimates reported in tables 8 and 9 and \bar{h}_{1940}^g is average log-housing consumption per group in 1940 and reported in table 13. Bootstrapped standard errors are in parenthesis.

B.2 Converting Durable expenditures into Service Flows in the CEX

This section follows Meyer and Sullivan (2017)'s data appendix very closely. For simplicity and because there are a couple of very minor differences with what they do, we describe the computation of service flows in this section.

Housing Services: For renters, housing consumption is defined as the annual rent they pay. For Homeowners, it is the self-reported rental equivalent for their house. Homeowner housing expenses such as mortgage interest, property tax, maintenance, repairs, insurance and other expenses are subtracted from total expenditures and replaced by the rental equivalent. For renters in public housing (only 2.37% of the sample), we impute a rental value using a quantile regression of the log rent on an urban indicator, region and state dummies, the SMSA status, number of bathrooms, bedrooms, half bathrooms and rooms, the presence of window or central air conditioning as well as region specific cubic trends. We impute the predicted 40th quantile of rent based on the house and geographic characteristics. This is justified by evidence from the PSID mentioned by Meyer and Sullivan (2017).

Vehicle Service Flows: The purchase price of new and used vehicles as well as finance charges are subtracted from expenditures and converted into a service flow. The service flow is computed as $\delta * (1 - \delta)^t * P$ where P is the original purchase price (in real terms) and t is the number of years since the vehicle was bought. The depreciation rate, δ is estimated by regressing log price of the car on its age at the time of purchase and make-model-year fixed effects. This regression is ran on a sample of all vehicles for which the purchase price is reported and which were bought within one year of the interview (this sample includes 172,160 vehicles). The coefficient from this regression is -0.17 . It is then converted into a depreciation rate by taking $\delta = 1 - \exp(-0.17) = 0.156$. There is a substantial number of vehicles in the sample for which the purchase price is not reported.

We impute a current market price for these vehicles by running the regression described below on the same estimation sample as above (sample of cars bought within 12 months of the interview with a reported purchase price). We run two separate regressions for the years 1980-2002 and 2003-2016 because of a change in the set of variables available in those years. For years 1980-2002, we regress log purchase price on a cubic in the age of the vehicle, an indicator for new/used, automatic transmission, power brake, power steers, air conditioning, diesel, an urban indicator, a quadratic in log total family expenditures, fixed effects for family size, age and education of the head of household and make-model-year fixed effects. For years, 2003-2016, the independent variables are: a cubic in the age of the vehicle, an indicator for new/used, an urban indicator, a quadratic in log total family expenditures, fixed effects for family size, age and education of the head of household and make-year fixed effects. The coefficients from those regressions are used to predict current market price where it is missing. The prediction is scaled up by the coefficient of a regression (again in the estimation sample) ,without constant, of the reported purchase price on the predicted price. This is because by Jensen's inequality - we ran the regression on log price - the predicted price *in levels* will tend to be an underestimate.

B.3 Derivation of exclusion restrictions in terms of δ_{0i} , δ_{1i}

We assume the data generating process in Eq. (2.2):

$$c_{i,t+1} = \delta_{0i} + \delta_{1i}c_{i,t}$$

Under this generating process the OLS parameters α and β are:

$$\beta = E[\delta_{1i}] + \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} + \frac{Cov((\delta_{1i} - E[\delta_{1i}])c_{it}, c_{it})}{Var(c_{it})} \quad (B.1)$$

$$\alpha = E[\delta_{0i}] - \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})}E[c_{it}] - \frac{Cov((\delta_{1i} - E[\delta_{1i}])c_{it}, c_{it})}{Var(c_{it})}E[c_{it}] \quad (B.2)$$

The prediction error is $\varepsilon_{it+1} = (\delta_{0i} - \alpha) + (\delta_{1i} - \beta) c_{it}$. Plugging the latter formula, Eq. (B.1) and Eq. (B.2) in the exclusion restriction, we can derive:

$$\begin{aligned} E[\varepsilon_{it+1} z_i^\ell] &= E[\delta_{0i} z_i^\ell] - E[\delta_{0i}] E[z_i^\ell] + E[(\delta_{1i} - E[\delta_{1i}]) c_{it} z_i^\ell] - E[(\delta_{1i} - E[\delta_{1i}]) c_{it}] E[z_i^\ell] \\ &\quad + \left(\frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} + \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})} \right) (E[c_{it}] E[z_i^\ell] - E[c_{it} z_i^\ell]) \end{aligned}$$

Using the fact that the covariance between two variables is the expectation of the product minus the product of expectations, we have:

$$\begin{aligned} E[\varepsilon_{it+1} z_i^\ell] &= Cov(\delta_{0i}, z_i^\ell) + Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell) \\ &\quad - \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} Cov(c_{it}, z_i^\ell) \\ &\quad - \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})} Cov(c_{it}, z_i^\ell) \end{aligned}$$

Evaluating whether this expression is equal to zero and dividing both sides by $Var(z_i^\ell)$, we obtain:

$$\begin{aligned} E[\varepsilon_{it+1} z_i^\ell] = 0 &\Leftrightarrow \\ \frac{Cov(\delta_{0i}, z_i^\ell)}{Var(z_i^\ell)} + \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell)}{Var(z_i^\ell)} &= \\ \left(\frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} + \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})} \right) \frac{Cov(c_{it}, z_i^\ell)}{Var(z_i^\ell)} & \end{aligned}$$

