



# DISSERTATION

*Submitted in partial fulfillment of the requirements  
for the degree of*

**DOCTOR OF PHILOSOPHY  
ECONOMICS**

*Titled*

**“ESSAYS ON PUBLIC FINANCE AND AUCTION THEORY”**

*Presented by*  
**Musab Murat Kurnaz**

*Accepted by*

Christopher Sleet

4/22/16

---

**Chair: Prof. Christopher Sleet**

---

**Date**

*Approved by The Dean*

Robert M. Dammon

5/5/16

---

**Dean Robert M. Dammon**

---

**Date**

DOCTORAL DISSERTATION

# **Essays on Public Finance and Auction Theory**

by

**Musab Murat Kurnaz**

April 8, 2016

Submitted to the Tepper School of Business  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
at

**Carnegie Mellon University**

Dissertation Committee:

Professor Christopher Sleet  
Associate Professor Laurence Ales  
Associate Professor Isa Hafalır  
Assistant Professor Hakki Yazıcı

## Abstracts

This dissertation contains three chapters and focuses on the optimal design of fiscal policy and multi-item auction mechanisms where there is an informational friction in the economy.

In the first chapter, I examine the optimal taxation of families in an environment in which (i) the earning abilities and child tastes of parents are private information, and (ii) child-rearing requires both parental time and goods. The optimal tax system combines an income tax schedule for childless families with tax credits for families with children. These components insure parents against low income and high taste for children draws respectively. The parental time and cost of goods involved in child-rearing have distinct impacts on the shape of optimal child tax credits. In the quantitative part, I estimate these costs and show that they translate into a pattern of optimal credits that is U-shaped in income. As a result, the credit to one (two) child families is decreasing over the first 40% (50%) of the income distribution. In addition, the credit for the second child is not equal to the credit for the first, owing to economies of scale in child-rearing. For median-income families, the credit for the second child equals 44% of the credit for the first child. Finally, I offer a simple linear-income dependent credit policy that achieves most of the welfare gain from the optimum.

In the second chapter (joint with Laurence Ales and Christopher Sleet), we consider the normative implications of technical change for tax policy design. A task-to-talent assignment model of the labor market is embedded into an optimal tax problem. Technical change modifies equilibrium wage growth across talents and the substitutability of talents across tasks. The overall optimal policy response is to reduce marginal income taxes on low to middle incomes, while raising those on middle to high incomes. The reform favors those in the middle of the income distribution, reducing their average taxes while lowering transfers to those at the bottom.

In the third chapter (joint with Isa Hafalır), we consider multi-unit discriminatory auctions where ex-ante symmetric bidders have single unit demands and resale is allowed after the bidding stage. When bidders use the optimal auction to sell the items in the resale stage, the equilibrium without resale is not equilibrium. We find a symmetric and monotone equilibrium when there are two units for sale, and, interestingly, show that there may not be a symmetric and monotone equilibrium if there are more than two units.

## Acknowledgments

I would like to express my profound gratitude to my advisor, Christopher Sleet. I have received a lot of support and encouragement from him during many “nonlinear” processes of research. He is in the group of great public finance economists in the world, where I want to be in my future career. In addition, I am deeply grateful to Laurence Ales. His constructive comments and helpful suggestions improved my work significantly. I also thank Isa Hafalır who provided constantly generous support. I am also indebted to Hakkı Yazıcı for his comments on my thesis.

I would like to thank my friends and colleagues at Tepper Ph.D. program too. Through talks and discussions we had, I have been exposed to various perspectives in economics.

I would like to express my special thanks to my parents, my brother, and my grandmother for unconditionally supporting my long schooling. Lastly, I deeply appreciate my wife Leman Kur-naz for believing in me and consistently encouraging me during the Ph.D. study. Without the support of her, I could have never finished my Ph.D.

# Contents

<b>1</b>	<b>Optimal Taxation of Families</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Institutional Background . . . . .	6
1.2.1	Child Tax Credit . . . . .	6
1.2.2	Earned Income Tax Credit . . . . .	7
1.2.3	Child and Dependent Care Tax Credit . . . . .	8
1.3	Model . . . . .	9
1.3.1	Family Problem . . . . .	10
1.3.2	The Government's Problem . . . . .	11
1.4	Quantitative Analysis . . . . .	18
1.4.1	Sample Selection . . . . .	18
1.4.2	Parental Time . . . . .	20
1.4.3	Cost of Goods . . . . .	21
1.4.4	Estimation of the Distribution of Earning Abilities . . . . .	22
1.4.5	Estimation of the Distribution of Child Taste . . . . .	22
1.4.6	Deriving the Optimal Tax System . . . . .	24
1.4.7	Proposal . . . . .	29
1.5	Robustness . . . . .	29
1.5.1	Age Analysis . . . . .	29
1.5.2	Detailed Cost Analysis . . . . .	31
1.5.3	Marital Status Analysis . . . . .	32
1.6	Conclusion . . . . .	32

<b>2</b>	<b>Technical Change, Wage Inequality and Taxes</b>	<b>35</b>
2.1	Introduction . . . . .	35
2.2	Evolution of the Occupational Wage Distribution: Stylized Facts . . . . .	41
2.3	Taxation with Imperfectly Substitutable Workers . . . . .	43
2.3.1	Physical Environment . . . . .	43
2.3.2	Optimal Policy . . . . .	45
2.4	Taxation, Assignment and Technical Change . . . . .	51
2.4.1	Physical Environment . . . . .	51
2.4.2	Tax Equilibria and the Government's Policy Problem . . . . .	53
2.4.3	An Example: Technical Change in the Two Talent Model . . . . .	57
2.4.4	Comparison to <b>Rothschild and Scheuer, 2013</b> . . . . .	60
2.5	Measuring Technical Change . . . . .	61
2.5.1	Determining Types and Tasks . . . . .	62
2.5.2	Estimating $b$ . . . . .	63
2.5.3	Estimating $a$ . . . . .	65
2.6	Quantitative Implications for Policy . . . . .	67
2.6.1	Selection of Other Parameters and Computational Method . . . . .	67
2.6.2	Optimal Tax Results . . . . .	68
2.7	Extension: INTRA-TASK WAGE DISPERSION . . . . .	72
2.8	Conclusion . . . . .	74
<b>3</b>	<b>Discriminatory Auctions with Resale</b>	<b>77</b>
3.1	Introduction . . . . .	77
3.2	Model . . . . .	80
3.2.1	Motivating Example . . . . .	81
3.3	Main Results . . . . .	83
3.3.1	When there are 2 units for sale . . . . .	84
3.3.2	When there are 3 or more units for sale . . . . .	88
3.4	Variations . . . . .	92

3.5	Conclusion . . . . .	95
<b>A</b>	<b>Appendix for Chapter 1</b>	<b>97</b>
A.0.1	Mechanism Design: Two-Dimensional Private Information . . . . .	97
A.0.2	Proof of Proposition 1 . . . . .	100
A.0.3	Family Income Elasticity . . . . .	102
A.0.4	Evolution of $b_n$ . . . . .	103
A.0.5	Figures and Tables . . . . .	106
<b>B</b>	<b>Appendix for Chapter 2</b>	<b>111</b>
B.1	Mechanism Design Formulation . . . . .	111
B.2	Proofs from Section 2.4 . . . . .	114
B.3	Continuous Talent-Continuous Task Model . . . . .	119
B.4	Empirical Implementation . . . . .	124
B.4.1	Data Set and Sample Selection . . . . .	124
B.4.2	Beyond Cobb-Douglas . . . . .	126
B.4.3	Occupational Ordering . . . . .	128
B.4.4	The demand for complex skills across occupations . . . . .	129
B.4.5	Alternative Productivity Function . . . . .	132
B.5	Counterfactuals . . . . .	133
B.6	Optimal policy under a Rawlsian objective . . . . .	136
B.7	A Model with Intra-occupational Wage Dispersion and a Heavy-tailed Talent Dis- tribution . . . . .	137
B.7.1	A general framework . . . . .	138
B.7.2	Reformulation of the general framework . . . . .	139
B.8	Intra-occupational task variation . . . . .	147
<b>C</b>	<b>Appendix for Chapter 3</b>	<b>149</b>
C.1	Appendix . . . . .	149

# List of Tables

1.1	Time Devoted to Market and Childcare . . . . .	20
1.2	Expenditures on child-rearing . . . . .	21
1.3	Estimation of Child Taste Distribution . . . . .	24
1.4	Credit Rates in Phase In and Phase Out . . . . .	29
2.1	Estimation of Productivity Function. . . . .	67
2.2	Optimal Tax Rates on Real Labor Income. . . . .	68
A.1	Earned Income Tax Credit Phase in and Phase out regions for 2014 . . . . .	110
B.1	Slope of the change of skill/ability intensity index over occupation . . . . .	131
B.2	Optimal Tax Rates on Real Labor Income, Alternate Case . . . . .	133
B.3	Optimal Tax Rates on Real Labor Income: Rawlsian Case. . . . .	137
B.4	Optimal Tax Rates on Real Labor Income. . . . .	145



# List of Figures

1.1	Earned Income Tax Credit in 2014 . . . . .	7
1.2	Government Spending for EITC and CTC . . . . .	8
1.3	Critical Child Taste Levels . . . . .	11
1.4	Increase in $T_1(\theta)$ by $dT$ . . . . .	14
1.5	Changes in Critical Child Taste Values . . . . .	16
1.6	Income-Fertility Relation . . . . .	20
1.7	Earning Ability Distribution . . . . .	22
1.8	Data: $V_n(\theta_j)$ , $\pi_n(\theta_j)$ , and $n(\theta_j)$ . . . . .	23
1.9	Probability distribution of child taste: $p(\beta)$ . . . . .	24
1.10	Current Tax System . . . . .	25
1.11	Optimum: $V_n(\theta_j)$ , $\pi_n(\theta_j)$ , and $n(\theta_j)$ . . . . .	26
1.12	Optimal Tax System . . . . .	27
1.13	Proposal Tax Credits . . . . .	30
1.14	Tax Credits: Age Analysis . . . . .	30
1.15	Tax Credits: Costs Analysis . . . . .	31
1.16	Tax Credits: Marital Status Analysis . . . . .	32
2.1	Evolution of average income by occupation over time. . . . .	42
2.2	Evolution in size of employment by occupation over time. . . . .	43
2.3	Evolution of $\log(b(v))$ across decades. . . . .	64
2.4	Occupations and $v$ . . . . .	65
2.5	LOWESS smoothed $\log(\text{wages})$ over talents. . . . .	66

2.6	Decomposing taxes. Solid curve: Mirrleesian tax, $\tau^M$ . Dashed curve: overall tax, $\tau^*$ .	69
2.7	Mirrlees term decomposition.	70
2.8	Evolution of the wage compression term.	71
2.9	Effort elasticity of tasks.	72
A.1	Evolution of the ratio of female income for different income groups	103
A.2	Evolution of $b_n$ for different income groups	104
A.3	Current Tax System	106
A.4	Optimal Tax System	107
A.5	Proposed Tax Credits	108
A.6	Tax Credits: Age Analysis	108
A.7	Tax Credits: Costs Analysis	109
A.8	Tax Credits: Marital Status Analysis	109
B.1	Estimates on entire sample.	125
B.2	Estimates of $b(v)$ for different $\varepsilon$ values.	127
B.3	Impact of changes in $\varepsilon$ on $b$ estimates and optimal taxes.	127
B.4	Rank order over decade.	128
B.5	Rank evolution of selected skills and abilities intensity across occupations. LOWESS smoothed.	130
B.6	Principal component analysis of the rank evolution of skill and ability intensity across occupations.	132
B.7	Relative changes in $\tilde{k}_v^*$ from the 1970s to the 2000s	134
B.8	Relative changes in task assignment $\tilde{\sigma}^*$ induced by the shift in $b$ from 1970 to 2000.	134
B.9	Equilibrium wage growth $\Delta w^* / w^*$ for the parameter combinations $(a_{70}, b_{70})$ , $(a_{70}, b_{00})$ , $(a_{00}, b_{70})$ and $(a_{00}, b_{00})$ .	135
B.10	Marginal tax effects.	136
B.11	Change in Marginal Taxes.	147
B.12	Average Leik's ordinal variation (LOV) index concerning the knowledge requirements of occupations.	148

# Chapter 1

## Optimal Taxation of Families

### 1.1 Introduction

How many children to have is an important decision for parents. The number affects child-rearing which requires significant resources of goods and time. These resources interact with parents' labor decision, and consequently affect family income. A large positive literature studies this interaction extensively. However, normative work exploring the policy implication of such interaction is sparse. Almost all governments consider the impact of child-rearing costs and provide some benefits for parents. However, there is no consensus on how these benefits should be structured. For example, the UK government has proposed cutting the benefits for third children born after 2017, while the US has gradually increased the credit-per-child rate between 2000 and 2010. These facts indicate the importance of family taxation and motivate the following questions: How should the government optimally tax families? Should child tax credits be part of an optimal tax system? What are the key forces shaping the credits? What are the quantitative implications of these forces for the US economy?

This paper focuses on the optimal design of income taxes and child tax credits. I make both theoretical and quantitative contributions. On the theoretical side, I explore the forces shaping optimal income taxes and child tax credits. The former is redistributive towards low-earning families. The latter reduces the income tax liabilities of those with children who are made monetarily worse off by child-rearing. On the quantitative side, I study the key forces behind the credits. While the goods cost reduces the welfare of low-earning families more relative to the high, the time cost re-

duces the welfare of the high-earning families more relative to the low. These impacts suggest that the goods cost is a motive for more provision to poor families, on the other hand, the time cost is a motive for more provision to the wealthier. As a result, the optimal child tax credits are U-shaped with respect to income.

I study a Mirrleesian environment in which families face shocks on earning abilities and tastes for children and decide how much income to generate and how many children to have. A higher ability shock decreases the cost of generating income and a higher taste for children increases the desire to have more children. Both shocks are families' private information. Facing a problem of asymmetric information, the redistributive government maximizes social welfare by choosing labor income taxes and child tax credits. Optimal taxes are characterized by a formula which links marginal income tax rates to the exogenous ability distribution, the redistributive motives of the government, and the sensitivity of family income to taxes. In addition, the formula has two *novel* terms introduced by the child choice. The first term measures the prevalence of different family sizes. This measure provides information on parents' underlying child tastes. The second term is the tax differences for families with  $n$  children and with  $n + 1$  children. This term is a reflection of the motive for redistributing to families with children whose wealth is reduced by child-rearing costs.

Since earning abilities and tastes for children are both private information, the government faces an informational friction along two dimensions. The two-dimensional friction creates some technical issues. Because of the issues, the literature on optimal taxation dealing with multidimensional screening is sparse. In this paper, I handle such issues by assuming that family welfare is separable in the shocks. The separability assumption facilitates the family problem in which families generate income after determining family size. The number of children to have is determined by an analysis on the marginal cost and the benefit of children. The benefit is purely driven by the tastes while the cost is measured by the impact of child-rearing on family consumption and income, and consequently on family welfare. Under the separability assumption, for a given family size, optimal consumption and income depend only on the families' earning abilities. Therefore, the marginal cost of a child is independent of families' tastes for children. The child tastes that equate the marginal benefit and the marginal cost of children are defined as *threshold* tastes. Using this definition, the two-dimensional friction is resolved by a pair of incentive constraints. Given

a family size, one in the pair prevents mimicking the earnings of other families. The other of the pair assures that the given family size is optimal.

The threshold tastes are a crucial concept in this paper. These thresholds provide a rationale for the tax difference terms in the optimal tax system. To grasp the intuition behind the terms, consider two families with same earning ability who would choose to have one child given a tax system. In addition, assume that their tastes for children are on the distinct thresholds. This implies that one family is indifferent between zero children and one child, while the other is indifferent between one child and two children. If the government raises the taxes of one-child families by a small amount, these families would be better off with zero and two children, respectively. As a result, these families would change their size and their new tax liabilities would depend on their new sizes. The differences between the new and the old liabilities would affect the total tax revenue, and hence the differences should be considered in an optimal tax system.

I calibrate my model to the US economy and quantitatively analyze the optimal tax system. First, I calculate families' earning abilities using the first-order conditions of their problem and the information about their income and tax brackets, which are taken from the March release of the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics. Using the weights of families provided by data, I derive the earning ability distribution. Second, I assume a particular distribution for child tastes and use maximum likelihood estimation to capture its parameters. To the best of my knowledge, few works estimate the distribution of tastes for children, and my paper is one of the first attempts to derive such a distribution.

Optimal child tax credits are shaped by child-rearing costs. I estimate the goods and time costs of child-rearing. These costs have distinct impacts on the shape of credits. On the one hand, the goods cost decreases the welfare of low-income families relatively more than the high. On the other, the time cost is more dominant for high-income families. These costs push the child credits up for low and high-income families, respectively. As a result, the optimal credits are U-shaped. The credit to one (two) child families decreases in the first 40% (50%) of the income distribution and increases in the rest. In contrast, the child tax credits in the US are constant for families with earnings less than a threshold level and decrease slowly after that level. The shape of the US child tax credit over income seems that the government focuses only on the impact of goods cost.

Considering the impact of time cost and shaping credits according to both impacts may improve social welfare.

In addition, I show that the optimal credits are not same for each child in a family because of economies of scale in child-rearing. The ratio of time costs of two children and one child is 1.55. In addition, goods cost of two children is 66% more than the goods cost of one child. Because of the scale, the credit to the second child is less than the first child for all families. In particular, the credit for the second child is 44% of the first child credit for the median income families. In contrast, the child tax credits in US are constant for each child.

I evaluate the potential welfare gain from implementing the optimum. First, the welfare gain from the optimum relative to the current tax system is 1.1% in terms of equivalent increase in consumption for all families. Next, I propose a tax system in which the income taxes are based on optimal taxes of childless families and the child tax credits are linear with respect to income. This proposal captures 87% of the welfare gain attained by the optimum. Another tax system, in which income taxes are same as in my proposal but the credits are constant and equal per child, can reach only 70% of the welfare gain. This suggests that income-dependent credits can improve social welfare significantly.

The remainder of the paper is organized as follows. After a brief review of the literature, I provide an institutional background for taxation of families in Section 1.2. I introduce the model in Section 1.3. I derive the optimal tax schedule and also show why the conventional tax formula should be adjusted with new terms. Section 1.4 quantitatively analyzes the model. I check the robustness of results in Section 1.5 and conclude with Section 1.6.

**Related Literature:** This paper links the literature on fertility theories to public finance literature. Most of the public finance literature abstracts from the child decision and the majority of work in the fertility theories abstracts from optimal taxation. My paper fills this gap.

There is a vast literature on fertility theories. The related works to my paper study well-known empirical evidence that fertility is negatively correlated with income. [Schultz \(1986\)](#) is an example of such works, which explicitly focuses on wages of spouses and relates the evidence with the changes in wage gap. Recently, [Jones, Schoonbroodt, and Tertilt \(2010\)](#) give a brilliant summary of fertility theories. They state that child-rearing costs are on the focus of many studies. Empirically,

Haveman and Wolfe (1995) work on the goods and the time costs. They find that these costs incurred by parents and the government are around 14.5% of 1992 GDP. Two-thirds of the costs is financed by parents. In addition, 82% of parental costs is goods cost which includes expenditures on food, housing, health care, and clothes. There are many studies that work on the impact of the goods cost. Golosov, Jones, and Tertilt (2007) and Hosseini, Jones, and Shourideh (2013) are recent examples. The former paper studies the efficiency of the future allocations and the latter focuses on the consumption inequality in the long run.

The second cost of child-rearing is parental time. Jones et al. (2010) state that parental time is a crucial ingredient to explain the negative correlation between fertility and labor income. This is mainly because the opportunity cost of time devoted to child-rearing is higher for the high-wage workers. As a result, high wage workers produce fewer children. In addition, time cost increases the labor sensitivity of parents to the wage changes which is a major component in the optimal tax system. Blundell, Meghir, and Neves (1993) estimate that married families with children have higher Frisch elasticity than married families without children. In this paper, I endogenize the income elasticity of parents and show that parents with more children have higher elasticity. This is mainly because more children requires more time and reduces available time for labor.

The child-rearing costs are important ingredients in my model. In contrast with many works, I study with both costs and show that their interaction with different family income levels is important to shape optimal policies.

My paper also contributes to the public finance literature, which is based on the trade-off between efficiency and equity. The trade-off arises because agents' earning abilities are their private information. In my paper, not only earning abilities but also tastes for children are families' private information, and hence the friction in the information is two dimensional. Because of the technical difficulties of multi screening problems, there are few works study such an environment.<sup>1</sup> Kleven, Kreiner, and Saez (2009) and Jacquet, Lehmann, and der Linden (2013) are notable exceptions. The former focuses on the jointness of family taxation in which primary earner's earning ability and secondary earner's work cost are families' private information. They show that marginal income tax rates of the primary earner should be smaller if his or her spouse works. The latter studies an

---

<sup>1</sup>Baron and Myerson (1982) and Rochet and Chone (1998) provide some additional requirements to solve a multi screening problem in the industrial organization literature.

environment in which workers have private information regarding their earning abilities and their taste of work and make a labor decision extensively and intensively. They provide a rationale for non-negative marginal rates. Unlike the studies above, I focus not only on the marginal rates but also on tax liabilities of families to study child tax credits. Moreover, both studies have only two categories of agents. In this paper, I derive optimal taxes for an arbitrary number of family sizes.

The other contribution of this paper to the public finance literature is to the studies which “tags” agents. In an interesting work, [Mankiw and Weinzierl \(2010\)](#) study optimal income taxation by considering agents’ heights. They notify that the income distribution of a particular height group is an informative tool for the government. One can consider that families are tagged according to their number of children in my paper. If children exogenously appeared in a family, the optimal tax formula would be very similar to that of [Mankiw and Weinzierl \(2010\)](#). However, the number of children to have is a choice in real life, and hence the tax formulas in the literature are not applicable.

In the next section, I briefly state the US government-oriented welfare programs.

## 1.2 Institutional Background

There are around 80 mean-tested federal programs providing for different needs of families such as cash, food, housing, medical care, and social services. Almost 50% of the budget for welfare programs is spent for families with children.<sup>2</sup> In this section, I give information about some of the cash programs: Child Tax Credit, Earned Income Tax Credit, and Child and Dependent Care Tax Credit. These are the main cash assistance programs provided to families with children.