A sufficient condition is:

$$\frac{Cov(\delta_{0i}, z_i^\ell)}{var(z_i^\ell)} = \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})} \frac{Cov(c_{it}, z_i^\ell)}{var(z_i^\ell)}$$

and

$$\frac{Cov\left((\delta_{1i} - E[\delta_{1i}])c_{it}, z_i^\ell\right)}{Var\left(z_i^\ell\right)} = \frac{Cov\left((\delta_{1i} - E[\delta_{1i}])c_{it}, c_{it}\right)}{Var\left(c_{it}\right)} \frac{Cov\left(c_{it}, z_i^\ell\right)}{var\left(z_i^\ell\right)}$$

Intuitively, it compares (focusing on the first of the two conditions) the coefficient of a regression of δ_{0i} on z_i^ℓ to the product of the coefficients resulting from regressing δ_{0i} on c_{it} and c_{it} on z_i^ℓ respectively. So the changes in expected δ_{0i} induced by conditioning on a different last name must be exactly equal to the change in expected δ_{0i} induced by a marginal change in c_{it} times the change in expected c_{it} induced by conditioning on a different last name. A concern arises if last names are very informative - through a large between surname variance for instance - about an omitted variable that correlates strongly with δ_{0i} or δ_{1i} :

$$\frac{Cov\left(u_i, z_i^\ell\right)}{var\left(z_i^\ell\right)} \gg \frac{Cov\left(u_i, c_{it}\right)}{c_{it}} \frac{Cov\left(c_{it}, z_i^\ell\right)}{var\left(z_i^\ell\right)}$$

and $Cov(u_i, \delta_{0i}) \neq 0$ or $Cov(u_i, \delta_{1i}) \neq 0$. This would lead to biased estimates when using last names as instruments. In practice, such concerns are relevant because surnames are informative about individuals ethnicity for instance. In our preferred estimates, we adjust for race.

Re-arranging again, we obtain another interesting condition - suggestive of the placebo tests performed in section 4:

$$E\left[\varepsilon_{it+1}z_i^\ell\right] = 0 \Leftrightarrow \frac{Cov\left(\delta_{0i}, z_i^\ell\right)}{Cov\left(c_{it}, z_i^\ell\right)} + \frac{Cov\left((\delta_{1i} - E[\delta_{1i}])c_{it}, z_i^\ell\right)}{Cov\left(c_{it}, z_i^\ell\right)} = \frac{Cov\left(\delta_{0i}, c_{it}\right)}{var\left(c_{it}\right)} + \frac{Cov\left((\delta_{1i} - E[\delta_{1i}])c_{it}, c_{it}\right)}{Var\left(c_{it}\right)}$$

With the associated sufficient condition:

$$\frac{Cov(\delta_{0i}, z^\ell)}{Cov(c_{it}, z_i^\ell)} = \frac{Cov(\delta_{0i}, c_{it})}{var(c_{it})}$$

and

$$\frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, z_i^\ell)}{Cov(c_{it}, z_i^\ell)} = \frac{Cov((\delta_{1i} - E[\delta_{1i}]) c_{it}, c_{it})}{Var(c_{it})}$$

B.4 Choice of the weighting matrix and asymptotic standard errors

In practice, we have to choose a weighting matrix combining the L moment conditions into a 2SIV estimator. We considered multiple options and decided to use the weighting matrix corresponding to a Two-Sample-Two-Stage-Least-Square estimator. Ridder and Moffitt (2003) derive the asymptotic distribution of the two-sample IV estimator under assumptions (A1)-(A4) (in their paper). Applying their formula to our context:

$$\sqrt{N_2} (\hat{B} - B) \rightarrow^d N(0, V(B))$$

with

$$V(B) = \{E[X^T Z] W E[Z^T X]\}^{-1} \cdot E[X^T Z] W \left\{ \lambda \text{Diag}(Var_\ell(c_{it+1})) + \text{Diag} \left(Var_\ell \left(\sum_g (\alpha^g + \beta^g c_{it}) D_i^g \right) \right) \right\} W E[Z^T X] \cdot \{E[X^T Z] W E[Z^T X]\}^{-1}$$

Where $\text{Diag}(\vec{A})$ is the diagonal matrix whose diagonal is vector \vec{A} , $Var_\ell(\cdot)$ is the within-last name ℓ variance and $\lambda = \lim_{N_1 \rightarrow \infty, N_2 \rightarrow \infty} \frac{N_2}{N_1}$ (assumption (A4) in Ridder and Moffitt (2003)).