### 1.2.1 Child Tax Credit

The Child Tax Credit (CTC) was enacted as a temporary provision in the Taxpayer Relief Act of 1997. A credit of \$400 is given to families for each qualifying children and the credit was refundable only for families with more than two children.<sup>3</sup> The credit has gradually increased to \$1,000 from 2001 to 2010 by the Economic Growth and Tax Relief Reconciliation Act of 2001. Moreover, the refundability is extended to all families. This refundable tax credit is called Additional Child

---

<sup>2</sup>Refer to Chart 3 of <http://budget.house.gov/uploadedfiles/rectortestimony04172012.pdf>

<sup>3</sup>See <http://www.irs.gov/uac/Ten-Facts-about-the-Child-Tax-Credit> for eligibility tests.



Tax Credit. If a family has less tax liability than their child tax credit, they may get the minimum of unclaimed credits and 15% of their income above \$3,000. Because of the changes in the eligibility conditions and the credit amount, the federal spending for CTC increased from 0.2% to 0.4% of GDP between 2000 and 2010. Currently, the credit decreases for high income families. For example, the credit is reduced by \$50 for each \$1,000 when aggregate gross income is above \$110,000 for married tax payers filing jointly. Finally, the credit has become permanent by the American Taxpayer Relief Act of 2012.

### 1.2.2 Earned Income Tax Credit

The Earned Income Tax Credit (EITC) is another program for working families. The literature on the EITC is voluminous and cannot be fully reviewed here. I refer to [Hotz and Scholz \(2003\)](#), and the references there. Here, I focus on how the credit differs with family size. The maximum credit and phase in and out rates drastically change with the number of children in families (see Table A.1 in Appendix A.0.5). Figure 1.1 plots the EITC for 2014. Families with more children are given more credits.

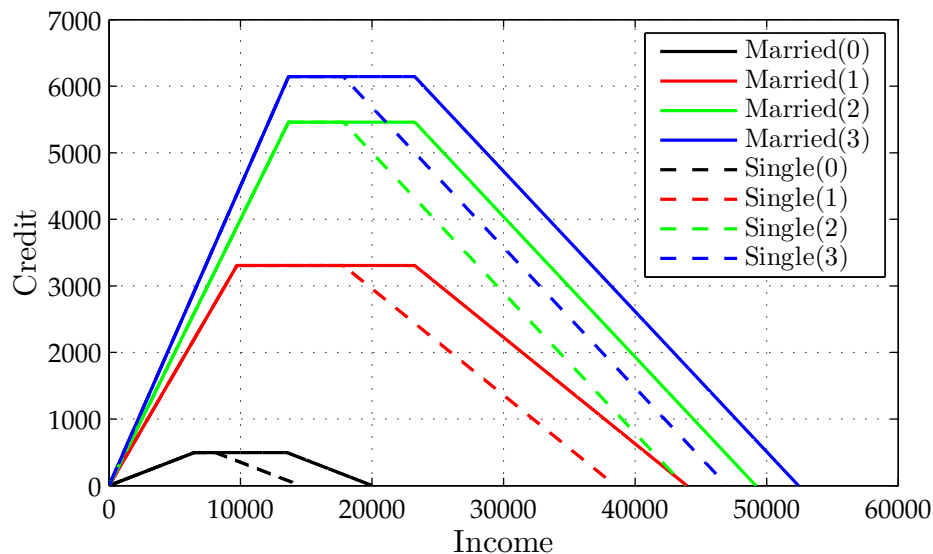


Figure 1.1: Earned Income Tax Credit in 2014  
Numbers in parenthesis represent the number of children in the family.

### 1.2.3 Child and Dependent Care Tax Credit

The Child and Dependent Care Tax Credit (CDCTC) program decreases the tax liability of families by 20% to 35% of child care expenditures for a qualifying child up to \$3,000 for up to two children. Also, \$5,000 from the salary can be excluded from adjusted gross income for child care if certain regulations are satisfied. The credit is non-refundable, and hence many low-income families do not participate in this program.<sup>4</sup> I refer to Blau (2003) for the history and effectiveness of the program.

To conclude this section, I focus on the functionality of these welfare programs. According to the Tax Policy Center, which is a joint venture of the Urban Institute and the Brookings Institution, 6.6 million families are qualified for the CDCTC in 2010,<sup>5</sup> more than 26 million taxpayers received the EITC in 2015,<sup>6</sup> and 38 million families claimed from the CTC in 2013.<sup>7</sup> More families benefit from the CTC because its eligibility requirement is more relaxed than the other welfare programs. Moreover, the CTC has become one of the most expensive welfare programs for the US government (see Figure 1.2).

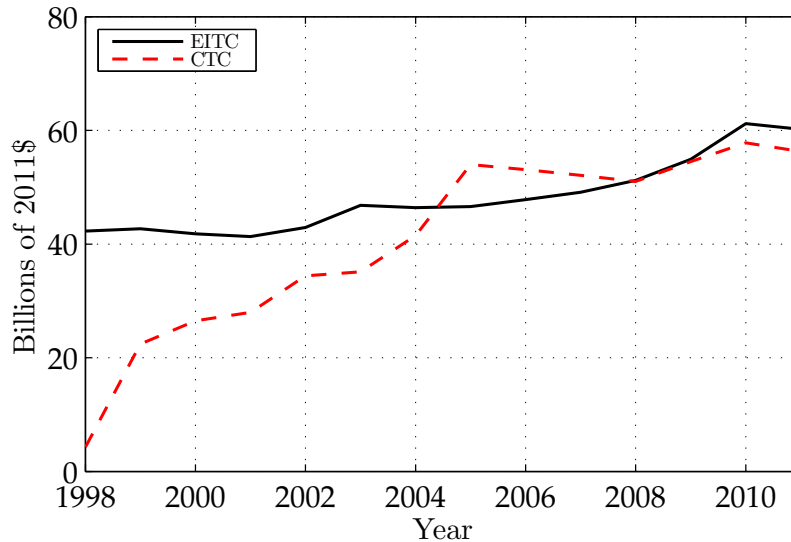


Figure 1.2: Government Spending for EITC and CTC

Source: [http://www.taxpolicycenter.org/taxfacts/Content/PDF/eitc\\_child\\_historical.pdf](http://www.taxpolicycenter.org/taxfacts/Content/PDF/eitc_child_historical.pdf).

In the following section, I introduce a static model, in which families face shocks on earning abilities and tastes for children and simultaneously decide how much income to generate and how

<sup>4</sup>See <http://www.irs.gov/taxtopics/tc602.html> for more details.

<sup>5</sup>See <http://taxpolicycenter.org/numbers/displayatab.cfm?Docid=2616>.

<sup>6</sup>See <http://www.taxpolicycenter.org/briefing-book/key-elements/family/eitc.cfm>

<sup>7</sup>See <http://www.taxpolicycenter.org/briefing-book/key-elements/family/ctc.cfm>.

many children to have. Child-rearing requires both goods and parental time, and these costs shape the optimal child tax credits.

### 1.3 Model

I consider a population of families, where the size is normalized to 1 and families have identical preferences over consumption  $c \in \mathbb{R}_+$ , earnings  $z \in [0, \bar{z}]$ , and number of children  $n \in \{0, 1, \dots, N\}$ . The families are characterized by two channels. Each family has an earning ability  $\theta$  distributed on  $(\underline{\theta}, \bar{\theta})$  in the population. The ability decreases the cost of earning  $z$  reciprocally:  $\frac{z}{\theta}$ . The second characterization is on the taste for children:  $\beta \sim (\underline{\beta}, \bar{\beta})$ . The benefit of  $n$  children,  $m(n, \beta)$ , separately increases the utility of the family. On the other hand, rearing  $n$  children requires goods (expenditures),  $e_n$ , and a fraction of parental time,  $b_n$ .

The family characteristics  $(\beta, \theta)$  are distributed according to a continuous density distribution over  $\mathcal{B} \times \Theta = [\underline{\beta}, \bar{\beta}] \times [\underline{\theta}, \bar{\theta}]$ . Let  $\Pi(\beta, \theta)$  be the cumulative distribution. I denote by  $P(\beta|\theta)$  the cumulative distribution of  $\beta$  conditional on  $\theta$ :  $\Pi(\beta, \theta) = \int_{\Theta} P(\beta|\theta) f(\theta) d\theta$  where  $f(\theta)$  is the unconditional distribution of  $\theta$ . Both  $\beta$  and  $\theta$  are families' private information.

Families report their income  $z$  and number of children  $n$  to the government, and the government constructs a nonlinear tax system:  $T(z, n)$ . Since  $n$  is binary, the system can be simplified with an  $N + 1$ -tuple tax vector:  $T_n(z)$  for  $n = 0, 1, \dots, N$ . I define the child tax credit to  $n^{th}$  children as:  $k_n(z) = T_{n-1}(z) - T_n(z)$  for  $n = 1, \dots, N$ . The total tax credit received by  $n$ -child families is:  $\sum_{j=1}^n k_j(z)$ . Note that credits are income dependent.

A family consumes  $c$ , which equals the net of income from taxes and expenditures for child raising:  $c = z - T_n(z) - e_n$ . The preference of a family is represented by:

$$\mathcal{U}(c, z, n, \theta, \beta) = u(c) - h(z, b_n, \theta) + m(n, \beta) \quad (1.1)$$

which satisfies Inada conditions:  $\lim_{z \rightarrow 0} \frac{\partial \mathcal{U}}{\partial z} = 0$  and  $\lim_{z \rightarrow \infty} \frac{\partial \mathcal{U}}{\partial z} \rightarrow -\infty$ , and Spence-Mirrlees condition:  $\frac{\partial}{\partial \theta} \left( -\frac{\frac{\partial \mathcal{U}}{\partial z}}{\theta \frac{\partial \mathcal{U}}{\partial c}} \right) \leq 0$ .

Note that child choice is discrete, and hence first-order conditions are not immediately applicable. Therefore, I solve the family problem in two steps.

### 1.3.1 Family Problem

Initially, families determine how many children to have. Second, consumption and income are chosen given the number of children. I use backward induction: Given  $n$ , the optimal income,  $z_n$ , and the optimal consumption,  $c_n$  should satisfy the first-order condition and the budget set:

$$u'(c_n) \left( 1 - \frac{\partial T_n(z_n)}{\partial z} \right) = \frac{\partial h(z_n, b_n, \theta)}{\partial z} \quad (1.2)$$

$$c_n = z_n - T_n(z_n) - e_n. \quad (1.3)$$

These equations imply  $c_n$  and  $z_n$  depend on child-rearing costs,  $b_n$  and  $e_n$ , but they are independent of taste for children,  $\beta$ . Next, I define the indirect utility of  $n$ -child families using optimal consumption and income:

$$V_n(\theta) := u(c_n) - h(z_n, b_n, \theta). \quad (1.4)$$

Using this definition, a  $\theta$ -ability family will have  $n$  children if and only if  $n$ - children choice provides the highest utility:

$$V_n(\theta) + m(n, \beta) \geq \mathcal{V}(\theta, \beta, n') := \max_{n'} \{V_{n'}(\theta) + m(n', \beta)\}.$$

This expression can be simplified by an analysis on the marginal cost and benefit of  $n$  children. Note that child rearing costs are captured by  $V_n(\theta)$ . This implies the marginal cost of  $n$  children equals  $V_{n-1}(\theta) - V_n(\theta)$ . In addition,  $m(n, \beta) - m(n-1, \beta)$  represents the marginal benefit of having  $n$  children. The family decides to have  $n$  children if and only if the marginal benefit of  $n$  children is larger than the marginal cost of  $n$  children while the marginal benefit of  $n+1$  children is less than the marginal cost of  $n+1$  children. Formally,  $(\beta, \theta)$  families decide to have  $n$  children if and only  $\beta \in (\beta_n(\theta), \beta_{n+1}(\theta))$ , where

$$\beta_n(\theta) := M^{-1}(V_{n-1}(\theta) - V_n(\theta)), \quad (1.5)$$

for  $n = 1, 2, \dots, N$  and  $M(\beta) := m(n, \beta) - m(n - 1, \beta)$ .<sup>8</sup> I assume that exogenous parameters satisfy  $\underline{\beta} = \beta_0 < \beta_1(\theta) < \dots < \beta_{N+1} = \bar{\beta}$ . This assumption satisfies that each  $n \in \mathcal{N}$  is chosen by a  $\theta$ -ability family. Since data provides that for all earning ability levels, there is no jump in family sizes, this assumption is valid.

In Figure 1.3, I illustrate the child choice graphically for  $n = 0, 1$ . For a particular ability level, the families with  $\beta \in (\beta_0, \beta_1(\theta))$  choose to have no children because the marginal benefit of one child is less than its costs. For  $\beta = \beta_1(\theta)$ , the benefit and cost having one child is equalized. When  $\beta \in (\beta_1(\theta), \beta_2(\theta))$ , the families decide to have one child because the marginal benefit of one child is higher than its cost and the marginal benefit of second children is less than the marginal cost of second children.

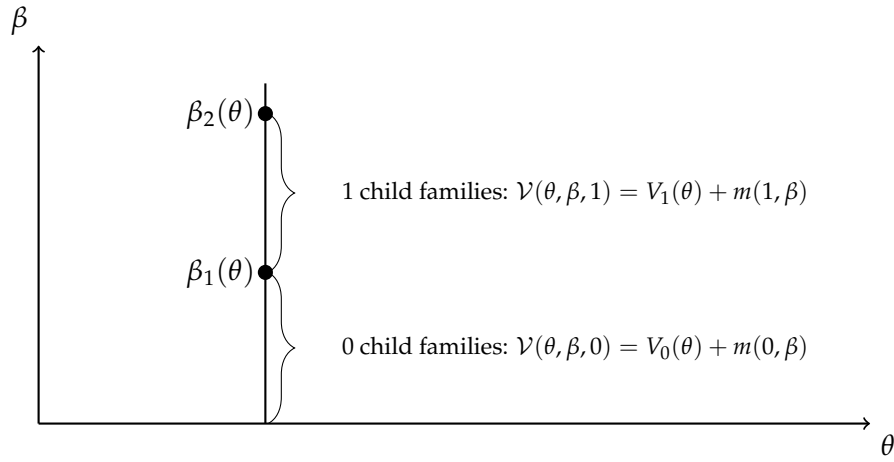


Figure 1.3: Critical Child Taste Levels

In the next subsection, I solve the government's problem using these threshold tastes to handle two dimensional friction in the information.

### 1.3.2 The Government's Problem

The government has a preference over the utilities of families,  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  which is increasingly weakly concave. Using this preference, the government maximizes social welfare. The concavity of  $\Psi$  creates an equity criterion in the government's objective. I also want to mention that this environment is equivalent to an environment in which the government is Utilitarian and  $\Psi$  is a concave transformation of utilities.

---

<sup>8</sup>I fix  $\beta_0(\theta) = \underline{\beta}$  and  $\beta_{N+1} = \bar{\beta}$ .

The characteristics of families are private information. To solve the problem of private information, the government uses a mechanism design.

### Mechanism Design Problem

To construct the optimal tax mechanism, I focus on implementation via direct mechanisms. In a direct mechanism, families report their characteristics to the government and the government optimally chooses consumption, number of children, and income for each families. In addition, these allocations also satisfy that families are not better off by pretending to be another family. Formally, the government solves:

$$\max_{c(\beta, \theta), n(\beta, \theta), z(\beta, \theta)} \int_{\Theta} \int_{\mathcal{B}} \Psi(\mathcal{U}((\beta, \theta))) p(\beta|\theta) f(\theta) d\beta d\theta \quad (\text{MDP})$$

subject to the incentive constraints

$$\mathcal{U}((\beta, \theta)) \geq \max_{(\tilde{\beta}, \theta') \in \mathcal{B} \times \Theta} \mathcal{U}((\beta, \theta), (\tilde{\beta}, \theta')) \quad \forall (\beta, \theta) \in \mathcal{B} \times \Theta \quad (1.6)$$

and the resource constraint

$$\int_{\Theta} \int_{\mathcal{B}} T(\beta, \theta) p(\beta|\theta) f(\theta) d\beta d\theta \geq \mathcal{G},$$

where  $T(\beta, \theta) = z(\beta, \theta) - c(\beta, \theta) - e_n \mathbb{1}(n(\beta, \theta))$  and  $\mathcal{G}$  is the government's expenditure.

Equation (1.6) states that the government should prevent mimicking via two channels, earning abilities and tastes for children. Consequently,  $\infty \times \infty$  possible deviations should be handled, which is hard to solve. To handle such deviations, I follow the arguments of the family problem solution and use the definitions of indirect utility and threshold tastes for children. First, for a given family size, the government uses a first-order approach to prevent deviation via earning abilities:

$$\dot{V}_n(\theta) := \frac{\partial V(\theta)}{\partial \theta} = -h\left(\frac{z_n}{\theta} + b_n\right) + h'\left(\frac{z_n}{\theta} + b_n\right) \frac{z_n}{\theta} \geq 0 \quad (1.7)$$

for all  $n = 0, 1, \dots, N$ . Second, the deviation in tastes are handled by (1.5). First of all, Equation (1.5)

directly implies that mimicking the tastes of families with different sizes does not make families better off. In addition, for a particular family size, mimicking the tastes for children does not alter the children choice and hence does not change the family utility. As a result, the possibility of double deviation in the Equation (1.6) can be handled via Equation (1.4) and (1.5).

I use the Equation (1.4) and (1.5) and adjust the objective function and constraints. In addition, the constraint (1.6) of the problem **MDP** is replaced by Equation (1.4), (1.5), and (1.7). This new problem is a sophisticated version of the original mechanism design problem, and without loss of generality, I call the new problem as the “pseudo-mechanism design problem”.

### Pseudo-Mechanism Design Problem

The government problem solves the following problem:

$$\begin{aligned} \max_{T_n(\theta)} & \int_{\Theta} \int_{\underline{\beta}}^{\beta_1(\theta)} \Psi(V_0(\theta) + m(0, \beta)) p(\beta|\theta) f(\theta) d\beta d\theta + \int_{\Theta} \int_{\beta_1(\theta)}^{\beta_2(\theta)} \Psi(V_1(\theta) - m(1, \beta)) p(\beta|\theta) f(\theta) d\beta d\theta \\ & + \dots + \int_{\Theta} \int_{\beta_N(\theta)}^{\bar{\beta}} \Psi(V_N(\theta) + m(N, \beta)) p(\beta|\theta) f(\theta) d\beta d\theta \end{aligned} \quad (\text{PMDP})$$

subject to Equation (1.4), (1.5), and (1.7) and the resource constraint:

$$\begin{aligned} & \int_{\Theta} \int_{\underline{\beta}}^{\beta_1(\theta)} T_0(\theta) p(\beta|\theta) f(\theta) d\beta d\theta + \int_{\Theta} \int_{\beta_1(\theta)}^{\beta_2(\theta)} T_1(\theta) p(\beta|\theta) f(\theta) d\beta d\theta \\ & + \dots + \int_{\Theta} \int_{\beta_N(\theta)}^{\bar{\beta}} T_N(\theta) p(\beta|\theta) f(\theta) d\beta d\theta \geq \mathcal{G} \end{aligned} \quad (1.8)$$

where  $T_n(\theta) := z_n(\theta) - c_n(\theta) - e_n$  for all  $n = 0, 1, \dots, N$ .

Note that the solution of **(MDP)** and **(PMDP)** are identical:

**Lemma 1.** *The solution of **(MDP)** equals to the solution of **(PMDP)**.*

*Proof.* See Appendix A.0.1. □

The solution of **(PMDP)** provides the optimal taxation of families. The optimal taxation is characterized by the marginal income tax rates for each family size:

**Proposition 1.** *The solution of (PMDP) satisfies the following differential equation*

$$\begin{aligned} \frac{T'_n(\theta)}{1 - T'_n(\theta)} = & \frac{1}{\varepsilon_n(\theta)} \times \frac{1}{\theta f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta))} \times \\ & \int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} (P(\beta_{n+1}|\theta') - P(\beta_n|\theta')) \right. \\ & \left. + \Delta T_{n-1}(\theta') p(\beta_n|\theta') \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta') p(\beta_{n+1}|\theta') \frac{\partial \beta_{n+1}}{\partial V_n} \right] u'(c_n(\theta)) f(\theta') d\theta' \quad (1.9) \end{aligned}$$

for  $n = 0, 1, \dots, N$  where  $z_n, T_n$  is continuous in  $\theta$ , and  $\varepsilon_n(\theta)$  is the elasticity of family income with respect to marginal taxes, and  $g_n(\theta)$  is the weight assigned by the government to  $\theta$ -ability families with  $n$  children, and  $\Delta T_n(\theta) := T_n(\theta) - T_{n+1}(\theta)$  are the tax difference terms.<sup>9</sup>

The formal proof is given by Appendix A.0.2. Here, to provide intuition, I follow Saez (2001) and show the heuristic proof of the Proposition 1. For simplicity, I focus on one-child families and assume  $u(c) = c$ .

Suppose that the government increases the taxes of one-child families with  $\theta' \geq \theta$  earning abilities by  $dT$  (see Figure 1.4). This change creates three effects. First, since the one-child families

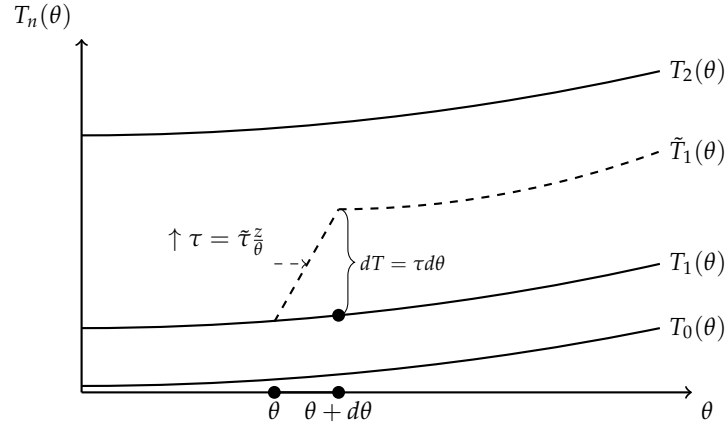


Figure 1.4: Increase in  $T_1(\theta)$  by  $dT$

consume less, there will be a welfare loss for the society by  $g_1(\theta')$  for each dollar of  $dT$  for all  $\theta' > \theta$  where

$$g_1(\theta) := \mathbb{E}_{\beta} \left[ \frac{\Psi'(V_1(\theta) + m(1, \beta))}{\lambda} | \beta_1(\theta) < \beta < \beta_2(\theta) \right].$$

<sup>9</sup>I let  $\Delta T_{-1}(\theta) = 0$  when  $n = 0$  and  $\Delta T_N(\theta) = 0$  when  $n = N$ .



$g_1(\theta)$  measures the average cost of taking an extra dollar more from  $\theta$  families with one-child in terms of the public good.<sup>10</sup> On the other hand, the government collects  $\$dT$  from all of these families, and hence the revenue increases. The total effect for a  $\theta'$  family is:  $dT(1 - g_1(\theta'))$ . The aggregate effect for all families with  $\theta' \geq \theta$  can be written as:

$$dG = dT \int_{\theta}^{\bar{\theta}} (1 - g_1(\theta')) \underbrace{[P(\beta_1|\theta') - P(\beta_2|\theta')]}_{\text{density of 1 child-families}} f(\theta') d\theta'.$$

Note that  $dG$  is a mechanical effect, which does not contain any behavioral responses. Next, I focus on the behavioral responses to  $dT$ .

Second effect is on the income decision of families whose abilities are in  $[\theta, \theta + d\theta]$ . To increase taxes by  $dT$ , the government should increase the marginal taxes of families with  $[\theta, \theta + d\theta]$  by  $\tau = \tilde{\tau} \frac{z_1}{\theta}$ , where  $\tilde{\tau}$  represents the change in the marginal tax rates on income (see Figure 1.4).<sup>11</sup> This increment creates a behavioral effect, i.e. the families in the small band decrease their income by  $dz = \frac{z_1 \varepsilon_1(\theta) \tilde{\tau}}{1 - T_1'(\theta)}$ , where  $\varepsilon_1(\theta) := \frac{\partial \log z_1}{\partial \log(1 - T_1'(z_1))}$  is the elasticity of income with respect to marginal tax rates. Combining the terms gives the first behavioral effect is:

$$dB_1 = -T_1'(\theta) dz f(\theta) d\theta = -dT \frac{T_1'(\theta)}{1 - T_1'(\theta)} \varepsilon_1(\theta) \theta \underbrace{[P(\beta_1|\theta) - P(\beta_2|\theta)]}_{\text{density of 1 child-families}} f(\theta).$$

If the number of children was exogenously given, there would not be any extra effect. Hence, if the original mechanism was optimal, these effects should sum up to zero:  $dG + dB_1 = 0$ . In this situation, the optimal tax formula would then be very similar to that of [Mankiw and Weinzierl \(2010\)](#).<sup>12</sup> However, the number of children to have is a choice in my set up, and  $dT$  affects the optimal number of children of families whose tastes for children are in the neighborhood of  $\beta_1(\theta)$  and  $\beta_2(\theta)$  (see Figure 1.5).

The one-child families whose tastes for children are in the neighborhood of  $\beta_1(\theta)$  prefer to have no children after the increase in their taxes. As a result, their tax liabilities are changed by:  $\Delta T_0(\theta') := T_0(\theta') - T_1(\theta')$  for all  $\theta' \geq \theta$ . For a particular  $\theta'$ , the effective change is:  $\Delta T_0(\theta') \frac{\partial \beta_1(\theta')}{\partial V_1(\theta')} p(\beta_1|\theta') f(\theta')$  where  $\frac{\partial \beta_1(\theta')}{\partial V_1(\theta')}$  is the mechanical effect of  $V_1(\theta')$  on  $\beta_1(\theta')$  and  $p(\beta_1|\theta') f(\theta')$  is the density of these

<sup>10</sup>See Equation (A.1) for a general definition.

<sup>11</sup>To change the marginal rates over abilities by  $\tau$ , the marginal rates on income should increase by  $\tilde{\tau}$ .

<sup>12</sup>Taxes for different categories can be considered as taxes for families with different sizes.

families. Similarly, one-child families in the neighborhood of  $\beta_2(\theta)$  prefers to have two children. For this case, the effective change is:  $\Delta T_1(\theta') \frac{\partial \beta_2}{\partial V_1} p(\beta_2|\theta') f(\theta')$ .<sup>13</sup>

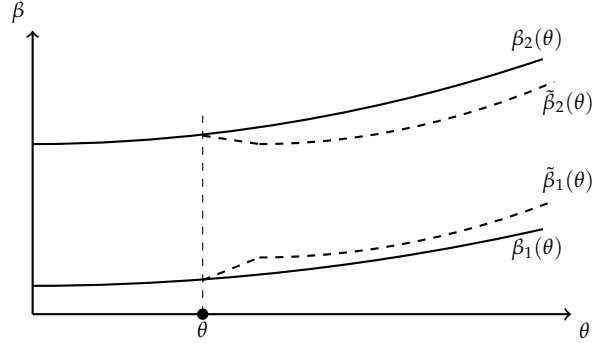


Figure 1.5: Changes in Critical Child Taste Values

The aggregate effect of the change of the family size is represented by:

$$dB_2 = \int_{\theta}^{\bar{\theta}} \left( \Delta T_0(\theta') p(\beta_1|\theta') \frac{\partial \beta_1}{\partial V_1} + \Delta T_1(\theta') p(\beta_2|\theta') \frac{\partial \beta_2}{\partial V_1} \right) f(\theta') d\theta'.$$

Together with the second behavioral effect, the original mechanism is optimal if  $dG + dB_1 + dB_2 = 0$ . This equality gives the equation in Proposition 1 when  $u(c) = c$  for  $n = 1$ . Note that, these procedures can be applied for any  $n = 0, 1, \dots, N$  to find the optimal marginal tax rates of families with  $n$  children.

Next, I state how the tax formula in Proposition 1 differ from the tax formulas in the literature.

**Novelty of the tax formula:** The tax formula in Proposition 1 varies from the conventional formulas of the literature in three ways. First, the elasticity component,  $\varepsilon_n$ , is *endogenous*. The endogeneity arises because time is perfectly substitutable between child care and market time. Time devoted to child care reduces the time devoted to labor and makes labor (income) more sensitive to tax changes. In the following lemma, I prove this for a particular case:

**Lemma 2.** Let  $u(c) = c$  and  $h(x) = \frac{x^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$ . The elasticity of income with respect to marginal tax rates is:  $\varepsilon_n(\theta) = \varepsilon(1 + \frac{b_n}{z_n/\theta})$ .

*Proof.* Define elasticity as  $\varepsilon_n := \frac{\log \partial z_n}{\log \partial(1-T'_n)} = \frac{1-T'_n}{z_n} \frac{\partial z_n}{\partial(1-T'_n)}$ . The first-order condition for income is:  $(1 - T'_n) = h'(\frac{z_n}{\theta} + b_n)$ . Taking the derivative with respect to  $(1 - T'_n)$  and rewriting yields:

<sup>13</sup>The new threshold tastes are represented by  $\tilde{\beta}_1(\theta)$  and  $\tilde{\beta}_2(\theta)$ . See Figure 1.5.

$$\varepsilon_n = \frac{h'(\frac{z_n}{\theta} + b_n)}{h''(\frac{z_n}{\theta} + b_n) \frac{z_n}{\theta}} = \varepsilon(1 + \frac{b_n}{z_n/\theta}). \quad \square$$

It is straightforward to see that  $\varepsilon_n(\theta)$  depends on  $z_n$ , and hence the elasticity of income of parents is endogenous. Moreover,  $\varepsilon_1(\theta) > \varepsilon_0(\theta)$  because childless families do not spend time on child care, i.e.  $b_0 = 0$ . This result is in line with [Blundell et al. \(1993\)](#) who find the labor elasticity of families with children is higher than those without children.

Second, a *novel* term, the density of family sizes appear in the formula:  $f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta))$ . This term is endogenous because the family size is a choice. The term provides information about the underlying tastes for children. The government knows that families with tastes in  $(\beta_n(\theta), \beta_{n+1}(\theta))$  will generate same income and will have same number of children if they face same marginal tax rates.

Third, the second *novel* term, the tax difference term  $\Delta T_n(\theta)$ , shows up in the formula. This is mainly because the government's redistributive motives are shaped not only by insuring low earning abilities but also by insuring families with more children. Families with more children are faced with the high child taste draws. Because of this draw, they produce children, and consequently their consumption and time to generate income are reduced. As a result, their welfare decreases. Hence, the government provides insurance for these families.

Next, I provide an interpretation of the conventional and novel terms that appear in the tax formula.

**Interpretation of the terms:** The interaction of the terms in Equation (A.2) is complex. Here, I go over term by term and provide a basic interpretation of each term. First, the elasticity,  $\varepsilon_n$ , is reciprocally correlated with the marginal taxes. Note that the marginal taxes create distortions on income decision and the distortions are higher for families with higher elasticity of income. The distortions create a deadweight loss for the economy and the government considers this loss and reduces the marginal taxes of those with higher income elasticity.

Second, the density of family sizes,  $f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta))$ , decreases the marginals. Intuitively, if the density is large, the impact of the distortions created by the marginal taxes will be large. Therefore, the government decreases marginal rates.

Third, when the benefit of increasing taxes,  $(1 - g(\theta))$ , rises, the government increases marginals. The intuition is straightforward.

Finally, I focus on the tax difference terms. The first term,  $\Delta T_{n-1}$ , tightens the incentive constraints (see Equation (1.5)). The government relaxes such constraints by decreasing the marginal rates of  $n$ -children families. On the other hand,  $\Delta T_n$  relaxes the incentives, and hence government increases marginal rates.<sup>14</sup>

In order to explore the forces behind the tax formula, I bring my model to the data.

## 1.4 Quantitative Analysis

In this section, I quantitatively examine the optimal taxation of families using the US data. Initially, I estimate the earning ability distribution and the child taste distribution. Using these estimates, I solve the optimal tax mechanism numerically.

According to the empirical labor market literature, the effect of non-labor income on labor is small (see [Blundell and Macurdy \(1999\)](#)). In addition, to understand the relationship between labor income and number of children, it is natural to eliminate the non-labor income effect on labor. Therefore, I assume that families have a quasi-linear preference in consumption:  $u(c) = c$ . Moreover, I assume that childless families a constant elasticity of income with respect to marginal rates:  $h(z, \theta) = \frac{\varepsilon}{1+\varepsilon} \left(\frac{z}{\theta}\right)^{\frac{\varepsilon}{1+\varepsilon}} \theta$  where  $\varepsilon := \frac{\partial \log z}{\partial \log(1-T')}$  is the elasticity of the total family income with respect to marginal taxes. The estimate for the elasticity of *family income* requires attention, because the literature on elasticity of income is based on individual levels. I study individual elasticities to figure out the family elasticity in [Appendix A.0.3](#). In the benchmark, I use  $\varepsilon = 0.56$ . Note that this number is quite close to the elasticity estimates in [Chetty \(2012\)](#), who creates a common confidence interval for the elasticities of different studies.

### 1.4.1 Sample Selection

To capture estimates for earning ability and child taste distribution, I use the March release of the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics.<sup>15</sup> I use the sample of 2005-2014 years. In these years, families report both their child tax credits and their marginal tax rates. Also, I make some sample restrictions on data.

---

<sup>14</sup>Note that  $\frac{\partial \beta_n}{\partial V_n} < 0$  and  $\frac{\partial \beta_{n+1}}{\partial V_n} > 0$ .

<sup>15</sup>Data is taken from: Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

First, I restrict the sample to two-spouse families in which both spouses are employed. This restriction eliminates potential time difference between one-spouse and two-spouse families. In addition, the employment status of spouses rules out the extensive margin decision, which helps to capture a fine estimate for elasticity of income. I also naturally assume that both spouses work at least 5 hours per week.

Second, I put lower and upper bound on the age of each spouse. The spouses are 35-45 years old. The age restriction helps in three ways: First, the age effect on income and children is eliminated. Empirical evidence suggests that earnings increase in the early ages (16-35) and become stabilized after the age of 35.<sup>16</sup> Moreover, early age households may postpone child decision because of socioeconomic factors. The possibility of this delay is filtered by the age restriction. Second, the fertility behavior can still evolve in this age range. Third, the probability of that some children have grown up and left the family is minimized. Age restriction is used by many works such as Docquier (2004), Jones and Tertilt (2008), and Jones et al. (2010). These positive works study the relationship between fertility and family income and put boundaries on the female ages to rule out the age effect.

Third, I remove families whose main source of income is not labor income. The total labor income of family should be 80% of total family income. Also, I focus only on families in which total labor income of each spouse is at least 80% of their total income (refer to Ales, Kurnaz, and Sleet (2015)). This assumption is constructed to validate the quasi-linear preference assumption and to capture a fine estimate for the family income elasticity.

Finally, I eliminate families who earn less than \$250 (see Heathcote, Perri, and Violante (2010b) for further details on CPS). The final sample has 37,165 families.

I plot the relationship between family labor income and the number of children in the family in Figure 1.6. The figure implies the well-known empirical evidence that the fertility rate is negatively correlated with family labor income.

In the next two subsections, I focus on the child-rearing costs and find estimates for  $b_n$  and  $e_n$ .

---

<sup>16</sup>See <http://www.bls.gov/news.release/wkyeng.t03.htm>

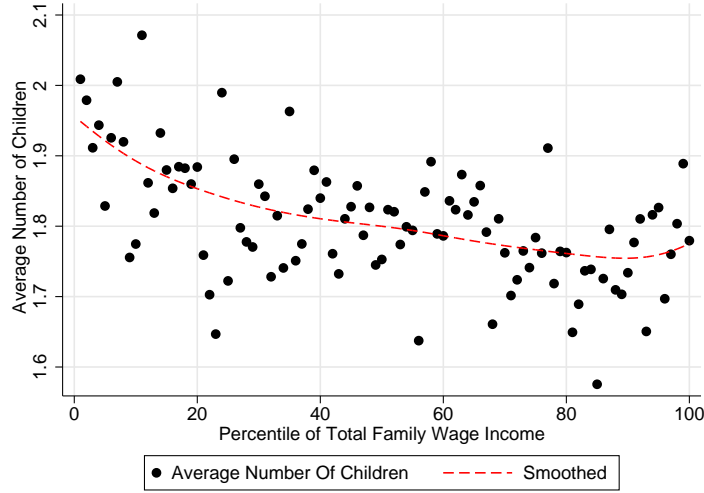


Figure 1.6: Income-Fertility Relation

Data: CPS 2005-2014. The sample is restricted to married (both spouses are present) households whose main source of income is labor. Total family wage income is converted to 2014\$ using CPI deflator. The age of the spouses is between 35 and 45. The sample size is 37,165 after all restrictions.

## 1.4.2 Parental Time

The assumption on the cost of earnings and child care suggests that time is normalized to one:  $h(0) = 0$  and  $h'(1) = 1$  (see [Kleven et al. \(2009\)](#)). Hence,  $b_n$  is the fraction of child care to the total labor time (market work and child care). To capture an estimate for  $b_n$ , I use the 2003 wave of American Time Use Survey (ATUS) sample which is also used by [Aguiar and Hurst \(2007\)](#). I restrict the sample using the criteria stated in Subsection 1.4.1. I show the time devoted to child care and market work in Table 1.1.

Category of labor:	0 child family	1 child families	$\geq 2$ children families
child care	0.2	5.7	9.4
market	57	54.9	54.9
$b_n \simeq$	0	0.09	0.14
sample size	158	247	547

Table 1.1: Time Devoted to Market and Childcare

Data: ATUS-2003. Each number in the second and the third row represents the weighted average hours per week devoted to the related category. The sample is restricted to married, 35-45 years aged, and working households who devote total time to market and child care at most 100 hours. Since  $b_2 \simeq b_{\geq 2}$ , I used the latter one.

Table 1.1 shows that there is an economics of scale in the time cost of child-rearing. More analysis on  $b_n$  can be found in Appendix A.0.4. In the benchmark, I assume one-child families devote 9% of their time to child care and two-child families spend 14% of their time for child-

rearing.

### 1.4.3 Cost of Goods

Haveman and Wolfe (1995) use Consumer Expenditure Survey (CEX) data and suggest that the goods cost is \$12,788 per child (in terms of 2014\$). Examples of such costs include expenditures on food, housing, transportation, clothing, and health care. More recently, a publication of the US Department of Agriculture, Lino (2014), analyzes the goods cost of child-rearing for families with different wealth.<sup>17</sup> This work particularly provides information on expenditures for children with different age. Using this information, I derive a range of expenditures for two-spouse families in Table 1.2.

Category of Families	Average Income	1 child	2 children	3 children
Low Income	39,989	11,597-13,211	19,847-21,137	23,710-24,683
Middle Income	84,114	16,442-19,014	28,149-30,426	33,444-35,273
High Income	189,443	27,095-32,646	47,793-52,234	56,104-58,022
$e_n \simeq$		\$12,000	\$20,000	\$24,000

Table 1.2: Expenditures on child-rearing

Table-1 and Table-8 of Lino (2014) is used. The first column categorizes families according to their income level. The second column presents the average income for each category. The last three columns represent the range of expenditures on child-rearing. The expenditures are converted to 2014\$.

Note that the ranges of expenditures for each category of families are small. In benchmark case, I follow low-income family expenditure and let  $e_1 = \$12,000$ ,  $e_2 = \$20,000$ , and  $e_3 = \$24,000$ . I interpret the extra costs for middle and high income families as a part of their family consumption. Note that  $e_2 \simeq e_3$ . In the numerical solution, I assume that families can have either 0 children or 1 child or 2+ children.

Using the estimates of child-rearing costs, I derive distribution of earning abilities and tastes for children in the next two subsections, respectively.

---

<sup>17</sup>See [http://www.cnpp.usda.gov/sites/default/files/expenditures\\_on\\_children\\_by\\_families/crc2013.pdf](http://www.cnpp.usda.gov/sites/default/files/expenditures_on_children_by_families/crc2013.pdf).

#### 1.4.4 Estimation of the Distribution of Earning Abilities

The quasi-linear preference structure allows me to find earning abilities of families (see Equation (1.2)):

$$\theta = \frac{z}{(1 - T')^\varepsilon - b_n}. \quad (1.10)$$

Note that CPS has information about family structure and detailed family income and taxes.<sup>18</sup> Given the complexity of state rates, I focus only on federal tax rates. In addition, I add earned income tax rates to the federal marginal tax rates.<sup>19</sup> Using  $b_n$  values from Table 1.1 and the weights of families given in the data, I derive the distribution of earning abilities and show it in Figure 1.7.

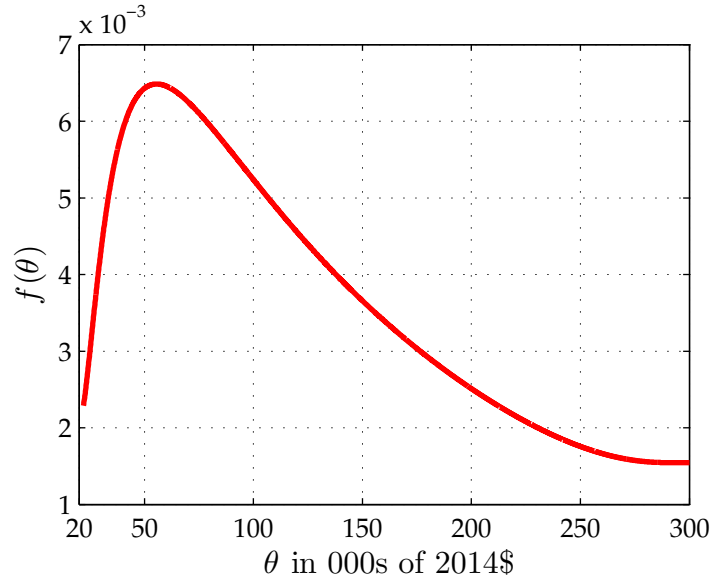


Figure 1.7: Earning Ability Distribution

#### 1.4.5 Estimation of the Distribution of Child Taste

An important contribution of this paper introduces a distribution of tastes for children to the literature. I assume that  $\beta \stackrel{i.i.d}{\sim} [0, \bar{\beta}]$  is distributed according to a power function, i.e. the cumulative

<sup>18</sup>The data I use contains information on characteristics of each spouse in a family. Also, types of income for each spouse are given in detail. Moreover, families also report their the child tax credits, federal marginal tax rates, and federal and state tax liabilities.

<sup>19</sup>The families report how much earned income credit they received. Yet, the data does not provide if credits are in the phase-in or out region. I use the information on EITC for years 2005-2014 to figure out the marginal effect of the credit.



density is  $P(\beta) = \left(\frac{\beta}{\bar{\beta}}\right)^\eta$ . The density of childless families equals  $P(\beta_1(\theta))$  for each  $\theta$ . Hence, I interpret  $\eta := \frac{\partial \log P(\beta_1)}{\partial \log \bar{\beta}}$  as the non-participation elasticity of zero-child families with respect to their tastes for children. Moreover, I assume  $m(n, \beta) = -(N - n)^p \beta$ . Within this framework, I need to estimate  $\eta$  and  $p$ .

I use Bernoulli maximum likelihood estimation to find the estimates. First, I derive percentiles of  $\theta$  distribution and calculate  $V_n(\theta_j)$  for each  $j$  -  $th$  percentile.<sup>20</sup> Second, I can calculate the fraction of  $n$ - child families:  $\pi_n(\theta_j)$ . In addition, I calculate the average number of children for each  $\theta_j$ :  $n(\theta_j)$ . I plot  $V_n(\theta_j)$ ,  $\pi_n(\theta_j)$  and  $n(\theta_j)$  in Figure 1.8.

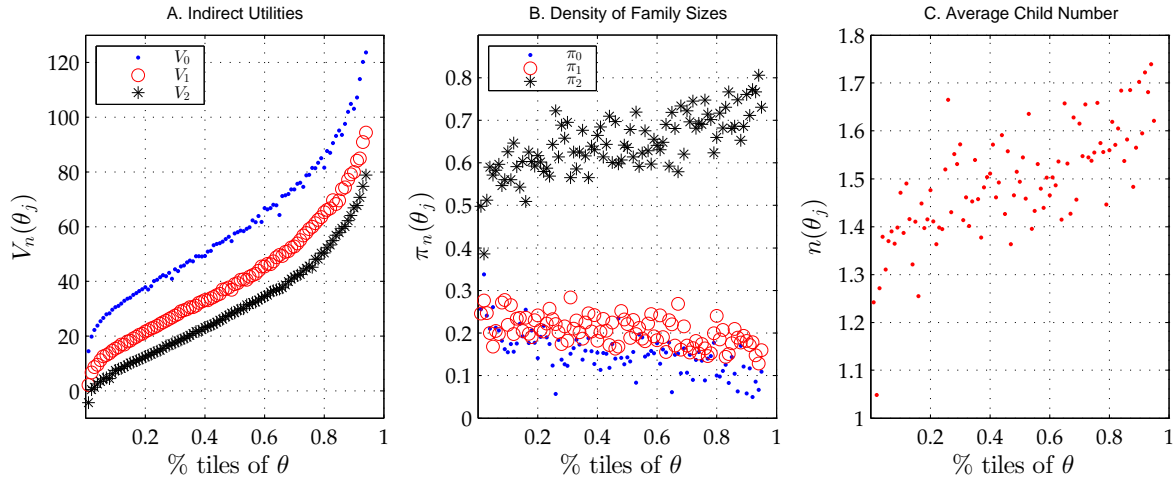


Figure 1.8: Data:  $V_n(\theta_j)$ ,  $\pi_n(\theta_j)$ , and  $n(\theta_j)$

Figures for  $V_n(\theta_j)$ ,  $\pi_n(\theta_j)$ , and  $n(\theta_j)$  respectively. Also within the sample restriction, the average number of children in the environment is 1.492. Note that families with more than two children are considered as they have two children.