The asymptotically “efficient” estimator would use an estimate of the “Optimal weighting matrix” which is the sample analog of:

$$\left\{ \lambda \text{Diag}(\text{Var}_\ell(c_{it+1})) + \text{Diag} \left(\text{Var}_\ell \left(\sum_g (\alpha^g + \beta^g c_{it}) D_i^g \right) \right) \right\}^{-1}$$

and we would use estimates (from a first step estimation using $W = \mathcal{I}$, of α^g and β^g) to compute the sample analog of the second diagonal matrix. This two-step estimator, using the estimate of the optimal matrix in the second step, has asymptotic variance:

$$V(B) = E[X^T Z] \cdot \left\{ \lambda \text{Diag}(\text{Var}_\ell(c_{it+1})) + \text{Diag} \left(\text{Var}_\ell \left(\sum_g (\alpha^g + \beta^g c_{it}) D_i^g \right) \right) \right\}^{-1} \cdot E[Z^T X]$$

This is the asymptotically efficient estimator if Y and X are split in two independent samples with common variables Z . Of course, there is a loss of efficiency compared to the optimal GMM that would be obtained if all variables were in the same sample.

The “imputation estimator” suggested by Klevmarken (1982) consists in running a linear first-stage in sample 1 to obtain $E[D_i^g | \{z_i^\ell\}]$ and $E[c_{it} D_i^g | \{z_i^\ell\}]$. In a second stage, it consists in regressing c_{it+1} (in sample 2) on the predicted values from the first-stage. It is numerically equivalent to regressing the last name average $\bar{y}_{\ell,t+1}$ on predicted values (weighted by last name size in sample 2) :

$$\bar{y}_{\ell,t+1} = \sum_g \left(\hat{\alpha}^g D_i^g + \hat{\beta}^g D_i^g \bar{y}_{g\ell,t} \right) + \eta_{\ell g}$$

The advantage of this estimator is that it is intuitively compelling and relates to the previous literature using last names to estimate intergenerational mobility. To our knowledge, all of the previous literature regresses last name averages on the previous period’s last name averages.

The asymptotic justification for the imputation estimator can be derived from the same exclusion restriction and the Law of Iterated Expectations. In the original exclusion restriction,

$$E \left[\left(c_{it+1} - \sum_{g=1}^G (\alpha^g + \beta^g c_{it}) D_i^g \right) z_i^\ell \right] = 0 \quad \forall \ell$$

Using the Law of Iterated Expectations, we can replace D_i^g with $E [D_i^g | \{z_i^\ell\}]$ and $D_i^g c_{it}$ with $E [D_i^g c_{it} | \{z_i^\ell\}]$:

$$E \left[\left(c_{it+1} - \sum_{g=1}^G (\alpha^g E [D_i^g | \{z_i^\ell\}] + \beta^g E [D_i^g c_{it} | \{z_i^\ell\}]) \right) z_i^\ell \right] = 0 \quad \forall \ell$$

The imputation estimator - also named TS2SLS - is easier to implement and Inoue and Solon (2010) prove that it has better finite sample properties because it corrects for differences in the empirical distributions of last names between the two samples.

B.5 Estimating Heterogeneity - Simulations and Placebo tests by size of Surnames

In this section we refer to the intersection of races and regions as “groups” and explore the variation in group specific housing consumption across surnames. The goal is to get a sense of whether the data allows consistent estimation of the group-specific parameters. The starting point is table 2.6 in the main text. Placebo tests show that our method of using last name as instruments to estimate group specific parameters then reconstruct the overall correlation does not compare well with the true individual-level correlations for a set of variables related to education, work and family. Table B.5 below focuses on education and provides more details. It compares the true OLS individual-level correlations to the estimated β^g for each group. We see that surname based estimates of heterogeneity tend to “blow up” compared to their individual level counterparts.

In order to understand whether the data contains enough variation in group specific log housing consumption to consistently estimate the heterogeneous parameters, we perform simulations. The simulations work as follow. First, we draw (with replacement) a sample of size N_1 from the 1940 Census - similar to what we would do for a bootstrap. From this sample, we generate the Y data

Table B.5: Comparison of True and Between last-name regression of Years of Education on Log Housing Consumption (Slope Coefficient by demographic group - 1940)

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.895	1.912	0.711	0.956	1.114	3.253	0.851	3.161	0.890	1.918
Black	0.923	5.233	0.744	1.231	0.648	0.678	0.949	2.841	0.914	1.555
Other	1.264	2.575	0.973	1.489	1.312	3.006	1.153	1.759	1.323	2.039
All	0.913	1.984	0.726	0.953	1.269	2.997	1.008	3.208	1.061	1.971

Columns:

(1) True OLS

(2) Between Last Names OLS - adjusted for covariates

(log housing consumption in 2012) by using the following model: $c_{it+1} = \alpha^g + \beta^g c_{it} + \varepsilon_{it}$ for known values of $\{\alpha^g\}$ and $\{\beta^g\}$.¹ We draw ε from a normal with mean zero and variance 1.4 (which is equal to the actual variance of log housing consumption in the 1940 data). Notice that because ε_{it} is drawn independently from α^g and β^g , the exclusion restriction in (2.6) holds. We then run the two sample surname level estimation. The Y matrix is the simulated data and the X matrix is the original 1940 Census data. Notice that, as explained above, we generate the Y data from a “bootstrapped” sample of X instead of the original X matrix. We do this in order to have two independent samples Y and X with common variables as opposed to having one sample of the same population containing the Y and X variables. We generate 200 different split samples and run the estimation in each of them. Table B.6 and B.7 report the mean and standard deviation of the estimates across the 200 estimations.