Note that the probability of having  $n = 0, 1, 2$  children are represented by  $P_0(\theta_j) := P(\beta_1(\theta_j))$ ,  $P_1(\theta_j) := P(\beta_2(\theta_j)) - P(\beta_1(\theta_j))$ , and  $P_2(\theta_j) := 1 - P(\beta_2(\theta_j))$ , respectively. The Equation (1.5) provides their values for each  $\theta_j$ . Next, I fix the upper bound with  $\bar{\beta} = 300$  and derive the Bernoulli maximum likelihood function:

$$\max_{\eta, p} \mathcal{L} = \prod_j P_0(\theta_j)^{\pi_0(\theta_j)} P_1(\theta_j)^{\pi_1(\theta_j)} P_2(\theta_j)^{\pi_2(\theta_j)}. \quad (1.11)$$

The estimates are given in Table 1.3 and I plot the distribution of child taste in Figure 1.9.

<sup>20</sup>Note that CPS has information on how much taxes a family pays. Hence, I can calculate  $V_n(\theta)$ .

$\eta$	$p$
0.31	4.82
( 0.83)	(13.50)

Table 1.3: Estimation of Child Taste Distribution

Standard errors in parenthesis. Also  $Cov(\eta, p) = 8.5$ .

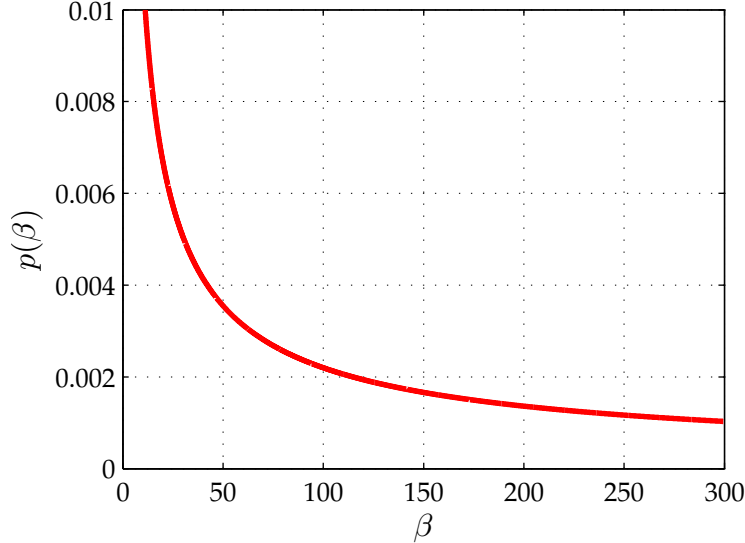


Figure 1.9: Probability distribution of child taste:  $p(\beta)$

#### 1.4.6 Deriving the Optimal Tax System

To solve the problem numerically, I find that the government collects \$13,412 per capita taxes from the sample I use, while the population produces \$109,421 per capita income. Hence, I set  $\mathcal{G} = \$13,412$  in my calculations. Before solving the optimal system, first I show the taxes of families under the current tax system in Figure 1.10. Panels A and B show the tax liabilities for low and high income families, respectively. Panels C and D suggest that the child tax credits are constant for the first seven quantiles. The credits decline after that and reach zero around the ninth quantile. Moreover, the marginal tax rates are higher for families with children at the first two quantiles. This is because the earned income tax credits fall from the plateau and increase the marginal rates (see Figure 1.1).

I solve the government's problem (i.e. **PMDP**) at my selected and estimated parameters using the GPOPS-II software.<sup>21</sup> Note that the government problem is an optimal control problem and

<sup>21</sup>GPOPS-II is a flexible software for solving optimal control problems. For additional details see [Patterson and Rao](#)

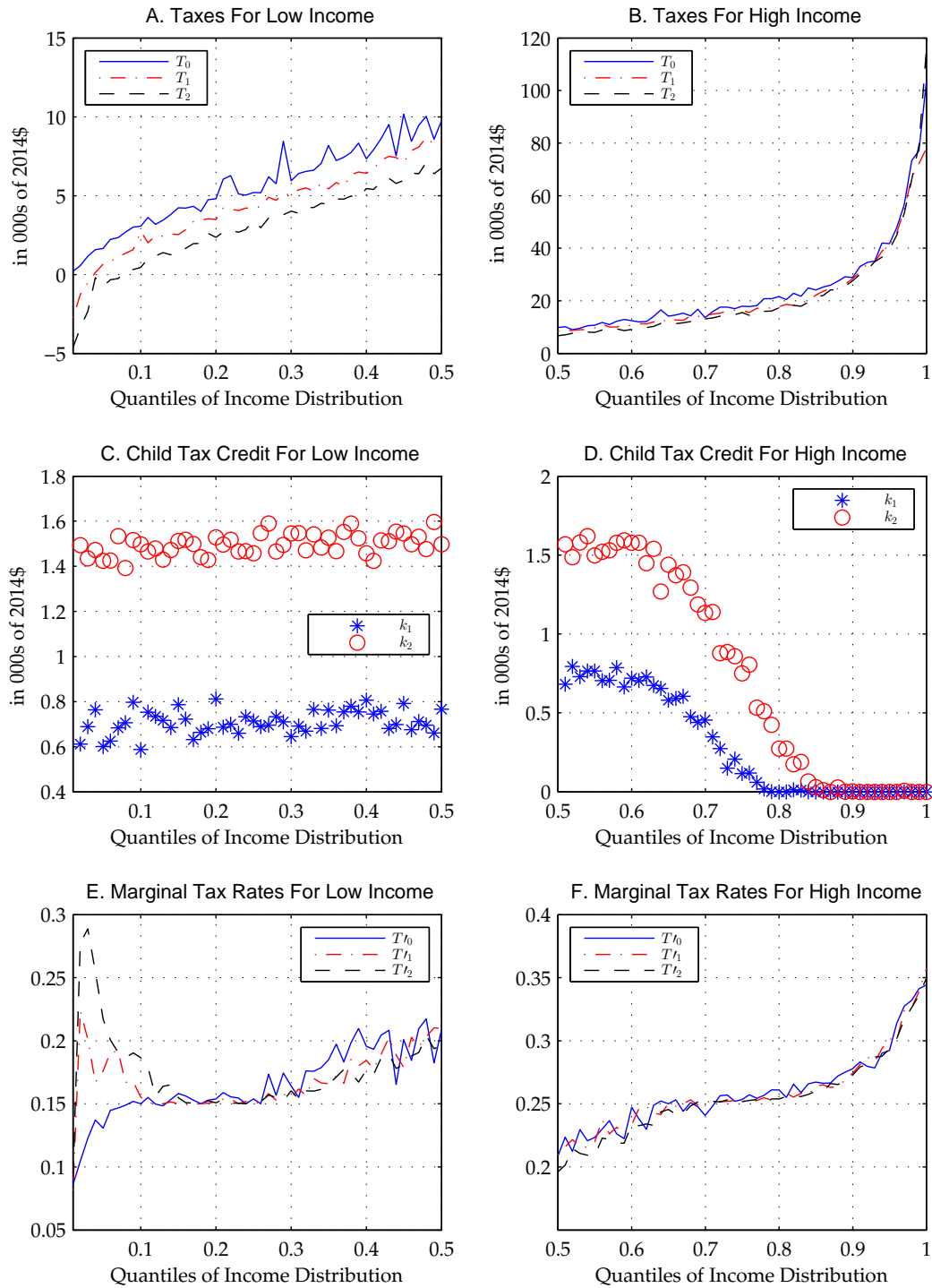


Figure 1.10: Current Tax System

Left (right) panel is for the left (right) side of the income distribution. The first row shows the actual taxes paid by families. The second row shows how much child tax credits they get. Note that  $k_1$  ( $k_2$ ) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families. See Figure A.3 in Appendix A.0.5 for \$ base taxes.

the Hamiltonian of the problem is stated in Appendix A.0.2.

First I plot the optimal indirect utilities and optimal family sizes in Figure 1.11. It is clear

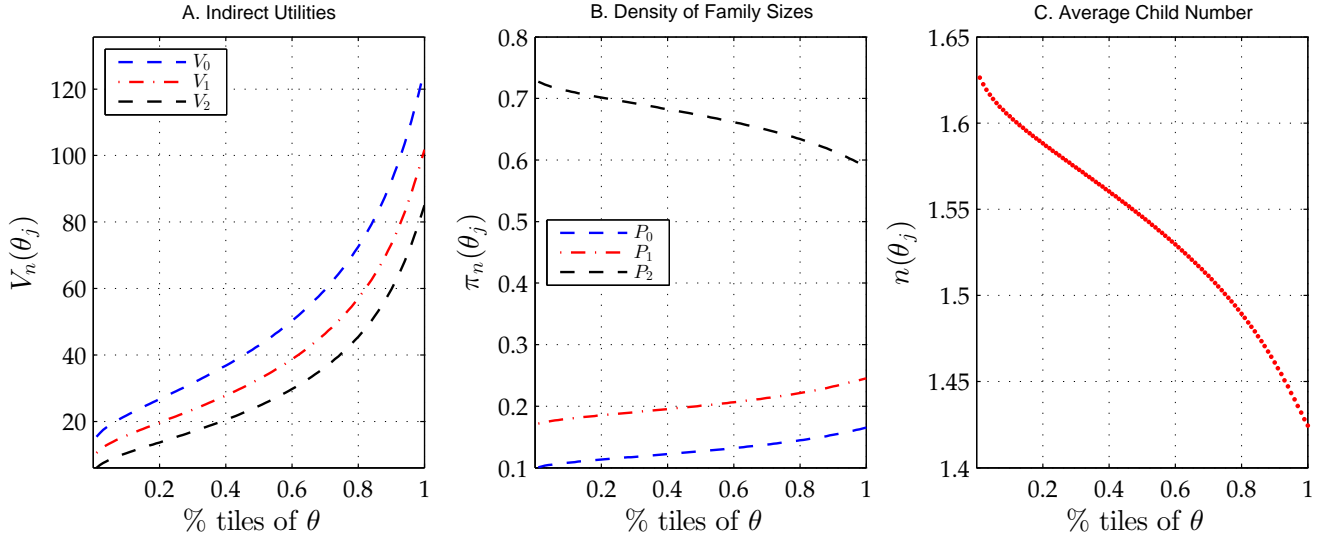


Figure 1.11: Optimum:  $V_n(\theta_j)$ ,  $\pi_n(\theta_j)$ , and  $n(\theta_j)$

Figures for  $V_n(\theta_j)$ ,  $\pi_n(\theta_j)$ , and  $n(\theta_j)$  respectively. Average number of children is 1.539 which is relatively 3% more than the current number children in the data.

that the current and the optimal indirect utilities of families are similar. However, the density of family sizes has distinct patterns across abilities (see Figure 1.8). The density of two child families decreases with earning abilities. The situation is reversed for childless families. The main reason behind this result is the time cost of child-rearing. The cost is relatively higher for families with higher earning abilities and the higher earning families produce fewer children. Interestingly, the density of the one-child family also increases. This result stems from economies of scale in child-rearing. Because of this scale, the credit received by one child families is higher than the average child credit received by two children families. As a result, the density of one child families increases with earning abilities.

Next, I numerically solve for the optimal tax system. First of all, the transversality conditions of Hamiltonian satisfy the conditions in Sadka (1976) and Seade (1977). As a result, the bottom and top of the incomes for each family sizes face zero marginal rates (see Panels E and F of Figure 1.12).

Panel E of Figure 1.12 shows that the government distorts the labor decision of two-child families more at the bottom. In return, these families receive a high subsidy via tax credits (see Panel A of Figure 1.12). This mainly stems from the effect of goods costs. The government provides

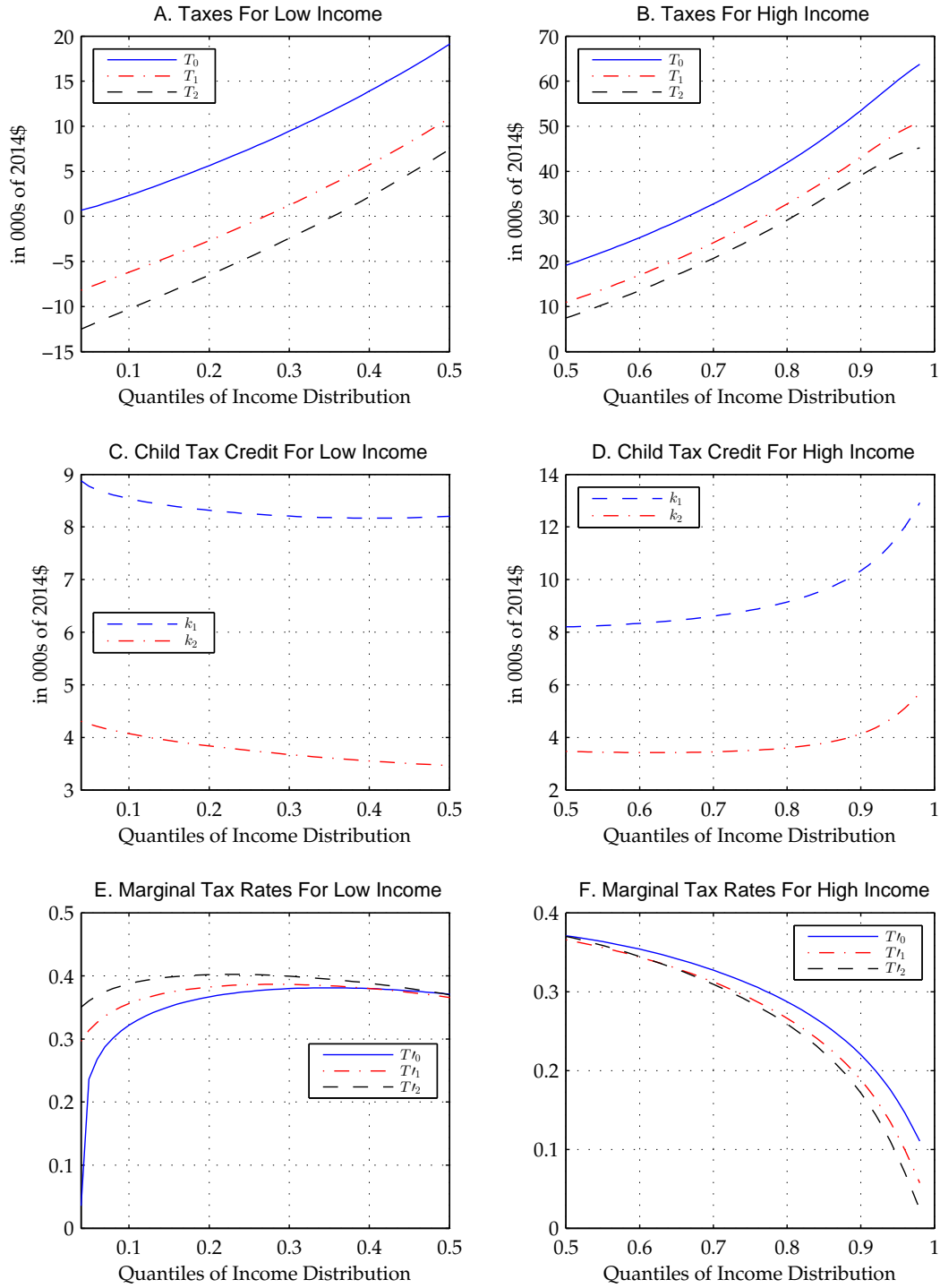


Figure 1.12: Optimal Tax System

Left (right) panel is for the left (right) side of the income distribution. The first row shows the actual taxes paid by families. The second row shows how much child tax credits they get. Note that  $k_1$  ( $k_2$ ) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families. See Figure A.4 in Appendix A.0.5 for \$ base taxes.

enough goods to low-income families to raise their children. This also provides incentives for the low-income families to produce more children and less income, because the cost of generating income is relatively higher than the cost of child-rearing for the bottom. On the other hand, the distortion is relatively less for families with children at the top of the income distribution (see Panel F of Figure 1.12). This is mainly because the government does not want to increase the distortion on the income decision of the families with children whose income is more elastic because of the time cost of the child-rearing. In addition, the government subsidizes these families to relieve their income loss due to the time cost of child-rearing (see Panel B of Figure 1.12).

Panels C and D of Figure 1.12 show the pattern of the child tax credit terms. Both credit terms ( $k_1(z)$  and  $k_2(z)$ ) are U-shaped and affected by the impacts of child-rearing costs. The left tick stems from the goods costs. Although, the goods costs decrease the consumption of all families with children, the decrease for low-income families is relatively higher. On the other hand, the time cost affects the consumption of high-income families more and creates the right tick. To grasp the intuition behind these results, consider two one-child families with  $\theta = 20,000$  and  $\theta = 100,000$ . In a *laissez-faire* economy, the goods costs ( $e_1 = \$12,000$ ) consume 60% and 12% of the family income, respectively. Hence, the decrease in the marginal utility of consumption because of the goods cost is higher for the family with low earning ability. As a result, the credits are pushed up for low income families. On the other hand, the virtual income losses of families due to the time cost ( $b_1 = 0.09$ ) are \$1,800 and \$9,000, respectively. This implies that the reduction on the marginal utility of consumption because of the time cost is higher for the family with higher earning ability with sufficient risk aversion in the preferences. As a result, the credits are pushed up for the high income families. Therefore, the credits are U-shaped.

These results suggest that the current US tax system ignores the time cost of parents. An adjustment on child tax credits and especially on the top-income earners can improve welfare. I find that the welfare gain from implementing the optimum is 1.1% in terms of equivalent increase in consumption for all families.

In the next subsection, I provide a simpler version of the optimal child tax credits. I create tax credits which are linear with respect to income.

### 1.4.7 Proposal

In this subsection, I propose a simple tax schedule. I let that the income taxes are determined by the optimal taxes of childless families. In addition, I propose a linear income dependent tax credits. The credits for the first (second) child linearly decrease in the first quartile, and are constant for the 25-65% (25-75%) of the income distribution. For the rest of income distribution, the credits linearly increase. I state the linear rates in Table 1.4. In addition, the minimum credits is determined by using the minimum values of the optimal credits.

Credit	Credit Rate	Phase In	Phase Out	Credit Rate
$\hat{k}_1$	-2%	\$46,700	\$110,000	3.7%
$\hat{k}_2$	-2.8%	\$46,700	\$139,000	2.1%

Table 1.4: Credit Rates in Phase In and Phase Out

\$46,700, \$110,000, and \$139,000 refer to 25 %, 65%, and 75% of the income distribution, respectively.

I plot the optimal and the proposed credits in Figure 1.13. With this proposal, almost 87% of the welfare gain attained by the optimum is captured. To understand how the income dependent taxes improve social welfare, I also consider a proposal in which the credit per child is same for all children and constant across income. This proposal only captures 70% of the welfare gain. As a result, the income dependent child tax credits can improve social welfare significantly.

In the next section, I check the robustness of the U-shaped tax credits. I relax restrictions on the sample and derive the optimal tax credits.

## 1.5 Robustness

In this section, I analyze robustness of U-shaped tax credits. First, I relax the age restriction on the sample in the following subsection. Next, I work on the types of the goods and time costs in detail. Finally, I study the optimal tax credits for single mothers.

### 1.5.1 Age Analysis

In this subsection, I relax age restriction of the sample. The minimum age of a spouse in a two-spouse family is relaxed to 25. I calculate the optimal credits and plot them in Figure 1.14. The

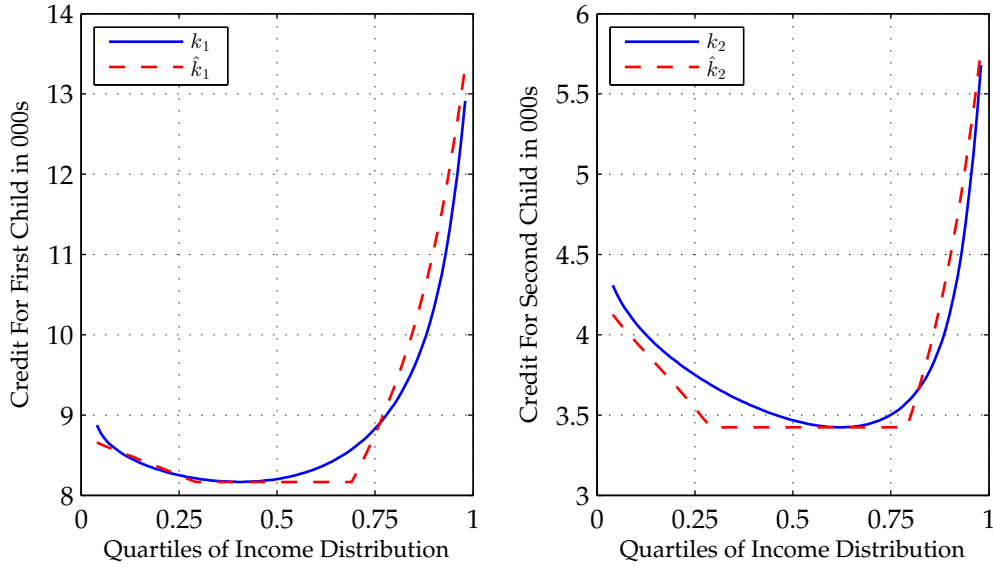


Figure 1.13: Proposal Tax Credits

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$.  $k_n$  is the optimal tax credits and  $\hat{k}_n$  is the proposed tax credit for  $n = 1, 2$ . See Figure A.5 in Appendix A.0.5 for \$ base credits.

tax credits are U-shaped for this sample. As a result, the credit shapes are quite robust without a restriction on ages of spouses.