The estimated parameters are all close to their “true” value. Estimates are unbiased and standard errors are small. The simulation exercise suggests that there is enough variation in the data and that unbiased consistent estimation of $\{\alpha^g\}$ and $\{\beta^g\}$ is not hopeless. We further explore variations in the data by performing the same simulations with restricted samples: one that contains only last names with less than 50 male heads of household in 1940 (table B.8) and another sample

1. In the simulations, we simply use the estimated values presented in tables 8 and 9

Table B.6: Simulations: Estimates versus true parameters (IGC)

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67 .	0.67 (0.001)	0.2 .	0.2 (0.001)	0.245 .	0.245 (0.002)	0.096 .	0.096 (0.002)	0.311 .	0.317 (0.001)
Black	0.071 .	0.071 (0.017)	-0.093 .	-0.092 (0.017)	0.206 .	0.206 (0.005)	0.024 .	0.023 (0.035)	0.08 .	0.082 (0.004)
Other	-0.119 .	-0.118 (0.019)	0.209 .	0.209 (0.011)	0.254 .	0.256 (0.009)	0.497 .	0.497 (0.005)	0.368 .	0.368 (0.004)
All	0.654 .	0.655 (0.001)	0.202 .	0.203 (0.001)	0.247 .	0.246 (0.002)	0.129 .	0.13 (0.002)	0.308 .	0.314 (0.001)

Columns:

(1) True parameter

(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

Table B.7: Simulations: Estimates versus true parameters (α constant)

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	3.248 .	3.247 (0.009)	5.889 .	5.889 (0.008)	5.575 .	5.576 (0.009)	6.583 .	6.58 (0.014)	5.181 .	5.133 (0.004)
Black	6.44 .	6.439 (0.1)	6.037 .	6.03 (0.09)	5.679 .	5.679 (0.023)	5.618 .	5.621 (0.198)	5.789 .	5.793 (0.023)
Other	7.942 .	7.936 (0.117)	5.47 .	5.47 (0.051)	5.174 .	5.167 (0.04)	4.713 .	4.714 (0.023)	5.335 .	5.351 (0.023)
All	3.39 .	3.377 (0.009)	5.891 .	5.891 (0.008)	5.59 .	5.591 (0.008)	6.456 .	6.45 (0.013)	5.237 .	5.189 (0.004)

Columns:

(1) True parameter

(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

Table B.8: Simulations: Estimates versus true parameters (IGC) - rare names only

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67 .	0.67 (0.002)	0.2 .	0.2 (0.001)	0.245 .	0.245 (0.002)	0.096 .	0.096 (0.002)	0.311 .	0.343 (0.001)
Black	0.071 .	0.071 (0.016)	-0.093 .	-0.092 (0.016)	0.206 .	0.206 (0.005)	0.024 .	0.021 (0.038)	0.08 .	0.095 (0.004)
Other	-0.119 .	-0.119 (0.016)	0.209 .	0.209 (0.01)	0.254 .	0.254 (0.01)	0.497 .	0.497 (0.006)	0.368 .	0.364 (0.004)
All	0.654 .	0.658 (0.002)	0.202 .	0.203 (0.001)	0.247 .	0.251 (0.002)	0.129 .	0.143 (0.002)	0.308 .	0.341 (0.001)

Columns:

(1) True parameter

(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

containing only last names with more than 50 male heads of households (table B.9). There is not much of a difference between table B.6 and B.8. The estimates remain unbiased and standard errors are reasonable albeit slightly larger when restricting the sample to large surnames only (see table B.9).

Because the estimator performs well in simulations, they do not bring much insight into why the placebo tests fail when trying to estimate regional heterogeneity. We also look at the behavior of the estimates - in the placebo test - as we iteratively eliminate a larger and larger share of the most common last names. Figure B.3 and B.4 show the result of these placebo tests for education and occupational score (regression of years of education and occupational score on log housing consumption in 1940). The x-axis indicates the percentile of last name size included in the sample. For example, at 20, the plotted coefficients are the surname- and individual- level coefficients computed in a sample that contains only the last names under the 20th percentile of last name size distribution. We can see on the pictures that getting rid of the larger surnames help bring the “placebo estimates” closer to their target value but the decrease in bias is relatively slow as we move down the surname size distribution.

A potentially interesting avenue for future research would be to use regression trees or random

Table B.9: Simulations: Estimates versus true parameters (IGC) - Large names only

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67 .	0.67 (0.004)	0.2 .	0.2 (0.005)	0.245 .	0.245 (0.004)	0.096 .	0.095 (0.009)	0.311 .	0.31 (0.002)
Black	0.071 .	0.077 (0.079)	-0.093 .	-0.09 (0.067)	0.206 .	0.205 (0.018)	0.024 .	0.027 (0.212)	0.08 .	0.081 (0.014)
Other	-0.119 .	-0.116 (0.092)	0.209 .	0.213 (0.067)	0.254 .	0.253 (0.032)	0.497 .	0.497 (0.013)	0.368 .	0.368 (0.01)
All	0.654 .	0.654 (0.004)	0.202 .	0.202 (0.005)	0.247 .	0.246 (0.004)	0.129 .	0.125 (0.009)	0.308 .	0.307 (0.002)

Columns:

(1) True parameter

(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

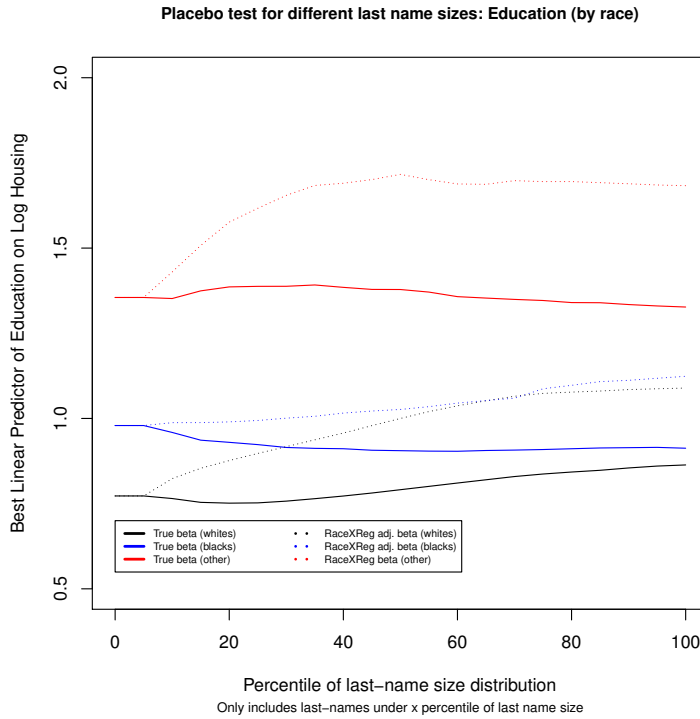


Figure B.3: Coefficient of regression of Education on log Housing Consumption - surname- versus individual-level by size of last name

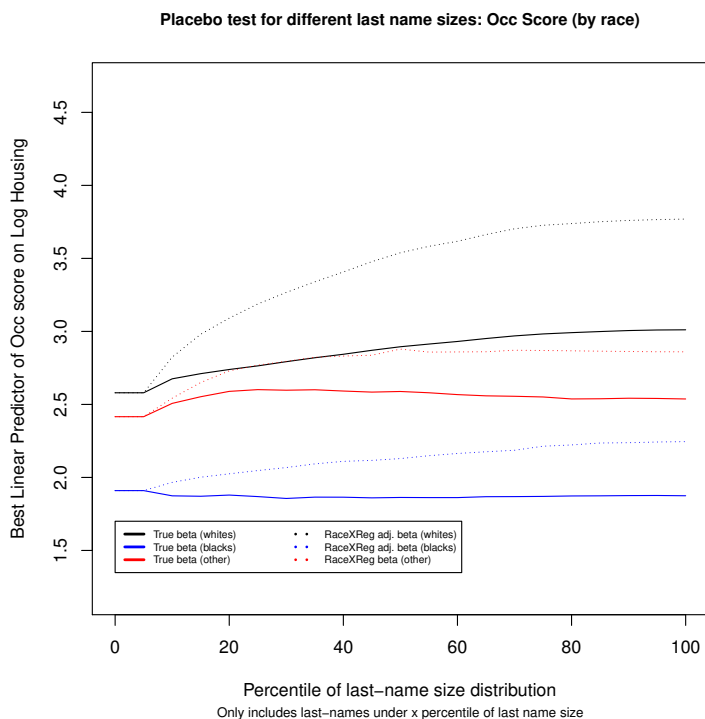


Figure B.4: Coefficient of regression of Occupational Score on log Housing Consumption - surname- versus individual-level by size of last name

forest types of methods from the machine learning literature to determine how to split the data into groups based on demographics.² The advantage of these methods is that it can divide the parameter space in a more efficient way when deciding on which covariates to adjust for. We leave this possibility for future research.

B.6 Derivation of OLS and IV coefficient when total consumption is proxied with c^*

We start with a few definitions:

². See Athey and Imbens (2016) and Athey and Wager (2017) for applications of trees and random forests in order to estimate heterogeneous treatment effects.

$$h_{it} = \gamma_{0t} + \gamma_{1t}c_{it} + \nu_{it}$$

$$h_{it+1} = \gamma_{0t+1} + \gamma_{1t+1}c_{it+1} + \nu_{it+1}$$

where h is log housing consumption, c is log total consumption and t and $t+1$ denote 1940 and 2012 respectively. We assume $cov(c_{it}, \nu_{it}) = 0$, $cov(c_{it+1}, \nu_{it+1}) = 0$ and $cov(c_{it}, \nu_{it+1}) = 0$. We form the following proxies for consumption in t and $t+1$:

$$c_{it}^* = -\frac{\gamma_{0t}}{\gamma_{1t}} + \frac{1}{\gamma_{1t}}h_{it}$$

$$c_{it+1}^* = -\frac{\gamma_{0t+1}}{\gamma_{1t+1}} + \frac{1}{\gamma_{1t+1}}h_{it+1}$$