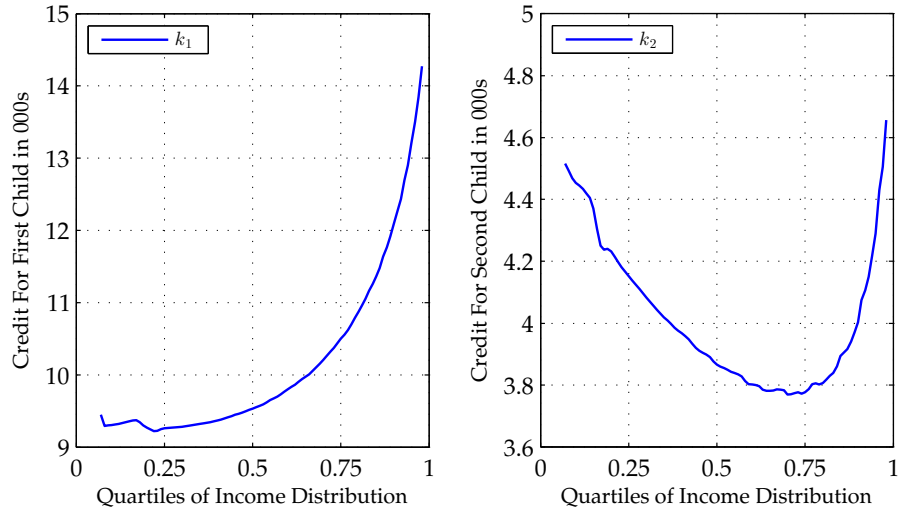


Figure 1.14: Tax Credits: Age Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$. See Figure A.6 in Appendix A.0.5 for \$ base credits.

In the next subsection, I focus on the details of child-rearing costs.



### 1.5.2 Detailed Cost Analysis

I focus on the details of the child-rearing costs. Some analysis suggests that not all types of time devoted to child raising are costly for parents. For example, [Godbey and Robinson \(1999\)](#) state that parents enjoy playing with their children and reading to their children. In this subsection, I set the time cost of child-rearing to basic child care activities such as looking after children, activities related children health. ATUS 2003 data provides such information on child-rearing. I set  $b_1 = 7\%$ ,  $b_2 = 11\%$ .

Next, I change the set of expenditures on the goods cost. Some might consider that not all types of goods cost are on the basic needs for children. For example, moving a bigger house, which is around 30% of the goods cost, can be considered as a non-required cost for child-rearing. I modify the set of goods costs for basic needs. The new set consists of the expenditure on food, clothing, health care and education. These costs are around 50% of the original goods costs (see [Lino \(2014\)](#)). As a result, I set  $e_1 = \$6,000$  and  $e_2 = \$10,000$ .

In Figure 1.15, I show that the optimal tax credits for this environment. Note that the credit amount are reduced due to the reduction in the child-rearing costs (see Figure 1.13). However, the credits are still U-shaped, and my results are robust.

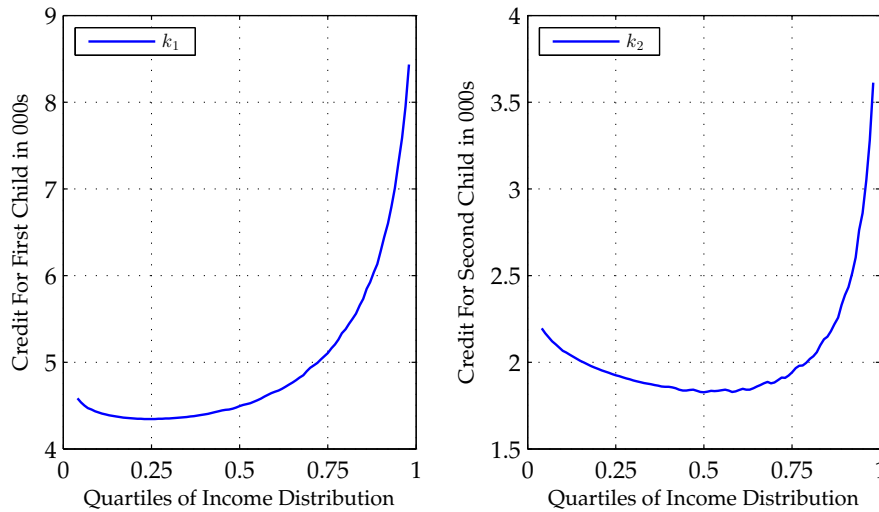


Figure 1.15: Tax Credits: Costs Analysis

$k_n$  is the optimal tax credits for  $n = 1, 2$ . The amounts are 1,000 in 2014\$. See Figure A.7 in Appendix A.0.5 for \$ base credits.

In the next subsection, I study if the marital status of households matters for the shape of tax

credits. Since there are few single fathers, I focus only on single mothers.<sup>22</sup>

### 1.5.3 Marital Status Analysis

In this subsection, I study the optimal taxation of single females. I make adjustment on child-rearing costs. The time costs for single mothers are set by  $b_1 = 10\%$  and  $b_2 = 19\%$ , and goods cost are set by  $e_1 = \$11,500$  and  $e_2 = \$18,500$  (see Table 8 in Lino (2014)). I pick the elasticity of income as  $\varepsilon = 0.8$  (see Blundell, Pistaferri, and Saporta-Eksten (2012)). The optimal tax credits are U-shaped (see Figure 1.16). Note that the credit amounts are larger than the case for married families. The main reason is that the time cost for singles is bigger than the time cost for marrieds.

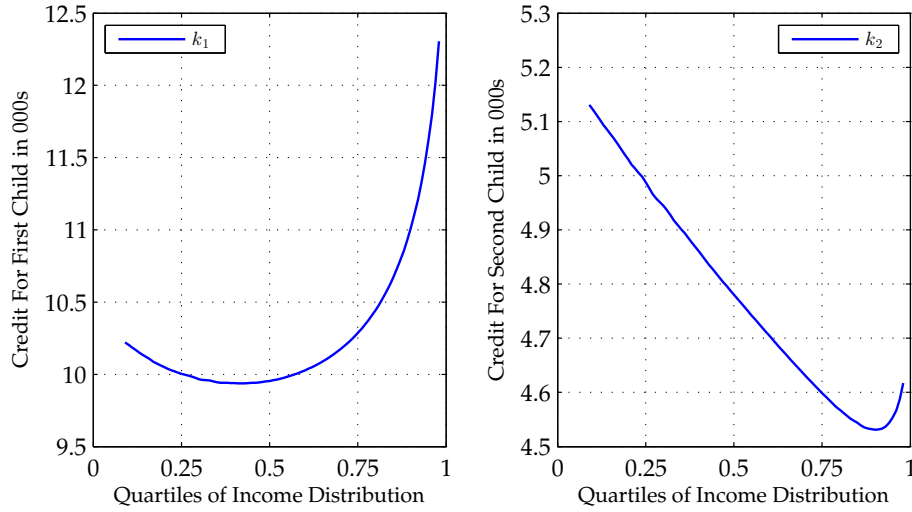


Figure 1.16: Tax Credits: Marital Status Analysis

$k_n$  is the optimal tax credits for  $n = 1, 2$ . The amounts are 1,000 in 2014\$. See Figure A.8 in Appendix A.0.5 for \$ base credits.

## 1.6 Conclusion

This paper studies optimal income taxation and child tax credits in a static Mirrlees model with heterogeneous shocks of child tastes and earning abilities. By facing these risks, families decide how much income to generate and how many children to have by considering child-rearing costs. The government aims to provide insurance against the shocks, which are families' private information. To do so, the government designs an optimal tax system which combines income taxes of

<sup>22</sup>See <https://www.census.gov/hhes/families/files/graphics/CH-1.pdf>.

childless families and child tax credits. The sufficient statistics for labor wedges and their relationship with child tax credits are derived.

Income taxes are designed to redistribute from high to low-income families and child tax credits decreases tax liabilities of parents who incur child-rearing costs. The child-rearing costs are crucial inputs on the shape of the child tax credits. The goods cost mostly affect the low-income families and drives the government's motives towards to poor families. On the other hand, time cost is the dominant cost for high-income families and increases provisions for the wealthier. As a result, the credits are U-shaped. Quantitatively, I find that the optimal credits are decreasing especially in the first half of income distribution and are increasing in the rest. In addition, the credit for the second child is less than the credit for the first child, because there is economies of scale in child-rearing.

This paper sheds light on the optimal income taxation including the child benefits for families who have multidimensional private information. I conclude by describing three extensions that I leave for future research. First, the paper abstracts from a dynamic setting. Such a setting can explain how the child benefit should be characterized by the age of the children. Moreover, two heterogeneous risks, the earning abilities and child tastes, can be linked with the age of the parents and, therefore, the effect of optimal taxes on the fertility age can be studied. Second, the paper abstracts from the child quality decision, which is positively correlated with parental time according to [Boca, Flinn, and Wiswall \(2013\)](#). Such a decision can explain why high-income families spend more time with their children (see [Guryan, Hurst, and Kearney \(2008\)](#)). Third, the costs of child-rearing can be endogenous. This endogeneity can help policy makers for designing the optimal provisions via costs. For example, policies that provide a high-quality child care in return of goods might be tempting for high-income families. This extension can also examine the current debate in the US on universal child care provisions for working parents.



## Chapter 2

# Technical Change, Wage Inequality and Taxes

### 2.1 Introduction

Technical change is inherently redistributive, complementing the labor of some whilst substituting for that of others. A large positive literature has analyzed its impact on the wage distribution. This literature has emphasized skill-biased technical change that favors the skilled over the unskilled and, more recently, has stressed the role of technical change in replacing “routine labor” in the middle of the wage distribution. However, while the positive literature documenting the redistributive nature of technical change is extensive, normative work exploring the policy implications of such change is not.<sup>1</sup> Our paper fills this gap. We explore how more than thirty years of technical change in the US has affected the policy recommendations that economic theory provides. Overall, we find that such change creates a rationale for a modest adjustment of optimal policy in a direction that favors middle income earners, reducing their average taxes while lowering transfers to those at the bottom. Optimal marginal taxes are reduced on incomes that are low (but not the lowest) and raised on incomes that are high (but not the highest). Although, the overall effects are moderate, they are the net effect of larger countervailing forces stemming from

---

<sup>1</sup>For a historical account of the relationship between skill and technology see [Goldin and Katz \(1998\)](#), [Autor, Katz, and Krueger \(1998\)](#) and the references therein. [Bresnahan, Brynjolfsson, and Hitt \(2002\)](#) look at firm level evidence connecting technology and the demand for skills. [Autor, Levy, and Murnane \(2003\)](#) argue that recent technical change has led to the replacement of “routine” labor in the middle of the wage distribution. [Autor, Katz, and Kearney \(2006\)](#) and [Goos and Manning \(2007\)](#) document “job polarization”: growth in low and high skill occupations.

technical change. First, such change directly modifies wage differentials across differently talented workers; second it alters the substitutability of talent across occupations and, hence, the sensitivity of wage differentials to taxes. The evolution of optimal policy depends upon the balance of these conflicting forces.

We make theoretical and quantitative contributions. On the theoretical side, we embed a talent-to-task assignment model into an optimal tax framework. The former has been used by labor and trade economists to analyze the implications of technical change for the structure of wages and employment. We show how the technological parameters emphasized in this work shape optimal tax formulas. On the quantitative side, we bring a parametric assignment model to the data; we estimate the key parameters and derive the implications of technical change from the 1970's to the present day for policy.

The normative tax literature largely focusses on the incentive to supply effort by perfectly substitutable and privately informed workers. An exception is [Stiglitz \(1982\)](#) who allows for imperfect substitutability between the effort of two different talents.<sup>2</sup> This assumption renders relative wages sensitive to the profile of effort across talents and, hence, tax policy. In particular, Stiglitz identifies a wage compression motive for subsidizing high and taxing low talents. By doing so the wages of high talents are compressed relative to low and the former's incentive constraints are relaxed. We begin our analysis with a Stiglitz-type environment in which the production function is defined directly over the imperfectly substitutable labor input of many different worker types. In this setting with minimal restriction on the production function, we derive a general formula for optimal taxation. The formula provides a framework for interpreting subsequent results. [Stiglitz \(1982\)](#)'s wage compression channel remains operative, but now takes a more complex form: the motive to tax a given talent type  $k$  at the margin depends, in part, on the elasticity of the relative wages of all pairs of adjacent talent types (ordered by wages) with respect to  $k$ 's effort. This setting suggests two ways in which technical change can influence optimal policy. First, factor augmenting technical change that is biased towards a subset of talents can do so by modifying relative wages and, hence, tightening or relaxing incentive constraints. Second, technical change that alters the effect of one talent type's effort on the relative wages of other talent types impacts policy by strengthening

---

<sup>2</sup>Other important exceptions include [Lockwood, Nathanson, and Weyl \(2014\)](#), [Rothschild and Scheuer \(2013\)](#), [Rothschild and Scheuer \(2014\)](#), [Rothschild and Chen \(2014\)](#) and [Slavík and Yazıcı \(2014\)](#).

or diluting the wage compression channel described above.

We next embed an assignment model into an optimal tax framework.<sup>3</sup> In the class of assignment models we consider talented workers have a comparative advantage in complex tasks and assortative matching of workers to tasks occurs. Such models omit an intensive effort margin, a societal motive for redistribution and explicitly private talent; the optimal tax framework adds these things.<sup>4</sup> In the equilibria of our embedded model, workers sort themselves efficiently across tasks *conditional on the effort of other workers*. This induces an indirect production function over the effort of different talents of the sort that our earlier analysis assumed. Technological parameters that determine relative task demand and the productivity of task-talent matches in the assignment framework are thus mapped to the variables and elasticities necessary for optimal tax analysis. In particular, the pattern of comparative advantage of talents across tasks shapes the sensitivity of relative wages to variations in the effort profile and, hence, policy. A local reduction in marginal taxes that induces a given talent type to increase its effort, depresses the (shadow) price of the task to which the type is assigned and, hence, the type's relative wage. Workers of this type offset this reduction by migrating into neighboring tasks, mitigating the impact on their original task's shadow price. However, the offset is partial since this migration erodes their productivity relative to neighboring talents. The greater is the comparative advantage of talented workers in complex tasks, the greater this erosion and the more sensitive are relative wages to task assignment. Thus, technical change that raises talent-complexity comparative advantage enhances the policymaker's ability to influence the wage structure through taxation. It strengthens the wage compression force identified in the more reduced form Stiglitz setting.<sup>5</sup>

We take our model to the data and quantify the implications of 30 years of technical change in

---

<sup>3</sup>The assignment framework originated with [Roy \(1950\)](#). Versions with a continuum of tasks, single dimensional talent and comparative advantage of talented workers in complex tasks were developed by [Sattinger \(1975\)](#) and [Teulings \(1995\)](#). Such models have proven to be a rich laboratory for analyzing the role of task-talent distributions and the productivity of task-talent matches in shaping the wage distribution. Recently, these models have been used to explore the implications of technical change that attaches to tasks (rather than talents), see [Costinot and Vogel \(2010\)](#), [Acemoglu and Autor \(2011\)](#) and [Autor and Dorn \(2013\)](#).

<sup>4</sup>[Rothschild and Scheuer \(2013\)](#) were the first to consider the optimal tax implications of an assignment model. They do so in the context of a Roy model, i.e. a model with two sectors and no explicit notion of comparative advantage. We elaborate below on the differences between our and their focus and approach.

<sup>5</sup>Migration of workers into neighboring tasks depresses the shadow prices of these tasks inducing the talents occupying them to migrate as well. A ripple effect is created and, so, an adjustment in one talent type's effort can induce reassignment of many types, affecting their relative wages and in the process relaxing and tightening many incentive constraints. However, the greater is talent-complexity comparative advantage the more contained the impact of a policy-induced effort adjustment.

the US for optimal policy. We treat information on occupations, incomes and hours in the Current Population Survey (CPS) as if it was generated by an equilibrium of our assignment model and use parametric assumptions and equilibrium restrictions to recover estimates of key technological parameters for the 1970's and the 2000's. To relate empirical occupations to the ordered set of tasks in our model, we order the former by the average wage paid. We recover an empirical proxy for the assignment of tasks to talents from the distribution of workers across occupations (ordered by wages). The estimation of parameters determining the demand for tasks is separated from those determining the productivity of task-talent matches by assuming a Cobb-Douglas technology for final goods as a function of tasks. This enables us to identify the demand parameters with occupational compensation shares. Parameters determining the productivity of talent-task matches and, hence, comparative advantage are recovered from the empirical assignment function and the distribution of wages across tasks using the envelope condition for wages implied by the model. After obtaining these estimates and supplementing them with calibrated preference parameters, we calculate optimal tax policies for the 1970's and 2000's.

We find evidence of relative reductions in demand for mid-level tasks and relative increases in demand for low and high level tasks. We also find evidence of a twisting of the talent-task productivity function, with low talent productivity catching up to high talent in simple tasks and falling behind in more complex ones. The latter is associated with significant increases in the comparative advantage of more talented workers in more complex tasks. Moving from the 1970's to the 2000's, we find that under our benchmark estimation/parameterization, optimal marginal tax rates rise at the very bottom of the income distribution, fall on low to middle level incomes, rise on higher ones before falling again at the very top of the income distribution. This change in policy favors those in the middle of the income distribution who pay lower average taxes; optimal transfers to workers at the first and second income deciles are reduced. The twisting of the productivity function is the main force at work. It has two effects. First, it suppresses wage variation at the bottom of the income distribution, while enhancing it at the top. This relaxes incentive constraints on low incomes, while tightening them on high ones; it is a force for reductions in optimal marginal taxes on the former and increases on the latter. These effects are slightly enhanced by the relative reduction of demand for mid-level tasks populated by mid-level talents. Second, there is a partially offsetting strengthening of the wage compression channel. Higher comparative advantage of talented



workers in complex tasks increases the policymaker’s motive to apply high marginal taxes on low talents. Such taxes deter low talent effort, raise low-level task prices and encourage higher talents into these tasks. The relative productivity of these task migrants is eroded, suppressing their wage premia and relaxing incentive constraints. A parallel strengthening of the policymaker’s motive to reduce marginal taxes on high talents occurs. Policy depends on the balance of these two forces. The first dominates at most incomes under our benchmark parametrization (except those in the extreme tails), but since the second dampens the first, the overall effect is modest.

The equilibrium of our baseline model does not exhibit intra-task wage dispersion or the payment of the same wage in multiple tasks (“wage overlap”). Thus, it cannot capture the policy implications stemming from these. At the end of the paper, in Section 2.7 (with details and elaboration in Appendix B.7), we describe an extension that permits non-degenerate and overlapping supports for intra-task wage distributions. This extension incorporates a second talent dimension, which impacts absolute advantage alone. We find that our results concerning the implications of technical change for policy are qualitatively robust to, but quantitatively dampened by this extension. We use it to obtain a lower bound on the responsiveness of policy to technical change.

The remainder of the paper proceeds as follows. After a brief literature review, Section 2.2 provides motivating facts. Section 2.3 gives optimal tax formulas for economies with imperfectly substitutable labor types and provides an initial discussion of the implications of technical change for policy. In Section 2.4 an assignment model is embedded into an optimal tax framework. An indirect production function over worker effort is derived and the parameters of the assignment model related to the relevant terms of the optimal tax formulas from Section 2.3. In addition, the implications of technical change for policy in a simple two talent model are discussed. Section 2.5 describes how the assignment model is used to identify estimates of technical change and reports these estimates. In Section 2.6, optimal policy for the 1970s and 2000s is computed and the implications of technical change for policy recovered. The tax formula from Section 2.3 is used to decompose and account for changes to optimal taxes. Section 2.7 describes a model extension that can accommodate intra-task wage variation; Section 2.8 concludes. Appendices contain proofs, robustness checks and extensions.

LITERATURE A contribution of our paper is to bring together the normative optimal taxation literature and a positive literature that analyzes the impact of technical change on the wage distribution. Both literatures are large. Many contributions to the latter have attributed increases in the skill premium to skill-biased technical change, formalizing this insight in what [Acemoglu and Autor \(2011\)](#) have called the “canonical model”, i.e. a model with imperfectly substitutable skilled and unskilled workers and factor-augmenting technical change directed towards the skilled.<sup>6</sup> Recently, a more nuanced view of the labor market has emerged that emphasizes growth in low and high wage occupations relative to those in the middle. It has spurred the development of assignment models that endogenize the joint distribution of workers across wages and occupations and in which technical change attaches to tasks rather than worker types. Examples include [Acemoglu and Autor \(2011\)](#) and [Autor and Dorn \(2013\)](#).

Most contributions to the normative literature focus on the incentive to supply labor in environments with privately known talent and perfectly substitutable labor. [Stiglitz \(1982\)](#) was the first to introduce imperfectly substitutable labor into such a setting. [Rothschild and Scheuer \(2013\)](#) (extended in [Rothschild and Scheuer \(2014\)](#)) were the first to introduce assignment.<sup>7</sup> [Rothschild and Scheuer](#) show that a worker’s ability to select her task mutes the regressivity of optimal taxes found by Stiglitz. They also show that optimal tax formulas are substantially complicated by additional terms stemming from wage overlap. The focus in [Rothschild and Scheuer \(2013\)](#) is on economies with two tasks and two dimensional talents.<sup>8</sup> In contrast, our baseline assignment model features a continuum of tasks and one dimensional talent. In our model a more talented worker is better at everything, but especially good at some things, with those things interpreted as more complex tasks. The restriction to one dimensional talents follows a tradition in labor economics initiated by [Sattinger \(1975\)](#) and adopted recently by the positive literature described above. Its adoption allows us to make contact with these recent contributions, to formulate the notions of talent and task complexity in a parsimonious way and to develop a strategy for bringing our model to the data. It permits a significant simplification of the tax formula in [Rothschild and Scheuer \(2013\)](#) (via the omission of wage overlap) and leads us to adopt a substantially different approach to an-

---

<sup>6</sup>Examples include [Acemoglu \(2002\)](#) and [Krusell, Ohanian, Ríos-Rull, and Violante \(2000\)](#).

<sup>7</sup>Also related is [Rothschild and Chen \(2014\)](#). In contrast to [Rothschild and Scheuer \(2013\)](#) and similar to us, this paper considers a model with a finite number of talents. It illustrates the difficulties of applying the method of [Rothschild and Scheuer \(2013\)](#) in such settings.

<sup>8</sup>This is generalized in [Rothschild and Scheuer \(2014\)](#) to  $K$  tasks and  $K$  dimensional talent.

alyzing the problem than that in [Rothschild and Scheuer \(2013\)](#). However, it cannot accommodate intra-task wage dispersion or wage overlap. In Section 2.7 and Appendix B.7 we provide an extension that can. [Lockwood et al. \(2014\)](#) also integrate tax considerations into an assignment setting. They focus on the externalities associated with certain assignments and characterize the optimal structure of corrective Pigouvian taxation.<sup>9</sup> We abstract completely from this tax motive.

[Slavík and Yazıcı \(2014\)](#) apply the logic of [Stiglitz \(1982\)](#) to capital taxation. In their paper they introduce two sorts of capital: buildings and machines. Following the skill premium literature, they assume a machine-skill (or machine-talent) complementarity. Thus, machines raise the marginal product of the talented relative to the untalented and, as in [Stiglitz \(1982\)](#), this dilutes incentives. It is socially desirable to deter the accumulation of machines. In quantitative work, [Slavík and Yazıcı \(2014\)](#) show that this creates a rationale for quite high rates of (machine) capital taxation. [Slavík and Yazıcı \(2014\)](#)’s contribution is complementary to ours. They endogenize technical change in the context of a two talent “canonical model” and develop policy implications. We treat technical change parametrically, but do so in a multi-talent/multi-task assignment setting.