Notice that $c_{it}^* = c_{it} + \frac{1}{\gamma_{1t}}\nu_{it}$ and $c_{it+1}^* = c_{it+1} + \frac{1}{\gamma_{1t+1}}\nu_{it+1}$. We also define a few correlation coefficients between variables of interest:

$$\rho_{\nu\epsilon} = \frac{cov(\nu_{it}, \epsilon_{it+1})}{\sqrt{var(\nu_{it})}\sqrt{var(\epsilon_{it+1})}}$$

$$\rho_{\nu\nu} = \frac{cov(\nu_{it}, \nu_{it+1})}{\sqrt{var(\nu_{it})}\sqrt{var(\nu_{it+1})}}$$

Let us assume that $0 \leq \rho_{\nu\nu} \leq 1$. The fact that $\rho_{\nu\nu}$ is potentially greater than 0 reflects the fact that the portion of housing that is not explained by total consumption may be partially correlated over generations (if taste for housing is hereditary for instance). We rule out data generating processes in which $\rho < 0$ as it seems implausible. Finally, when estimating the housing Engel curve equations in the CEX data, we obtain an estimate of:

$$R_{hc,t}^2 = \frac{\gamma_{1t}^2 \text{var}(c_{it})}{\text{var}(h_{it})}$$

$$R_{hc,t+1}^2 = \frac{\gamma_{1t+1}^2 \text{var}(c_{it+1})}{\text{var}(h_{it+1})}$$

B.6.1 Potential OLS bias

Running OLS is impossible in our dataset because we do not have individual linkages across time. For the sake of intuition, it is however useful to consider the probability limit of the OLS coefficient (in the intergenerational elasticity of consumption regression) when using housing consumption as a proxy. The probability limit of the OLS estimator is:

$$\begin{aligned} \beta_{OLS} &= \frac{\text{cov}(c_{it+1}^*, c_{it}^*)}{\text{var}(c_{it}^*)} \\ &= \frac{\text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}}\nu_{it}, \beta c_{it} + \varepsilon_{it+1} + \frac{1}{\gamma_{1t+1}}\nu_{it+1}\right)}{\text{var}(c_{it}^*)} \end{aligned}$$

By definition of ε_{it+1} being the prediction error in the population regression, $\text{cov}(c_{it}, \varepsilon_{it+1}) = 0$. Also, by assumptions on the Engel curve equations, $\text{cov}\left(c_{it}, \frac{1}{\gamma_{1t+1}}\nu_{it+1}\right) = 0$ and $\text{cov}\left(\frac{1}{\gamma_{1t}}\nu_{it}, \beta c_{it}\right) = 0$. As a result, we have:

$$\beta_{OLS} = \beta \frac{\text{var}(c_{it})}{\text{var}(c_{it}^*)} + \frac{\frac{1}{\gamma_{1t}} \text{cov}(\nu_{it}, \varepsilon_{it+1})}{\text{var}(c_{it}^*)} + \frac{\frac{1}{\gamma_{1t}} \frac{1}{\gamma_{1t+1}} \text{cov}(\nu_{it}, \nu_{it+1})}{\text{var}(c_{it}^*)}$$

Let us focus on the second term which can be re-arranged as:

$$\begin{aligned}
\frac{\frac{1}{\gamma_{1t}} \text{cov}(\nu_{it}, \varepsilon_{it+1})}{\text{var}(c_{it}^*)} &= \rho_{\nu\varepsilon} \sqrt{1 - R_{hc,t}^2} \sqrt{\frac{\text{var}(\varepsilon_{it+1})}{\text{var}(c_{it}^*)}} \\
&= \rho_{\nu\varepsilon} \sqrt{1 - R_{hc,t}^2} \sqrt{R_{hc,t}^2} \sqrt{\frac{\text{var}(\varepsilon_{it+1})}{\text{var}(c_{it})}}
\end{aligned}$$

Under the assumption of a stationary distribution of consumption, $\text{var}(c_{it}) = \text{var}(c_{it+1})$ and the term is equal to $\rho_{\nu\varepsilon} \sqrt{1 - R_{hc,t}^2} \sqrt{R_{hc,t}^2} \sqrt{1 - R_{cc}^2}$ where R_{cc}^2 is the R^2 of the regression of c_{it+1} on c_{it} . This expression underlies a specific threat to consistency of OLS estimates when using housing as a proxy: there may be a correlation between ν and ε . For instance, if housing is a vehicle for asset accumulation, families who own and consume a lot of housing relative to their total consumption (large ν_{it}) in 1940 may end up with a larger total consumption in 2012 than predicted by their 1940 total consumption. Notice however that we use both renters and owners in 1940 and there is presumably no reason to believe that high ν renters in 1940 would systematically have a larger ε in $t + 1$. From now on, we assume $\rho_{\nu\varepsilon} = 0$ but keep in mind that housing as a mechanism of asset accumulation may generate a bias. As made clear by the expression above, the size of the bias is lower for very large $R_{hc,t}^2$ (housing is such a strong proxy for total consumption that variance in ν is small).³ Under the assumption that $\rho_{\nu\varepsilon} = 0$:

$$\begin{aligned}
\beta_{OLS} &= \beta \frac{\text{var}(c_{it})}{\text{var}(c_{it}^*)} + \frac{\frac{1}{\gamma_{1t}} \frac{1}{\gamma_{1t+1}} \text{cov}(\nu_{it}, \nu_{it+1})}{\text{var}(c_{it}^*)} \\
&= \beta \frac{\gamma_{1t}^2 \text{var}(c_{it})}{\text{var}(h_{it})} + \rho_{\nu\nu} \frac{\frac{1}{\gamma_{1t+1}} \sqrt{\text{var}(\nu_{it})} \sqrt{\text{var}(\nu_{it+1})}}{\frac{1}{\gamma_{1t}} \text{var}(h_{it})} \\
&= \beta R_{hc,t}^2 + \rho_{\nu\nu} \frac{\gamma_{1t}}{\gamma_{1t+1}} \sqrt{1 - R_{hc,t}^2} \sqrt{1 - R_{hc,t+1}^2} \sqrt{\frac{\text{var}(h_{it+1})}{\text{var}(h_{it})}}
\end{aligned}$$

Let us assume that distributions are stationary (and Engel curves are time invariant) so that $R_{hc,t}^2 = R_{hc,t+1}^2 = R_{hc}^2$, $\text{var}(h_{it+1}) = \text{var}(h_{it})$ and $\gamma_{1t} = \gamma_{1t+1}$. We obtain:

3. The bias is also larger when the variance of ε is large relative to consumption - that is when 1940 actual consumption is not a strong predictor of 2012 consumption.

$$\beta_{OLS} = \beta R_{hc,t}^2 + \rho_{\nu\nu} (1 - R_{hc}^2)$$

This expression has two component. First, the use of a proxy for total consumption leads to attenuation bias because the proxy is a (classical) error ridden version of the true regressor. The attenuation bias stronger if the R-squared of the Engel curve regression is lower. Note that R_{hc}^2 can be estimated in the CEX data and one can therefore correct for attenuation bias in OLS. Second, intergenerational correlation in ν - preference for housing for instance - leads to a bias in the estimate. Using the assumption that $0 \leq \rho_{\nu\nu} \leq 1$ we can bound the size of this second term by $0 \leq \rho_{\nu\nu} (1 - R_{hc}^2) \leq 1 - R_{hc}^2$. As reported in figure 2, the R-squared of the Engel curve regression in the CEX data is 0.57.

B.6.2 Bounding the IV bias

In practice, we run an IV regression using last-names as instruments. In this section, we re-explore the potential bias arising from the use of housing consumption as a proxy when the estimator uses last-names as an instrument. We could follow the matrix notation of section 4, but using covariance formula seem to make the intuition more transparent. In the following reasoning, think of each variable (c, c^*, h, z^ℓ) as being de-meanded. In addition, if we are looking for a “group-specific” β^g instead of an overall β , we can think of these covariances as being over group g population only (and de-meanded by the group-specific mean). Under these assumptions, we can write the IV estimator $\hat{\beta}$ as the ratio of weighted sums of covariances.⁴

$$\hat{\beta} = \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it+1}^*, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}^*, z_i^{\ell})}$$

We can re-write this ratio of weighted covariances as:

4. The weight on each covariance - indexed by last name ℓ - depends on the specific covariance matrix in the GMM estimation or is a function of the distribution of last names in each sample for the Two Sample Two Stage Least Square estimator that we use in the estimations.

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it+1}^*, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}^*, z_i^{\ell})} \\
&= \frac{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it+1} + \frac{1}{\gamma_{1t+1}} \nu_{it+1}, z_i^{\ell}\right)}{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}} \nu_{it}, z_i^{\ell}\right)} \\
&= \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}} \nu_{it}, z_i^{\ell}\right)} \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it+1}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})} + \frac{\frac{1}{\gamma_{1t+1}} \sum_{\ell} \omega_{\ell} \text{cov}(\nu_{it+1}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}} \nu_{it}, z_i^{\ell}\right)} \\
&= \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}} \nu_{it}, z_i^{\ell}\right)} \beta + \frac{\frac{1}{\gamma_{1t+1}} \sum_{\ell} \omega_{\ell} \text{cov}(\nu_{it+1}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}\left(c_{it} + \frac{1}{\gamma_{1t}} \nu_{it}, z_i^{\ell}\right)}
\end{aligned}$$

We use the fact that $\frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it+1}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})} = \beta$ which follows from the assumptions made in section 4 and in particular the exclusion restrictions that guarantee consistency of the last-name level estimates (when using actual *total* consumption). Now assume that family “preferences” for housing evolve according to an AR(1) process:

$$\nu_{it+1} = \delta + \rho_{\nu\nu} \nu_{it} + \tilde{\nu}_{it+1}$$

with $0 \leq \rho_{\nu\nu} \leq 1$ and $\tilde{\nu}_{it+1}$ independent from ν_{it} and z_i^{ℓ} . In that case, we can re-write the above formula for $\hat{\beta}$ as:

$$\hat{\beta} = \Omega \beta + (1 - \Omega) \frac{\gamma_{1t}}{\gamma_{1t+1}} \rho_{\nu\nu}$$

$$\text{and } \Omega = \frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell}) + \frac{1}{\gamma_{1t}} \sum_{\ell} \omega_{\ell} \text{cov}(\nu_{it}, z_i^{\ell})}.$$