[Heathcote, Storesletten, and Violante \(2014\)](#) analyze optimal income tax progressivity in a rich dynamic environment. They assume imperfectly substitutable skills, but do not explicitly model tasks. Our model is static, but we add assignment and, hence, endogenize the substitutability of skills and relate it to technical change. In addition, [Heathcote et al. \(2014\)](#) restrict optimal taxes to a parametric class, we do not.

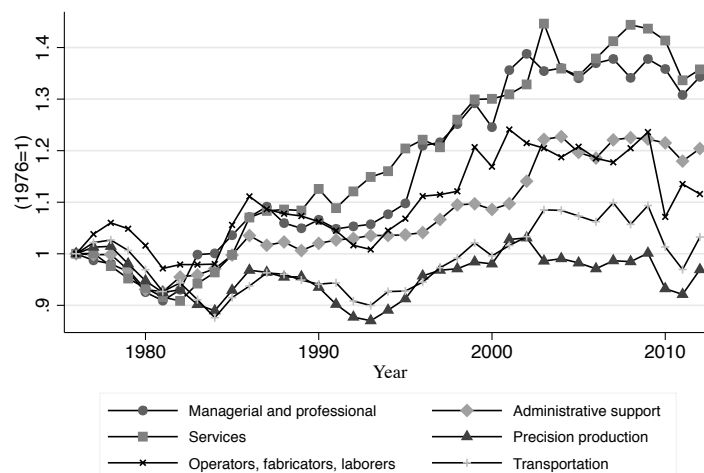
## 2.2 Evolution of the Occupational Wage Distribution: Stylized Facts

We first document some stylized facts that motivate our analysis. Figure 2.1 displays changes in average incomes across (1-digit) occupations from the 1970’s to the present.<sup>10</sup> The figure indicates considerable variation in the experience of different occupations, with some exhibiting significant average income growth and others stagnating. Moreover, occupations with slow average income growth were predominantly middle income in the 1970’s, while fast growers were mainly low or high income at that time. For example, precision production, craft and repair workers had a mid-

---

<sup>9</sup>[Rothschild and Scheuer \(2014\)](#) also incorporate this motive into their theoretical work.

<sup>10</sup>The data is taken from the March survey of the Current Population Survey (CPS). See Appendix B.4 for additional details on the data and our sample selection.



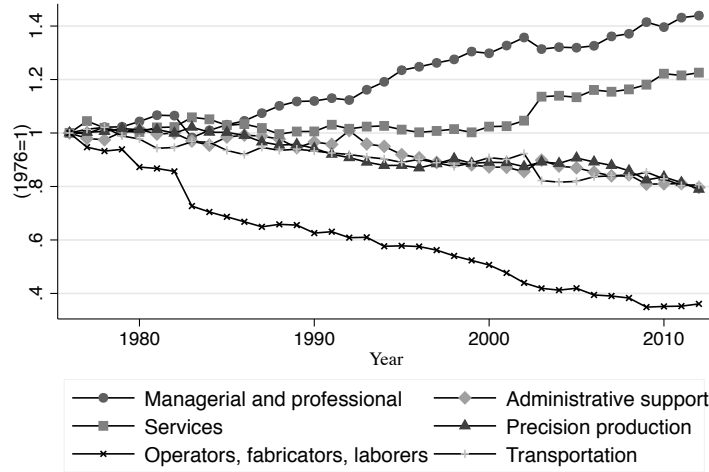
Source: March CPS.

Figure 2.1: Evolution of average income by occupation over time.

level income of \$33,109 in 1975 (all incomes are expressed in 2005 dollars) and negligible income growth subsequently. In contrast, the two occupations with the fastest growing average incomes, services and managerial and professional, had average incomes in the mid-1970's of \$12,912 and \$40,013, placing them at opposite extremes of the income distribution. Such occupational polarization, with the middle growing more slowly than the extremes, is not confined to earnings; it is also present in various measures of occupational size and demand. Figure 2.2, displays changes in the share of employment of different occupations over time.<sup>11</sup> Here managerial/professional and service related occupations that are concentrated in the extremes of the income distribution are expanding in size, while mid-income level occupations operators and fabricators (mostly employed in manufacturing) are shrinking over time.

Overall, the picture that emerges from the CPS (and other data sources) is one in which high wage and low wage occupations are growing in size and in average compensation relative to middle ones. If talent is imperfectly substitutable across occupations, then these varied occupational fortunes suggest varied fortunes for differently-talented workers. In the remainder of the paper we consider the optimal policy response to such events.

<sup>11</sup>See, inter alia, [Goos and Manning \(2007\)](#), [Acemoglu and Autor \(2011\)](#) and [Autor and Dorn \(2013\)](#) for related evidence.



Source: March CPS.

Figure 2.2: Evolution in size of employment by occupation over time.

## 2.3 Taxation with Imperfectly Substitutable Workers

Mirrlees (1971)'s model of optimal taxation assumes that workers of different types are perfect substitutes and that final output is a weighted sum (or integral) of worker efforts, with the weights given by private productivities. Stiglitz (1982) allows for a more general production function. He assumes that workers are one of two imperfectly substitutable types and interprets these types as "low" and "high" skilled. In this section, we generalize Stiglitz (1982) to  $K$ -types, but place no interpretation on a worker's type (the nature of which is defined implicitly by the production function). In this (and compared to later sections reduced form) context we discuss implications of technical change for taxes.<sup>12</sup>

### 2.3.1 Physical Environment

**WORKERS** A continuum of workers has identical preferences over consumption  $c \in \mathbb{R}_+$  and effort  $e \in [0, \bar{e}]$  described by a utility function  $U : \mathbb{R}_+ \times [0, \bar{e}] \rightarrow \mathbb{R}$ . The function  $U$  is assumed to be concave, twice continuously differentiable on the interior of its domain, with for each  $e \in [0, \bar{e}]$ ,  $U(\cdot, e)$  increasing and for each  $c \in \mathbb{R}_+$ ,  $U(c, \cdot)$  decreasing and strictly concave. First and

<sup>12</sup>Much of the optimal tax literature is cast in terms of a continuum of types. This literature maintains the linear production function assumption. Although versions of the results that we give below are available for continuum economies, for general constant returns to scale production functions, their derivation requires leaving the framework of optimal control and maximizing an infinite-dimensional Lagrangian directly. To avoid technical complications that do not generate additional economic insight we do not do this.

second partial derivatives of  $U$  are denoted  $U_x$  and  $U_{xy}$  with  $x, y \in \{c, e\}$ .  $U$  satisfies the Inada conditions: for all  $c > 0$ ,  $\lim_{e \downarrow 0} U_e(c, \cdot) = 0$  and  $\lim_{e \uparrow \bar{e}} U_e(c, \cdot) = -\infty$ . In addition,  $U$  satisfies the Spence-Mirrlees single crossing property:  $-U_e(c, y/w) / \{w U_c(c, y/w)\}$  is decreasing in  $w$ . Workers are partitioned across a finite number of “types”  $K \geq 2$  with a fraction  $\pi_k$  of workers being of type  $k \in \{1, \dots, K\}$ . The fraction of workers with type less than or equal to  $k$  is denoted  $\Pi_k = \sum_{j=1}^k \pi_j$ .

Workers sell their labor to firms and pay taxes on the income that they earn. Let  $T : \mathbb{R}_+ \rightarrow \mathbb{R}$  denote an income tax function.<sup>13</sup> A worker of type  $k$  receiving wage  $w_k$  solves the problem:

$$\max_{\mathbb{R}_+ \times [0, \bar{e}]} U(c, e) \quad \text{s.t.} \quad c \leq w_k e - T(w_k e). \quad (2.1)$$

**TECHNOLOGY** A representative competitive firm hires workers of all types. The firm uses a production function  $F : \mathbb{R}_+^K \rightarrow \mathbb{R}_+$  defined directly on the labor inputs of the different types. The firm solves:

$$\max_{\mathbb{R}_+^K} F(e_1 \pi_1, \dots, e_K \pi_K) - \sum_{k=1}^K w_k \pi_k e_k,$$

where  $e_k$  is the common effort level of workers of type  $k$ .  $F$  is assumed to be a continuously differentiable, constant returns to scale function with  $k$ -th partial derivative  $F_k$ . At this stage, we place no further restrictions on  $F$ . In classical Mirrlees models  $F(e_1 \pi_1, \dots, e_K \pi_K) = \sum_{k=1}^K a_k e_k \pi_k$  for some positive constants  $\{a_k\}$  and workers of different types are perfectly substitutable. However, we allow for and focus upon worker types that are imperfect substitutes in production. Since  $F$  defines what it means for a worker to be of one type or another, the economic nature of a worker’s type is for the moment left implicit.<sup>14</sup>

**TAX EQUILIBRIUM** Let  $G \in \mathbb{R}_+$  be a fixed public spending amount. Given  $G$ , a *tax equilibrium* is an income tax function  $T : \mathbb{R}_+ \rightarrow \mathbb{R}$ , an allocation  $\{c_k, e_k\}_{k=1}^K$  and a wage profile  $\{w_k\}_{k=1}^K$  such that (i) for each  $k = 1, \dots, K$ ,  $(c_k, e_k)$  solves (2.1), (ii) for each  $k = 1, \dots, K$ ,  $w_k = F_k(e_1 \pi_1, \dots, e_K \pi_K)$  and (iii) the goods market clearing condition holds:  $G + \sum_{k=1}^K c_k \pi_k \leq F(e_1 \pi_1, \dots, e_K \pi_K)$ . Let  $\mathcal{E}$  denote the set of tax equilibria (given  $G$ ), which we take to be non-empty.

<sup>13</sup>We restrict attention to non-stochastic tax functions. See Hellwig (2007) for sufficient conditions for such mechanisms to be socially optimal in utilitarian settings.

<sup>14</sup>The firm’s problem determines relative levels of efforts across types. The scale of the representative firm is determined in equilibrium.

### 2.3.2 Optimal Policy

A government attaches Pareto weight  $g_k$  to workers of type  $k$ , with weights normalized to satisfy  $\sum_{k=1}^K g_k = 1$ . It selects a tax equilibrium to solve:

$$\sup_{\mathcal{E}} \sum_{k=1}^K U(c_k, e_k) g_k. \quad (\text{PP})$$

Let  $T^*$  and  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  denote an optimal tax equilibrium. Define the corresponding (optimal) marginal tax rate at income  $q_k^* := w_k^* e_k^* > 0$  to be:<sup>15</sup>

$$\tau_k^* = 1 + \frac{U_e(c_k^*, e_k^*)}{w_k^* U_c(c_k^*, e_k^*)}.$$

To characterize optimal tax equilibria, we follow the conventional procedure of recovering optimal allocations from a mechanism design problem. Subsequently, prices and (optimal) taxes are determined to ensure implementation of this allocation as part of a tax equilibrium. The mechanism design problem associated with (PP) is:

$$\sup_{\{c_k, e_k\}_{k=1}^K \in \{\mathbb{R}_+ \times [0, \bar{e}]\}^K} \sum_{k=1}^K U(c_k, e_k) g_k \quad (\text{MDP})$$

s.t. for each  $k, j \in \mathcal{K} := \{(l, m) \in \{1, \dots, K\}^2, l \neq m\}$ ,

$$\eta_{k,j} : \quad U(c_k, e_k) \geq U\left(c_j, \frac{F_j(e_1 \pi_1, \dots, e_K \pi_K)}{F_k(e_1 \pi_1, \dots, e_K \pi_K)} e_j\right) \quad (2.2)$$

and

$$\chi : \quad F(e_1 \pi_1, \dots, e_K \pi_K) \geq G + \sum_{k=1}^K c_k \pi_k. \quad (2.3)$$

In (MDP) the government selects a report-contingent allocation of consumption and effort  $\{c_k, e_k\}_{k=1}^K$  that induces each worker to truthfully report its type  $k$  and produce the associated income  $q_k = F_k(e_1 \pi_1, \dots, e_K \pi_K) e_k$ . Incentive constraints that ensure the optimality of truthful reporting are given in (2.2) with corresponding Lagrange multipliers  $\eta_{k,j}$ . If type  $k$  claims to be of type  $j$  she

<sup>15</sup>(PP) does not uniquely determine  $T^*$ . However,  $T^*$  may be chosen to be directionally differentiable in which case:  $\partial T_-^*(q_k^*) \leq \tau_k^* \leq \partial T_+^*(q_k^*)$ , where  $\partial T_-^*(q_k^*)$  and  $\partial T_+^*(q_k^*)$  are left and right derivatives of  $T^*$  at  $q_k^* > 0$ . If  $T^*$  is (chosen to be) differentiable at  $q_k^*$ , then its derivative at that point equals  $\tau_k^*$ .

must reproduce the corresponding income  $q_j = F_j(e_1\pi_1, \dots, e_K\pi_K)e_j$ . The effort cost to her of doing so is  $\frac{F_j(e_1\pi_1, \dots, e_K\pi_K)}{F_k(e_1\pi_1, \dots, e_K\pi_K)}e_j$ . Thus, the  $(k, j)$ -th incentive constraint (2.2) depends upon the *entire profile* of worker efforts via the  $(k, j)$ -th shadow wage ratio. We refer to a  $(k, j)$ -incentive constraint as local if  $j = k - 1$  or  $j = k + 1$ , local downwards if  $j = k - 1$  and local upwards if  $j = k + 1$ . The final restriction (2.3) in (MDP) is the resource constraint with corresponding multiplier  $\chi$ .

Towards understanding how technical change shapes policy, we give a proposition that relates optimal taxes to  $F$ . This proposition is a consequence of a more general result given in the Appendix. In the latter, we show that when worker types are ordered consistently with optimal wages and incomes, then only local downwards  $(k, k - 1)$  or upwards  $(k, k + 1)$  incentive constraints bind. In the main text we follow the common convention of assuming that only the former are binding and then verifying this assumption in numerical calculations.<sup>16</sup> To state the proposition (and its generalization in the appendix) it is convenient to re-express the constraints in (MDP) in the form  $\mathcal{G}(\{c_k, e_k\}_{k=1}^K) \geq 0$ , where  $\mathcal{G} : \mathbb{R}_{++}^{2K} \rightarrow \mathbb{R}^{K(K-1)+1}$  combines the constraint functions from (2.2) and (2.3). Problem (MDP) satisfies a (Mangasarian-Fromowitz) constraint qualification at  $\{c_k, e_k\}_{k=1}^K \in \mathbb{R}_{++}^{2K}$  if there is an  $x \in \mathbb{R}^{2K}$  such that  $\nabla \mathcal{G}(\{c_k, e_k\}_{k=1}^K)x < 0$ , where  $\nabla \mathcal{G}(\{c_k, e_k\}_{k=1}^K)$  is the Jacobian of  $\mathcal{G}$  at  $\{c_k, e_k\}_{k=1}^K$ . Let  $\eta_{k,j}^*$  and  $\chi^*$  denote the optimal (Karush-Kuhn-Tucker) multipliers associated with the incentive and resource constraints. Finally, let  $\Delta_e U_c(c', e'; \delta) := \frac{U_c(c', e' + \delta) - U_c(c', e')}{\delta}$  denote a finite difference approximation to the derivative of  $U_c$  with respect to  $e$  at  $(c', e')$  and define  $\Delta_e U_e$  analogously.

**Proposition 2.** *Let  $T^*$  and  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  denote an optimal tax equilibrium with worker types indexed so that  $w_k^* = F_k(e_1^*\pi_1, \dots, e_K^*\pi_K)$  is non-decreasing in  $k$ . Assume that  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  is interior (i.e. in  $\mathbb{R}_{++}^{2K}$ ), that  $\mathcal{G}$  satisfies the constraint qualification at  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  and that the local upwards incentive constraints are non-binding, i.e.:*

$$U(c_k^*, e_k^*) > U(c_{k+1}^*, q_{k+1}^*/w_k^*), \quad \text{where } q_{k+1}^* := w_{k+1}^* e_{k+1}^*. \quad (\text{NUIC})$$

<sup>16</sup>A general formula with possibly binding upwards incentive constraints is supplied in the Appendix.



Optimal tax rates then satisfy:

$$\frac{\tau_k^*}{1 - \tau_k^*} = \underbrace{\frac{1 - \Pi_k}{\pi_k} \frac{\Delta w_{k+1}^*}{w_{k+1}^*} \Psi_k^* \mathcal{H}_k^*}_{\text{Mirrlees}} + \underbrace{\sum_{j=1}^{K-1} \mathcal{M}_{k,j}^* \phi_{k,j}^*}_{\text{Wage Compression}}, \quad (2.4)$$

where  $\Delta w_k^* := w_k^* - w_{k-1}^*$ ,

$$\Psi_k^* := \sum_{j=k}^{K-1} \mathcal{N}_{k,j}^* \left\{ 1 - \frac{g_{j+1} U_c(c_{j+1}^*, e_{j+1}^*)}{\pi_{j+1} \chi^*} \right\} \left( \frac{U_c(c_k^*, e_k^*)}{U_c(c_{j+1}^*, e_{j+1}^*)} \right) \frac{\pi_{j+1}}{1 - \Pi_k},$$

with  $\mathcal{N}_{k,j}^* := \prod_{i=k+1}^j \frac{U_c(c_i^*, q_i^*/w_{i+1}^*)}{U_c(c_i^*, e_i^*)}$ ,

$$\mathcal{H}_k^* := - \frac{\Delta_e U_c(c_k^*, e_k^*; q_k^*/w_{k+1}^* - e_k^*)}{U_c(c_k^*, e_k^*)} e_k^* + \frac{\Delta_e U_e(c_k^*, e_k^*; q_k^*/w_{k+1}^* - e_k^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_j^*} e_k^* + 1,$$

$\mathcal{M}_{k,j}^* := \frac{U_c(c_k^*, e_k^*)}{U_e(c_k^*, e_k^*)} \frac{U_e(c_j^*, q_j^*/w_{j+1}^*)}{U_c(c_j^*, e_j^*)} \frac{q_j^*}{w_{j+1}^*} \frac{1 - \Pi_j}{\pi_j} \frac{\pi_j}{\pi_k} \Psi_j^*$  and cross relative wage elasticities  $\phi_{k,j}^* := \frac{e_k^*}{w_{j+1}^*/w_j^*} \frac{\partial w_{j+1}^*/w_j^*}{\partial e_k} (e_1^*, \dots, e_K^*)$ .

*Proof.* See the Appendix. □

The right hand side of the optimal tax formula (2.4) is the sum of two terms, which we label “Mirrlees” and “Wage Compression”.

**MIRRLEES TERM** The Mirrlees term in (2.4) is quite standard in optimal tax analyses. We very briefly review and interpret its four components.<sup>17</sup>  $\mathcal{H}_k^*$  is a discrete approximation to  $\frac{1 + \mathcal{E}_{u,k}}{\mathcal{E}_{c,k}}$ , where  $\mathcal{E}_{c,k}$  and  $\mathcal{E}_{u,k}$  are, respectively, the compensated and uncompensated labor supply elasticities at  $(c_k^*, e_k^*)$ . If worker preferences are additively separable, this reduces to one plus (a discrete approximation to) the reciprocal of the Frisch elasticity. Incentive-compatibility considerations require that if worker type  $k$  receives an increment in consumption all higher types  $j = k + 1, k + 2, \dots, K$  receive an increment in utility sufficient to deter them from reporting a lower type.  $\Psi_k^*$  captures the net societal cost of such a redistribution; it weighs the cost of extracting resources from the population at large against the benefits of raising the welfare of higher income types.  $\frac{1 - \Pi_k}{\pi_k}$  is the reciprocal of the type hazard. This plays an important role in conventional optimal tax analysis

<sup>17</sup>For detailed discussion of these components in a continuous-type setting see [Salanié \(2011\)](#).

since, if types have compact support (as in the current finite setting), it implies zero marginal taxes at the maximal income. However, it is unaffected by technical change and, thus, is less central to our analysis. In contrast, the wage growth (across types) term  $\Delta w_{k+1}^*/w_{k+1}^*$  is endogenous and important in what follows. To understand its role consider the local downwards incentive constraint:

$$U(c_{k+1}, e_{k+1}) \geq U(c_k, e_k w_k / w_{k+1}) \quad (2.5)$$

As noted previously, the wage ratio  $w_{k+1}/w_k$  appears on the right hand side of this inequality. Higher values of this ratio reduce the effort that a  $k+1$ -th type worker must exert to mimic a  $k$ -type. Consequently, they tighten the incentive constraint and lead to greater distortions of allocations. Higher wage growth across the  $k$  and  $k+1$  types is, other things equal, a force for higher marginal taxes on the  $k$ -th type.<sup>18</sup>

**WAGE COMPRESSION TERM** The second term in (2.4) does *not* appear in standard optimal tax equations that are derived from models with linear production functions and exogenous wages. In settings with non-linear production functions, such as ours, the effort of the  $k$ -th worker type can affect the marginal rate of transformation and, hence, the ratio of wages between the  $j$  and  $j+1$ -th types. Following the logic of the previous paragraph, more compressed wage ratios relax incentive constraints and to the extent that the effort of a given type enhances such compression it should be encouraged through taxation. Conversely, larger values of the cross relative wage-effort elasticities  $\phi_{k,j}^*$  imply that wage differentials are increasing in the effort of the  $k$ -th type and, hence, this type's effort should be deterred via higher marginal taxes. [Stiglitz \(1982\)](#) identifies this wage compression channel in a two type model. In that case there is only one binding incentive constraint and  $-\phi_{1,1}^* = \phi_{2,1}^* = 1/\mathcal{E}^*$ , where  $\mathcal{E}^*$  is the elasticity of substitution between the two worker types (i.e.  $\frac{\frac{w_2}{e_2}}{\frac{w_1}{e_1}} \frac{\partial e_2}{\partial \frac{w_2}{e_2}} \frac{e_1}{w_1}$ ) at the optimum. Assuming this is positive, compression of wages between the two types, requires that the effort of the high (resp. low) type should be relatively encouraged (resp. discouraged).