First of all, we see that if $\sum_{\ell} \text{cov}(\nu_{it}, z_i^{\ell}) = 0$, the estimator is consistent even if housing consumption was a noisy proxy for total consumption.⁵ That result is not surprising, it is well-known that valid instruments (in particular, instruments uncorrelated with the “noise” in the proxy variable) fix the attenuation bias in regressions plagued with classical measurement error. Notice that two (strong) assumptions were made to arrive at the formula. First, we rule out correlations be-

5. A sufficient condition would be $\text{cov}(\nu_{it}, z_i^{\ell}) \neq 0 \quad \forall \ell$.

tween ε_{it+1} and ν_{it} . This is implied by the assumptions made in section 4 and in particular by the assumption that $\frac{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it+1}, z_i^{\ell})}{\sum_{\ell} \omega_{\ell} \text{cov}(c_{it}, z_i^{\ell})} = \beta$. If last-names were correlated with ν_{it} , which in turn would be correlated with ε_{it+1} , exclusion restriction (2.6) would fail (the population project error would be correlated with last-names). In the above reasoning, we have also ruled out correlation between ν_{it+1} and z_i^{ℓ} .

In case last names are correlated with preference for housing in 1940, the estimate recovers a weighted sum between the true intergenerational elasticity of consumption, β , and the intergenerational persistence in housing preferences $\rho_{\nu\nu}$. If there is no persistence in preferences, then the estimator is biased downward while it is biased upward if persistence in preferences is higher than elasticity of consumption. The weight on the intergenerational elasticity of consumption is proportional to the R-squared of the Engel curve regression. Intuitively, if housing consumption is a strong proxy, in the sense that the Engel curve regression in the CEX has a large R-squared, then correlation in housing preferences across generations matter less.

B.7 Estimates of Propensity Scores and Inverse Mills Ratio

We aim to estimate the conditional probability that descendants of each head of household in 1940 owns a home in 2012 (and as a result show up in our DataQuick sample). Recall that we assume:

$$B_{i,t+1} = 1 \Leftrightarrow B_{i,t+1}^* = \tilde{X}_{it}\Gamma + \tilde{\eta}_{i,t+1} \geq 0$$

where \tilde{X}_{it} is a vector of variables, in the 1940 sample, that includes at least c_{it} and the constant 1 - both potentially interacted with D_i^g . We further assume joint normality of the error terms:

$$\begin{pmatrix} \tilde{\varepsilon}_{i,t+1} \\ \tilde{\eta}_{i,t+1} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\varepsilon}^2 & \sigma_{\varepsilon\tilde{\eta}} \\ \sigma_{\varepsilon\tilde{\eta}} & \sigma_{\tilde{\eta}}^2 \end{pmatrix} \right)$$

If we observed individual intergenerational linkages between observations of the two samples, we would simply estimate a Probit model. To overcome the lack of individual intergenerational linkages, we use - again - last names as instruments. The exclusion restriction is:

$$E \left[\tilde{\eta}_{i,t+1} z_i^\ell \right] = 0 \Leftrightarrow$$

$$E \left[\left(B_{i,t+1} - \Phi \left(\tilde{X}_i \Gamma \right) \right) z_i^\ell \right] = 0$$

We minimize a quadratic form in the sample analog of this exclusion restriction. The sample analog of $E \left[B_{i,t+1} z_i^\ell \right]$ is formed using the Census count of individuals with “common last names” and a count, for each of those surnames of the numbers of individuals owning a primary residence in the DataQuick sample. We note here that this measure is not perfect. The Census counts indicate the total number of individuals with a given last names (it only includes last names with at least 100 members in 2010) irrespective of their age and status within the household. Ideally, we would like to use the number of heads of households, within each last name, - or only the adult population - as the denominator in order to obtain a rate of home ownership. Notice that this should not be a problem if the difference between the “ideal” homeownership rate and our calculated rate does not systematically vary across the income distribution. The propensity scores are always used in estimation as a ratio - in the inverse Mills ratio and in the “re-weighting” of the 1940 sample by $\frac{\Phi(\tilde{X}_i \hat{\Gamma})}{B}$ - so that adjusting the propensity scores by a constant factor would not change the results. Coming back to the objective function. Denoting the Census count of individuals for each last name in 2010 by $\{N_\ell^{Census}\}$ and the number of individuals owning their primary residence in the Dataquick sample by $\{N_\ell^{Dataquick}\}$, the sample analog of $E \left[B_{i,t+1} z_i^\ell \right]$ is $\frac{1}{\sum_\ell N_\ell^{Census}} \sum_\ell N_\ell^{Dataquick}$.

The sample analog of $E \left[\Phi \left(\tilde{X}_i \Gamma \right) z_i^\ell \right]$ is:

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \Phi \left(\tilde{X}_i \Gamma \right) z_i^\ell$$

Define $G(\Gamma)$ a $L \times 1$ vector whose i^{th} element is $\Phi \left(\tilde{X}_i \Gamma \right)$. We form a 2SIV estimator by minimizing a quadratic in the sample analogs to the exclusion restrictions.