---

<sup>18</sup>The terms  $\frac{1-\Pi_k}{\pi_k}$  and  $\frac{\Delta w_{k+1}^*}{w_{k+1}^*}$  may be consolidated as:  $\frac{1-\Pi_k}{\pi_k} \frac{\Delta w_{k+1}^*}{w_{k+1}^*}$ . In the continuous limit the latter reduces to  $\text{Haz}(w) = \frac{1-\Xi(w)}{\zeta(w)w}$ , where  $\Xi$  and  $\zeta$  are the wage distribution and density functions and, following the usage of [Saez \(2001\)](#),  $\text{Haz}(w)$  is the wage hazard ratio. In the continuous setting, the impact of a change in wage growth across types  $\frac{1}{w^*(k)} \frac{\partial w^*}{\partial k}(k)$  on marginal taxes may be understood via its impact on  $\text{Haz}$ . Specifically, an increase in  $\frac{\partial w^*}{\partial k}(k)$  reduces  $\zeta(w^*(k))$  (the “fraction” who will be distorted by a marginal tax) relative to  $1 - \Xi(w^*(k))$  (the fraction who will be undistorted and will pay higher average taxes). It is, therefore, a force for higher marginal taxes at  $w^*(k)$ .

Since the first term in (2.4) is zero for  $k = K = 2$ , this translates into an optimal marginal income subsidy for high types and an enhanced marginal income tax for low types.

**THE FORM OF  $F$**  The functional form for  $F$  plays an important role in shaping wage growth across types  $\Delta w_{k+1}^*/w_{k+1}^*$ , the cross relative wage-effort elasticities  $\phi_{k,j}^*$  and, hence, optimal taxes. If  $F$  is a weighted sum of type efforts, as in the classical Mirrlees model, then  $\Delta w_{k+1}^*/w_{k+1}^*$  is treated as structural and invariant to policy, while each  $\phi_{k,j}^*$  is set to zero. A more general alternative is to allow  $F$  to be a CES function.<sup>19</sup> This assumption permits policy to affect  $\Delta w_{k+1}^*/w_{k+1}^*$ , but continues to treat the elasticity of substitution and, hence, the relative wage-effort elasticities as structural. It also places strong restrictions on the latter requiring that they equal:

$$\phi_{k,j}^* = \begin{cases} -\frac{1}{\mathcal{E}} & j = k \\ \frac{1}{\mathcal{E}} & j = k - 1 \\ 0 & \text{otherwise,} \end{cases} \quad (2.6)$$

where  $\mathcal{E}$  is the elasticity of substitution between the effort of worker type pairs. Thus, for each worker type  $k$ , the elasticities  $\phi_{k,j}^*$  are non-zero only locally (i.e. a variation in a type's effort only affects its wage relative to others, it does not affect the relative wage of other type pairs) and all elasticities  $\phi_{k,k}^*$  and  $\phi_{k,k-1}^*$  take common values independent of  $k$ . These features have led to some resistance amongst labor and public finance economists to the use of CES production functions in modeling labor demand. For example, [Salanié \(2011\)](#) asserts: "It is, unfortunately, quite difficult to specify a production function that models the limits to factor substitution with an infinite number of factors." (Chapter 4, p.111). He emphasizes that the substitutability of similar and dissimilar worker types may be quite different, but that such differences cannot be accommodated under the CES assumption.

**TOWARDS AN ASSIGNMENT ECONOMY** [Rothschild and Scheuer \(2013\)](#) and its generalization [Rothschild and Scheuer \(2014\)](#) consider assignment economies in which workers choose tasks as well as effort. In their setting a worker's multidimensional type gives his or her productivity in all

---

<sup>19</sup>See [Heathcote et al. \(2014\)](#) for an analysis of optimal taxation within a special class of tax functions that makes such an assumption in a rich dynamic setting.

tasks. If, in the latter paper, the distribution over worker types places all mass on types that have positive productivity in only one task, then it reduces to a  $K$ -type Stiglitz economy with workers effectively “locked” into particular tasks. Consequently, results similar to Proposition 2 would emerge in Rothschild and Scheuer (2014) under this restriction. Below, however, we show that Proposition 2 is much more generally applicable. In particular, in Section 2.4, we use an assignment framework to micro-found a production function  $F$  defined directly over worker efforts. In this setting workers are not locked into tasks. Instead, the allocation of workers to tasks is efficient given worker effort and  $F$  is the upper envelope to a family of production functions indexed by worker task choices. Proposition 2 is applicable and, importantly, relative wage-effort elasticities  $\phi_{k,j}^*$  are influenced by policy and are no longer structural.

**TECHNICAL CHANGE** The formulas in Proposition 2 point to several channels through which technical change can influence optimal policy. Most simply, if technical change raises the return to effort of all workers equally at a given effort profile, then it does not directly affect wage growth over types  $\Delta w_{k+1}^*/w_k^*$  or the responsiveness of relative wages to effort  $\phi_{k,j}^*$  in (2.4). Such “type-neutral” technical change impacts policy only insofar as it affects labor supply elasticities and relative marginal rates of substitution across workers.<sup>20</sup> If, on the other hand, technical change augments the effort of a subset of workers, then, in general, it does affect wage growth over types. Specifically, if  $F$  is a CES function of the form  $F(e_1\pi_1, \dots, e_K\pi_K) = A[\sum_{k=1}^K D_k e_k^{\frac{\varepsilon-1}{\varepsilon}}]^{\frac{\varepsilon}{\varepsilon-1}}$ , then:

$$\frac{\Delta w_{k+1}}{w_{k+1}} \approx -\log\left(\frac{w_k}{w_{k+1}}\right) = -\log\left(\frac{D_k}{D_{k+1}}\right) + \frac{1}{\varepsilon} \log\left(\frac{e_k}{e_{k+1}}\right) \quad (2.7)$$

and technically induced variations in the log relative CES weights  $\{\log \frac{D_k}{D_{k+1}}\}$  additively translate the map from efforts to wage growth over types. Such variations, by modifying the productivity of one type of worker relative to another at a given effort profile, relax or tighten incentive constraints and, hence, elicit an optimal tax response. They do not affect the responsiveness of relative wages to effort, i.e. they leave the elasticities  $\phi_{k,j}$  unaltered (at the fixed values given in (2.6)). For more general production functions (such as the induced  $F$  in the next section), technical change can influence the sensitivity of wages to the effort profile as well. In particular, by reducing substitutability between skills, technical change can enhance the impact of variations in relative

<sup>20</sup>In our later numerical work, we shut this channel down by restricting to utility functions:  $\log c + h(e)$ .

labor supplies on relative wages and, hence, the policymaker's influence over the wage distribution. This strengthens the wage compression motive and is a further channel via which technical change can influence optimal policy (and a channel that is absent under the CES specification).

## 2.4 Taxation, Assignment and Technical Change

We now consider optimal taxation in a framework with task assignment. As noted in the introduction, assignment-based frameworks have been used in the positive literature to formalize the impact of technical change on the distribution of workers across wages and occupations. As we show below they imply and, hence, micro-found an indirect production function over worker efforts. Consequently, we are able to relate key elasticities in the optimal tax equation (2.4) to deeper structural parameters that describe the relative demand for tasks and the way in which tasks and talent interact. We interpret changes in these parameters as technical change and conclude this section by deriving implications of such change for optimal policy in a very simple assignment model.

### 2.4.1 Physical Environment

As before workers are partitioned across types  $1, \dots, K$  with a fraction  $\pi_k$  being of type  $k$ . Types are now explicitly identified with talents. In addition, there is a continuum of tasks  $v \in [\underline{v}, \bar{v}]$  differentiated by complexity. Workers can choose which task to work (exert effort) in; they cannot work in multiple tasks. They face a schedule of task-specific wages  $\omega : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$ , with  $\omega(v)$  the wage per unit of effective labor paid in task  $v$ . A worker of talent  $k$  has productivity  $a_k(v) \in \mathbb{R}_+$  in task  $v$ . If she exerts effort  $e$  in this task her effective labor is  $a_k(v)e$  and her income is  $\omega(v)a_k(v)e$ . The worker chooses her consumption, effort and task to solve:

$$\sup_{\mathbb{R}_+ \times [0, \bar{e}] \times [\underline{v}, \bar{v}]} U(c, e) \quad \text{s.t.} \quad c \leq \omega(v)a_k(v)e - T(\omega(v)a_k(v)e). \quad (2.8)$$

The productivity functions  $\{a_k\}$ ,  $a_k : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$ , play a key role in the subsequent analysis. The following condition is imposed upon them.

**Assumption 1.** *The functions  $a_k : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$ ,  $k \in \{1, \dots, K\}$  are continuous and satisfy (i) (Weak*

comparative advantage) for each  $k \in \{1, \dots, K-1\}$  and  $v', v \in [\underline{v}, \bar{v}]$  with  $v' > v$ ,  $\log a_{k+1}(v') - \log a_k(v') \geq \log a_{k+1}(v) - \log a_k(v)$  and (ii) (Absolute advantage) for each  $k \in \{1, \dots, K-1\}$ ,  $a_{k+1} > a_k$ .

By Assumption 1(i)  $a$  is a weakly log super-modular function of talent and task and higher talents have a weak comparative advantage in more complex tasks. In the subsequent analysis this assumption is often strengthened to strict log super-modularity: for  $k \in \{1, \dots, K-1\}$ , and  $v', v \in [\underline{v}, \bar{v}]$  with  $v' > v$ ,  $\log a_{k+1}(v') - \log a_k(v') > \log a_{k+1}(v) - \log a_k(v)$ . This stronger condition ensures assortative matching of tasks and talents in equilibrium. Assumption 1(ii) implies that more talented types have an absolute advantage in all activities. It is not essential for all of our results, but it guarantees that wages are strictly increasing in talent. Hence, the orderings over talent and wages conform and there is no “wage pooling” (multiple talents earning the same wage).

**Remark 1** (Interpreting  $a$ ). *The function  $a$  captures the idea that different workers may be more or less effective at performing specific tasks or using task-specific capital. Combined with Assumption 1 it formalizes the notions of talent and task complexity. More talented workers are better at all tasks and are especially good at more complex ones. Relatedly, more complex tasks are more talent-intensive. The formulation of production here follows that in the assignment literature, e.g. Costinot and Vogel (2010), with the important addition of an intensive effort margin.<sup>21</sup>*

Later we allow for the possibility that  $a$  may change over time. We interpret such change as technical progress and allow it to depend upon both worker talent and task complexity. In particular if, for each  $v$  and  $k' > k$ ,  $\log \frac{a_{k'}(v)}{a_k(v)}$  increases, then technical progress is talent-biased; if for each  $k$  and  $v' > v$ ,  $\log \frac{a_k(v')}{a_k(v)}$  increases, then it is complexity-biased and if for each  $k' > k$ ,  $v' > v$ ,  $\log \left( \frac{\frac{a_{k'}(v')}{a_k(v')}}{\frac{a_{k'}(v)}{a_k(v)}} \right)$  increases, then it is biased towards high talent-high complexity matches. In the latter case, it enhances the comparative advantage of talent in complex tasks and reduces the substitutability of talent across tasks.  $\square$

The task choices of workers imply a distribution of workers and, hence, effective labor across tasks. Let  $\Lambda_k$  denote a distribution of  $k$ -th talent workers over tasks with density  $\lambda_k$ . If  $k$ -th talent workers exert effort  $e_k$ , then the supply of effective labor in task  $v$  is:

$$\sum_{k=1}^K \lambda_k(v) a_k(v) e_k.$$

<sup>21</sup>The assignment literature refers to a worker’s innate productive attribute as “skill”. Since skills are endogenous, we prefer the word talent. Our model could be reinterpreted as one in which workers exert effort partly or wholly in acquiring skills rather than working.

A representative firm hires effective labor to perform tasks and combines task output to produce final output. Let  $l : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$  denote an allocation of effective labor across tasks and let  $\mathcal{L}$  denote the set of such allocations (with  $\mathcal{L}$  restricted to ensure the integrals defined below in (2.9) are well defined). Output is assumed to equal effective labor in each task  $v$ . Final output  $Y$  is produced from task output and, hence, from an allocation of effective labor  $l$  using a CES-technology:

$$Y = H(l) := \begin{cases} A \left\{ \int_{\underline{v}}^{\bar{v}} b(v) l(v)^{\frac{\varepsilon-1}{\varepsilon}} dv \right\}^{\frac{\varepsilon}{\varepsilon-1}} & \varepsilon \in \mathbb{R}_+ \setminus \{1\}, \\ A \exp \left\{ \int_{\underline{v}}^{\bar{v}} b(v) \ln l(v) dv \right\} & \varepsilon = 1, \end{cases} \quad (2.9)$$

where  $A > 0$  and  $b : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_{++}$  is a continuous function such that if  $B(v) := \int_{\underline{v}}^v b(v') dv'$ , then  $B(\bar{v}) = 1$ . Let  $\omega : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$  be the wage per unit of effective labor in each task  $v$ . The firm solves:

$$\max_{l \in \mathcal{L}} H(l) - \int_{\underline{v}}^{\bar{v}} \omega(v) l(v) dv. \quad (2.10)$$

**Remark 2** (Interpreting  $b$ ). *The function  $b$  weights task output in the final good aggregator. Variations in  $b$  may be interpreted as stemming from technological or preference-based variations in demand for different task outputs. We do not explicitly model capital. However, the model may be extended in this direction, in which case the production functions in (2.9), under the assumption  $B(\bar{v}) \in (0, 1)$ , can be reinterpreted as indirect production functions for labor across tasks after the substitution of optimal capital. The parameter  $b(v)$  is then interpreted as the sensitivity of final output with respect to the labor input in task  $v$ . It is influenced not only by variations in demand for different tasks, but also variations in the capital/labor intensity of tasks. Such variations are stressed by [Acemoglu and Autor \(2011\)](#) who emphasize the automatization of middle complexity tasks. A further possibility is that  $b$  captures the extent to which workers purchase task output in domestic markets, produce it at home or purchase it in foreign markets. Shifts in  $b$  for some tasks may reflect the substitution of market for home production as in [Buera and Kaboski \(2012\)](#) or domestic for foreign production as in [Grossman and Rossi-Hansberg \(2008\)](#).*

## 2.4.2 Tax Equilibria and the Government's Policy Problem

In the assignment setting, the definition of a tax equilibrium is modified as follows.<sup>22</sup>

---

<sup>22</sup>As before, we constrain the set of mechanisms available to the government to ones that deterministically condition upon worker incomes. This assumption is standard in the literature and to a first approximation describes current tax

**TAX EQUILIBRIUM** Let  $G$  be a fixed public spending amount. Given  $G$ , a *tax equilibrium* is an income tax function  $T : \mathbb{R}_+ \rightarrow \mathbb{R}$ , an allocation  $\{l, \{c_k, e_k, \lambda_k\}_{k=1}^K\}$  and a wage profile  $\omega$  such that (i) for each  $k = 1, \dots, K$ ,  $(c_k, e_k)$  and  $v$  in the support of  $\Lambda_k$  solves the  $k$ -th worker's problem at  $T$  and  $\omega$ , (ii)  $l$  solves (2.10) at  $\omega$ , (iii) the final goods market clears:

$$G + \sum_{k=1}^K c_k \pi_k \leq H(l), \quad (2.11)$$

and (iv) labor markets clear, for all  $v \in [\underline{v}, \bar{v}]$ ,

$$l(v) = \sum_{k=1}^K \lambda_k(v) a_k(v) e_k, \quad (2.12)$$

and for all  $k = 1, \dots, K$ ,

$$\pi_k = \int_{\underline{v}}^{\bar{v}} \lambda_k(v) dv. \quad (2.13)$$

Again, let  $\mathcal{E}$  denote the set of tax equilibria. Proposition 3 below characterizes tax equilibria. It contains the simple, but important result that *conditional on effort* assignment in a tax equilibria maximizes output.

**Proposition 3.** *Let Assumption 1 hold. Let  $\{l, \{c_k, e_k, \lambda_k\}_{k=1}^K\}$  and  $\omega$  be, respectively, the allocation and wage profile of a tax equilibrium. Then there is a tuple of threshold tasks  $\{\tilde{v}_k\}_{k=1}^{K-1}$  such that:*

$$\lambda_k(v) = \begin{cases} 0 & v \in [\underline{v}, \tilde{v}_{k-1}) \cup (\tilde{v}_k, \bar{v}] \\ \frac{b(v)^\varepsilon a_k(v)^{\varepsilon-1}}{B_k(\tilde{v}_{k-1}, \tilde{v}_k)^\varepsilon} \pi_k & v \in (\tilde{v}_{k-1}, \tilde{v}_k), \end{cases}$$

where  $B_k(\tilde{v}_{k-1}, \tilde{v}_k) := \left[ \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv \right]^{\frac{1}{\varepsilon}}$ . All workers of talent  $k$  earn a common wage  $w_k = \omega(v) a_k(v)$ ,  $v \in [\tilde{v}_{k-1}, \tilde{v}_k]$ . Relative wages are given by:

$$\frac{w_{k+1}}{w_k} = \frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)} = \frac{\frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{\{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}}}{\frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{\{\pi_k e_k\}^{\frac{1}{\varepsilon}}}}. \quad (2.14)$$

*Conditional on the effort profile  $\{e_k\}$ , the equilibrium allocation of talent to tasks maximizes output.*

codes. In our setting, it implies that the government cannot observe the task a worker does or the amount of task output. The former may reasonably reflect the inherent difficulties in distinguishing between a worker's formal job description and the tasks that the worker actually performs.



*Proof.* See Appendix B.2. □

Efficiency of assignment (in the sense of output maximization) *conditional on effort* implies that output is given by the following indirect production function over efforts:

$$F(\pi_1 e_1, \dots, \pi_K e_K) = \sup \left\{ A \left\{ \sum_{k=1}^K B_k(\tilde{v}_{k-1}, \tilde{v}_k) \{e_k \pi_k\}^{\frac{\varepsilon-1}{\varepsilon}} \right\}^{\frac{\varepsilon}{\varepsilon-1}} \mid \text{s.t. } \underline{v} \leq \tilde{v}_1 \leq \dots \leq \tilde{v}_{K-1} \leq \bar{v} \right\}. \quad (2.15)$$

With  $F$  determined in this way, the environment effectively reduces to that in Section 2.3 and the government's problem to (PP). Recovery of an optimal tax equilibrium can be decomposed into two steps. The outer step is simply (PP) at the induced production function  $F$ ; the embedded inner step solves the assignment problem (2.15) at each candidate effort allocation  $\{e_k\}$  and, hence, evaluates  $F$  at  $\{e_k\}$ .<sup>23</sup>

In contrast to Section 2.3, the production function  $F$  is micro-founded; changes in parameters of this production function can be related to changes in the demand for tasks  $b$  and the productivity of task-talent matches  $\{a_k\}$ . The inner step assignment problem is essentially the same as those considered in Teulings (1995), Costinot and Vogel (2010) and Acemoglu and Autor (2011) (with the distinction that the supply of each talent's labor is selected as part of an optimal tax equilibrium rather than being pinned down parametrically).<sup>24</sup> Solving the assignment problem at an effort profile  $\{e_k\}$  reduces to finding a sequence of task thresholds  $\{\tilde{v}_k\}_{k=1}^{K-1}$  satisfying the discrete boundary value problem:

$$\frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)} = \frac{\frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{\{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}}}{\frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{\{\pi_k e_k\}^{\frac{1}{\varepsilon}}}}, \quad (2.16)$$

with  $\tilde{v}_0 = \underline{v}$  and  $\tilde{v}_K = \bar{v}$ .

An immediate consequence of Proposition 3 and the absolute advantage condition Assumption 1(ii) is that  $\frac{w_{k+1}}{w_k} = \frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)} > 1$ . Consequently, talents are strictly ordered by equilibrium wages and “wage pooling” (the payment of the same wage to different talent types) does not oc-

<sup>23</sup>In a tax equilibrium, a worker reproducing the income and paying the taxes of a less talented type will exert less effort in the task that pays her the best wage, she does not move to the task of the less talented whose income she mimics. Thus, worker task (and wage) choice is independent of the effort she exerts and the income she earns in the task. The counterpart of this in the decomposition just described is the incentive constraint in the outer step which depends on relative wages and only via them on task choice.

<sup>24</sup>In fact the analysis on p. 758-60 of Costinot and Vogel (2010) in which the labor input across “skills” is changed in particular ways represents a partial exploration of the indirect production function.

cur.<sup>25</sup> Proposition 2 identifies relative wage-effort elasticities  $\phi_{k,j}$  as key determinants of the wage compression channel and, hence, marginal taxes. If each  $\log(a_{j+1}/a_j)$  is differentiable, then in a tax equilibrium the terms  $\phi_{k,j}$  can be expressed as:

$$\phi_{k,j} = -\frac{\partial \log(w_{j+1}/w_j)}{\partial \log e_k} = \begin{cases} -\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j} \prod_{l=k}^{j-1} \left( \frac{\partial \log \tilde{v}_{l+1}}{\partial \log \tilde{v}_l} \right) \frac{\partial \log \tilde{v}_k}{\partial \log e_k} & j \geq k \\ -\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j} \prod_{l=j}^{k-2} \left( \frac{\partial \log \tilde{v}_l}{\partial \log \tilde{v}_{l+1}} \right) \frac{\partial \log \tilde{v}_{k-1}}{\partial \log e_k} & j < k \end{cases}. \quad (2.17)$$

Thus, elasticity  $\phi_{k,j}$  depends upon the local comparative advantage of talents  $j$  and  $j+1$   $\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j}$  at the threshold  $\tilde{v}_j$ , the sensitivity of the  $k-1$ -th or  $k$ -th task threshold to the effort of the  $k$ -th talent  $\frac{\partial \log \tilde{v}_k}{\partial \log e_k}$  and the sensitivity of thresholds intermediate between  $j$  and  $k$  to one another  $\frac{\partial \log \tilde{v}_{l+1}}{\partial \log \tilde{v}_l}$ .

Only under very special conditions is the induced production function  $F$  a CES function. One such case occurs when each  $\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j} = 0$ ,  $\phi_{k,j} = 0$ , talents are perfectly substitutable across tasks and  $F$  is linear. Another<sup>26</sup> occurs when the  $a_k$  functions are indicators for the sub-intervals  $[\underline{v}, \tilde{v}_1]$ ,  $(\tilde{v}_1, \tilde{v}_2], \dots, (\tilde{v}_{K-1}, \bar{v}]$ . Then workers are as substitutable as the tasks into which they are locked. For more general cases, however, relative wage-effort elasticities are functions of technological parameters and the effort profile  $\{e_k\}$  and, hence, indirectly policy. Thus, they are not structural. In the appendix, we prove:

**Lemma 3.** *Each  $\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}}$ ,  $\frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j}$  and  $\frac{\partial \log \tilde{v}_k}{\partial \log e_k}$  is positive. Each  $\frac{\partial \log \tilde{v}_{k-1}}{\partial \log e_k}$  is negative. If  $\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j} > 0$ , then  $\phi_{k,j} < 0$  if  $j \geq k$  and  $\phi_{k,j} > 0$  if  $j < k$ . In addition,  $\phi_{k,k} \in [-1/\varepsilon, 0]$  and  $\phi_{k,k-1} = [0, 1/\varepsilon]$ .*

*Proof.* See Appendix B.2. □

The economics behind Lemma 3 is straightforward. Consider a small increase in  $e_k$  (perhaps in response to a policy change). This raises output in tasks  $[\tilde{v}_{k-1}, \tilde{v}_k]$ , placing downward pressure on  $[\tilde{v}_{k-1}, \tilde{v}_k]$ -shadow prices and, hence, the wage  $w_k$  of talent  $k$  workers. These workers respond by populating tasks that are both below  $\tilde{v}_{k-1}$  and above  $\tilde{v}_k$ . This task migration moderates, but does not fully offset the impact of the increase in  $e_k$  on  $w_k$ . As  $k$ -talents move into less complex tasks in which they have a comparative disadvantage relative to  $k-1$ -talents and more complex tasks in

<sup>25</sup>Assumption 1(ii) (i.e. global absolute advantage of more talented types across the entire task space) is sufficient, but not necessary for this result. Local absolute advantage of successive talents  $k+1$  at each task boundary  $\tilde{v}_k$  is enough.

<sup>26</sup>Although, this case is not consistent with talent-complexity comparative advantage (except when  $K = 2$ ), smoothness or continuity of the  $a_k$  functions.

which they have a comparative disadvantage relative to  $k + 1$ -talents so  $w_k/w_{k-1}$  falls and  $w_{k+1}/w_k$  rises. Moreover, as  $k$ -talents spill into neighboring tasks, output of these tasks increases, depressing their shadow prices and inducing neighboring talents to migrate into new tasks. Workers of talent  $k + 1$  move into tasks above  $\tilde{v}_{k+1}$ , while workers of talent  $k - 1$  talents move into tasks below  $\tilde{v}_{k-1}$ . A ripple effect is created with each task threshold  $\tilde{v}_j$  above  $k$  rising and each threshold below  $k$  falling. Since relative wages between adjacent talents are determined by productivity ratios at thresholds (i.e. by  $a_{j+1}(\tilde{v}_j)/a_j(\tilde{v}_j)$ ), an effort change by talent  $k$  workers can affect relative wages across the whole spectrum of talents and be a motive for encouraging or discouraging that talent's effort.

Expressions for the threshold elasticities  $\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}}$ ,  $\frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j}$  and  $\frac{\partial \log \tilde{v}_k}{\partial \log e_k}$  are given in the proof of Lemma 3. They point to the role of the parameters  $b$  and  $a$  in influencing the sensitivity of task choices and, hence, relative wages to a given talent's effort. Suppose that workers of talent  $j$  migrate into more complex tasks either because they have increased their effort or because the tasks that they originally performed have been encroached upon by  $j - 1$  talents. If there is much demand and, hence, high  $b$ -values for tasks immediately above  $\tilde{v}_j$ , then these tasks will soak up this migration with little change in the threshold  $\tilde{v}_j$ . Conversely, if  $b$ -values in this neighborhood are low, then talent  $j$ -workers will migrate further up through the task set pushing  $\tilde{v}_j$  to a new possibly much higher level. In the former case, the impact on the  $w_{j+1}/w_j$  wage differential will be muted; in the latter case, it will be enhanced. Turning to the  $a$  function, an increase in the comparative advantage of talent in complex tasks, raises  $\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j}$  and, hence, the sensitivity of relative wages to task threshold adjustment. The resulting upwards pressure on  $\phi_{k,j}$  is dampened by the deterrence to task migration and task threshold adjustment and, hence, lower values for  $\frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j}$  and  $\frac{\partial \log \tilde{v}_k}{\partial \log e_k}$ , created by higher comparative advantage.

### 2.4.3 An Example: Technical Change in the Two Talent Model

We now make some of the preceding observations more precise in the context of a simple two talent example. For concreteness, we label these talents low ( $k = L$ ) and high ( $k = H$ ) rather than 1 and 2. We restrict preferences to be quasi-linear in consumption,  $U(c, e) = c - \frac{e^{1+\gamma}}{1+\gamma}$ , with  $\gamma > 0$  and denote the government's Pareto weights by  $g_k$ ,  $k \in \{L, H\}$ . To create a motive for redistribution to low skills, we assume  $g_L > \pi_L$ . In this case, the Mirrlees and Wage Compression components can

be consolidated to give the (Stiglitz) optimal tax functions:

$$\frac{\tau_L^*}{1 - \tau_L^*} = \left( \frac{g_L}{\pi_L} - 1 \right) \left\{ 1 - \left( \frac{1}{\mathcal{W}^*} \right)^{1+\gamma} \left\{ 1 - \frac{1}{\mathcal{E}^*} \right\} \right\} \geq 0, \quad (2.18)$$

$$\frac{\tau_H^*}{1 - \tau_H^*} = \left( \frac{g_H}{\pi_H} - 1 \right) \left( \frac{1}{\mathcal{W}^*} \frac{e_L^*}{e_H^*} \right)^{1+\gamma} \frac{1}{\mathcal{E}^*} \leq 0, \quad (2.19)$$

where  $\mathcal{W}^* = \frac{w_H^*}{w_L^*}$  is the optimal talent premium and the substitutability of talents at the optimum is completely described by the elasticity of substitution  $\mathcal{E}^*$ . In the assignment setting, both  $\mathcal{W}^*$  and  $\mathcal{E}^*$  are endogenous. If Assumption 1 is maintained, then in an optimal tax equilibrium the set of tasks is partitioned at a threshold  $\tilde{v}^*$ , with low talents working in tasks below  $\tilde{v}^*$  and high talents working in tasks above. The talent premium satisfies  $\mathcal{W}^* = \frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)}$ , while the elasticity of substitution between the labor of the two talents is given by  $\mathcal{E}^* = \mathcal{E}(\tilde{v}^*, a, b)$ , where

$$\mathcal{E}(\tilde{v}; a, b) := -\frac{\partial \log \frac{e_H}{e_L}}{\partial \log \frac{w_H}{w_L}} = \varepsilon + \frac{1}{\frac{\partial \log \frac{a_H}{a_L}}{\partial v}(\tilde{v})} \left[ \frac{b_H(\tilde{v})}{B_H(\tilde{v})} + \frac{b_L(\tilde{v})}{B_L(\tilde{v})} \right] \geq \varepsilon, \quad (2.20)$$

$B_L(\tilde{v}) := \int_{\underline{v}}^{\tilde{v}} b(v)^\varepsilon a_L(v)^{\varepsilon-1} dv$ ,  $B_H(\tilde{v}) := \int_{\tilde{v}}^{\bar{v}} b(v)^\varepsilon a_H(v)^{\varepsilon-1} dv$  and for  $k \in \{L, H\}$ ,  $b_k(\tilde{v}) := b(\tilde{v})^\varepsilon a_k(\tilde{v})^{\varepsilon-1}$ .

Equation (2.20) makes explicit the role of task migration in raising the elasticity of substitution between talents above that of task outputs:  $\frac{\partial \log \frac{a_H}{a_L}}{\partial v}(\tilde{v})$  is the local comparative advantage of high talents in the neighborhood of the threshold task  $\tilde{v}$ . If this term equals  $\infty$ , then workers are substitutable as the interval of tasks into which they are locked. Otherwise, their ability to migrate across tasks enhances their substitutability. Equation (2.20) highlights the dependence of the elasticity of substitution on technological parameters and its (implicit) dependence on policy.

The workers' equilibrium first order conditions in this setting together with (2.14) gives:

$$\mathcal{W}^* = \frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)} = \left( \frac{B_H(\tilde{v}^*)}{B_L(\tilde{v}^*)} \right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} \left( \frac{\pi_L}{\pi_H} \right)^{\frac{\gamma}{1+\gamma\varepsilon}} \left( \frac{1 - \tau_H}{1 - \tau_L} \right)^{\frac{1}{1+\gamma\varepsilon}}. \quad (2.21)$$

Equation (2.21) gives the threshold  $\tilde{v}^*$  and relative wages as functions of  $a$  and  $b$  and relative marginal taxes. Substituting for optimal marginal taxes from (2.18) and (2.19) reduces the system

of equilibrium equations to a single equation in one unknown,  $\tilde{v}^*$ :

$$\frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)} = \left( \frac{B_H(\tilde{v}^*)}{B_L(\tilde{v}^*)} \right)^{\frac{\gamma\epsilon}{1+\gamma\epsilon}} \left( \frac{\pi_L}{\pi_H} \right)^{\frac{1+\gamma}{1+\gamma\epsilon}} \times \left\{ \frac{g_H - (\pi_H - g_H) \left( \frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)} \right)^{(1+\gamma)(\epsilon-1)} \left( \frac{B_H(\tilde{v}^*)}{B_L(\tilde{v}^*)} \right)^{-(1+\gamma)\epsilon} \frac{1}{\mathcal{E}^*}}{g_L - (g_L - \pi_L) \left( \frac{a_L(\tilde{v}^*)}{a_H(\tilde{v}^*)} \right)^{1+\gamma} \left( \frac{\mathcal{E}^*-1}{\mathcal{E}^*} \right)} \right\}^{\frac{1}{1+\gamma\epsilon}}. \quad (2.22)$$

It follows easily from (2.22) that if  $\epsilon \geq 1$  (so that goods, and, hence, efforts of different talents are gross substitutes) and if  $\mathcal{E}(\cdot; a, \cdot)$  is (locally) constant, then complexity-biased perturbations of  $b$  that raise  $B_H/B_L$  lead to increases in  $\tilde{v}^*$  and  $\mathcal{W}^*$ . Intuitively, increases in the relative demand for more complex tasks raise the relative shadow price of such tasks and encourage less talented workers to migrate into them ( $\tilde{v}^*$  rises). However, such task-upgrading erodes the comparative advantage of low talents and the talent premium ( $\mathcal{W}^*$ ) rises. These effects are mitigated by adjustments in relative efforts that occur in response to wage adjustments and that are reinforced by changes to tax policy. Overall, a rise in  $B_H/B_L(\cdot)$  is associated with a higher talent premium, a tightening of the incentive constraint between low and high talents and higher marginal taxes on low talents.

Sufficient conditions for the elasticity of substitution  $\mathcal{E}^*$  to be constant in response to a shift in task demand are rather stringent.<sup>27</sup> In general, it may rise or fall as a direct effect of the change in  $b$  or the indirect effect of changes in  $\tilde{v}^*$  on  $\frac{1}{\frac{\partial \log \frac{a_H}{a_L}}{\partial \tilde{v}}(\tilde{v}^*)} \left\{ \frac{b_H(\tilde{v}^*)}{B_H(\tilde{v}^*)} + \frac{b_L(\tilde{v}^*)}{B_L(\tilde{v}^*)} \right\}$  in (2.20). These changes may reinforce or offset the responses just described. To the extent that  $\mathcal{E}^*$  is increased, the government's ability to compress wage differentials and relax incentive constraints is reduced. It is correspondingly encouraged to reduce relative taxation of low talents and to permit a further increase in the talent premium. The reverse is true if  $\mathcal{E}^*$  falls.

Turning next to the consequences of variation in  $a$ , suppose that  $\log \frac{a_H(v)}{a_L(v)} = \alpha_1 + \alpha_2(v - \underline{v})$  so that  $\alpha_1$  controls the absolute advantage of high talents (in the lowest task) and  $\alpha_2$  controls their comparative advantage in more complex tasks. If  $\epsilon > 1$  and  $\frac{b_H(\tilde{v}^*)}{B_H(\tilde{v}^*)} + \frac{b_L(\tilde{v}^*)}{B_L(\tilde{v}^*)}$  is locally constant, then small technologically induced increases in  $\alpha_2$  will, from (2.22), both raise the talent premium and reduce the elasticity of substitution  $\mathcal{E}^*$ .<sup>28</sup> Low talent marginal taxes  $\tau_L^*$  will rise both because  $\mathcal{W}^*$

<sup>27</sup>For example, if  $\epsilon = 1$ ,  $b$  is constant and equal to one and the remaining parameters are such that  $\tilde{v}^* = 1/2$ , then  $\mathcal{E}^*$  is locally constant.

<sup>28</sup>In this case the task threshold  $\tilde{v}^*$  falls: the increased productivity of high talents in complex tasks reduces the

risers and because the wage compression channel is enhanced via the reduction in  $\mathcal{E}^*$ : as workers become less substitutable, the government is encouraged to offset the rise in the talent premium by discouraging low talent effort through taxation. Increases in  $\alpha_1$  work in a related way, but absent any reinforcing adjustment in  $\mathcal{E}^*$ . As in the case of complexity-biased perturbations in the  $b$  functions, adjustments in the  $\frac{b_H(\tilde{v}^*)}{B_H(\tilde{v}^*)} + \frac{b_L(\tilde{v}^*)}{B_L(\tilde{v}^*)}$  term (either direct through changes to the  $b_k$  functions or indirect through adjustments to  $\tilde{v}^*$ ) may work to reinforce or dampen these effects.

**SUMMARY** Technical change that increases the talent wage premium and reduces the substitutability of talents is associated with higher optimal marginal taxes on low talents. Change that increases both the talent income premium and the substitutability of talents is associated with lower marginal subsidies on high talents. In general, the technical parameters  $a$  and  $b$  influence both talent premia and talent substitutability directly and indirectly through endogenous task assignment. The analysis is more complicated in settings with multiple talents. Such settings are, however, essential for exploring the policy implications of recently documented polarizing shifts in the pattern of wages and employment across occupations.

#### 2.4.4 Comparison to Rothschild and Scheuer, 2013

We briefly describe the connections between our model and that of Rothschild and Scheuer (2013). Our model features a continuum of tasks and a finite set of talents, but it is readily reformulated as one with a continuum of tasks and talents (see Appendix B.3). In both formulations our assumptions ensure that the ordering over talents translates directly into an ordering over wages. Consequently, the pattern of (local) binding incentive constraints over talents is easily inferred and consumption and effort allocations can be solved directly as functions of talent. We use our approach to relate optimal taxes to the indirect production function  $F$ , relative wage elasticities and, hence, properties of the task-talent productivity function  $a$ .

In contrast, Rothschild and Scheuer (2013) consider an environment with a finite number of tasks in which an agent's type is her productivity in each task and is, thus, multidimensional. In this case, the structure of binding incentive-compatibility conditions across allocations expressed as functions of type is complicated. However, such conditions become quite standard if con-

---

relative shadow price of such tasks and encourages high talents to downgrade their tasks. Despite some erosion of their comparative advantage, their relative wages rise.

sumption and effort allocations are reformulated as functions of wages. The cost of this reformulation is a rather complicated joint restriction on allocations and the (endogenous) distribution over wages. To solve such a problem [Rothschild and Scheuer \(2013\)](#) propose a quite different inner-outer method than that used here. In the inner step the allocation of labor across tasks and, hence, the wage distribution is fixed and an optimal incentive-compatible consumption-effort profile (over wages) consistent with this allocation is found. In the outer step, the labor allocation and the wage distribution are determined. [Rothschild and Scheuer \(2013\)](#) use this approach to relate optimal taxes to the impact of effort on the wage distribution. Their formula, thus, provides an alternative perspective on the forces shaping tax policy in an endogenous wage environment to ours.

[Rothschild and Scheuer \(2013\)](#)'s model permits intra-task wage dispersion and task-specific wage distributions with overlapping support. It thus allows the implications of these things for policy to be explored, ours does not. On the other hand, our model connects directly to the technical change literature in labor economics. It underpins an empirical strategy for quantifying the effect of technical change on optimal policy described in Sections [2.5](#) and [2.6](#). Thus, the models are complimentary. In Appendix [B.7](#), we present a general formulation that nests our model and that of [Rothschild and Scheuer \(2013\)](#) and makes transparent the alternative approaches taken. It then specializes that formulation to one intermediate between our model and theirs. This formulation incorporates intra-task wage dispersion and wage overlap, while preserving our approach to formalizing the impact of technical change.

## 2.5 Measuring Technical Change

In this section, we measure the extent of technical change in the US. Our data source is the Current Population Survey (CPS).<sup>29</sup> We proceed as if this data was generated by a (possibly sub-optimal) tax equilibrium and use parametric assumptions and equilibrium restrictions from our model to identify and estimate the technological parameters  $a$  and  $b$  in the 1970's and 2000's. In Section [2.6](#), we calculate optimal tax equilibria at these estimated parameters.

---

<sup>29</sup>Further details of our use and treatment of the data are given in Appendix [B.4.1](#).

### 2.5.1 Determining Types and Tasks

**MAPPING EMPIRICAL OCCUPATIONS TO ORDERED SETS OF TASKS** The CPS categorizes workers into distinct occupations; our sample contains  $M = 302$  occupations. The CPS also provides information on worker earnings and hours worked from which a measure of wages can be imputed. Our model involves an interval of tasks ordered by complexity. We identify tasks with empirical occupations and use the average wage paid in each occupation to infer its complexity. In so doing, we utilize the model's implication that task wages are rising in task complexity. We normalize the task space to  $[v, \bar{v}] = [0, 1]$  and sub-divide this interval into  $M$  sub-intervals of length  $\Delta v = \frac{1}{M}$ ,  $\mathcal{V}_m = [v_{m-1}, v_m]$ . We calculate the imputed average wage in each occupation using 1970's data and rank occupations according to this wage. The  $m$ -th ranked occupation is then mapped to the  $m$ -th subinterval  $\mathcal{V}_m$ .<sup>30</sup>

We use data on the skill content of occupations contained in the O\*NET database to (partially) corroborate our inferred complexity ordering over occupations. The O\*NET database provides a detailed description of the skill (35 distinct skills are considered) and ability (52 distinct abilities are considered) content of each occupation.<sup>31</sup> We recover from the O\*NET a single index describing the importance of each skill and ability for each occupation.<sup>32</sup> We then calculate the correlations of these skill/ability indices with our wage imputed rank. We find that the three most correlated skills (correlation in parenthesis) are: *complex problem solving* (0.66); *critical thinking* (0.62) and *judgement and decision making* (0.61). The three most correlated abilities are: *deductive reasoning* (0.63); *inductive reasoning* (0.60) and *written comprehension* (0.57). The least correlated skill is *equipment maintenance* (-0.07), while the least correlated abilities are: *stamina* (-0.33) and *trunk strength* (-0.37). These correlations suggest that the average wage paid in an occupation is informative about that occupation's complexity.

---

<sup>30</sup>We keep this ranking over occupations fixed. In doing so, we follow the precedent of [Acemoglu and Autor \(2011\)](#). Fixing the ranking allows us to unambiguously identify an index  $v$  with a physical occupation and to interpret variations in the parameters  $a$  and  $b$  as occurring in a given physical occupation rather than at a given complexity index whose physical interpretation is shifting. However, there is some reranking of occupations over time in the data. In [Appendix B.4.3](#) we describe the implications of using current rather than the 1970's wage ranking for our estimates of the  $a$  and  $b$  functions and for optimal taxes.

<sup>31</sup>The O\*NET database contains 974 occupations. We relate these to the occupations contained in CPS in two steps. We first map the occupations in our sample to the Standard Occupation Classification of the 2000 census. We then map these occupations to those in the 19<sup>th</sup> release of the O\*NET. A small number of occupations are recoded manually. We thank Giovanni Gallipoli for directing us towards the O\*NET.

<sup>32</sup>Specifically, the index is the product of the importance and level measures in O\*NET.



RECOVERING THE EMPIRICAL ASSIGNMENT FUNCTION  $\tilde{v}$  The model in Section 2.4 featured a finite number of talents; this facilitated the derivation of analytical results. However, for the remainder of the paper we find it convenient to treat worker talent symmetrically with task complexity and to assume that workers are distributed uniformly across an interval of talents,  $k \in [\underline{k}, \bar{k}]$ .<sup>33</sup> Thus, a worker's talent should now be interpreted as an index (and a rank), the implications of which for productivity are captured by the function:  $a : [\underline{k}, \bar{k}] \times [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$ . Although the distribution over the (ordinal) talent index is uniform, the distribution over (cardinal) productivities is not: it is induced endogenously by  $a$  and by the assignment of talent to tasks. The set  $[\underline{k}, \bar{k}]$  is normalized to  $[0, 1]$ .

The continuous analogue of the task thresholds  $\{\tilde{v}_k\}$  is a task assignment function:  $\tilde{v} : [\underline{k}, \bar{k}] \rightarrow [\underline{v}, \bar{v}]$ . This function is strictly increasing in our model. Denote its inverse by  $\tilde{k}$ . Under the assumption that workers are distributed uniformly across talent indices,  $\tilde{k}$  is the distribution of workers across tasks. Consequently, we treat the distribution of workers across ordered occupations as the empirical counterpart of  $\tilde{k}$  and  $\tilde{v}$  to be the inverse of this.

### 2.5.2 Estimating $b$

It is well known that the elasticity of substitution between goods and factor augmenting technical progress ( $\varepsilon$  and  $b$  in our case) cannot be separately identified from data on outputs, inputs and marginal products - an observation that goes back to [McFadden, Diamond, and Rodriguez \(1978\)](#). In our baseline case, we restrict the elasticity of substitution between task outputs,  $\varepsilon$ , to be one (so that the final good production function is Cobb-Douglas) and identify  $b(v)$  with the share of total compensation paid to workers in task  $v$ .<sup>34</sup> Thus, estimates of  $b$  may be calculated from compensation shares independently of knowledge of the  $a$ 's. Specifically, under the Cobb-Douglas restriction, the firm's first order conditions from the continuous-talent version of (2.10), imply for almost all  $(k, v)$ :

$$\omega(k, v) = Y \frac{a(k, v)b(v)}{y(v)}. \quad (2.23)$$

---

<sup>33</sup>The convenience is two fold. First, since occupational (task) data is discrete, assuming a continuous set of talents avoids having to deal with talent groups that are distributed across adjacent occupations. Second, it allows us to apply numerical optimal control methods to solve the problem. A formal statement of the continuous talent-continuous task model can be found in Appendix B.3.

<sup>34</sup>The quantitative implications of alternative assumptions for  $\varepsilon$  are considered in Appendix B.4.2.

In the continuous talent setting, task output is given by  $y(v) = a(\tilde{k}(v), v)e(\tilde{k}(v))\tilde{k}_v(v)$ , with  $\tilde{k}_v$  the derivative of  $\tilde{k}$ . Combining this with (2.23) and integrating over  $\mathcal{V}_m$  gives total labor income in occupation  $m$  in terms of the  $b$ -function:

$$\int_{\mathcal{V}_m} \omega(\tilde{k}(v), v)e(\tilde{k}(v))\tilde{k}_v(v)dv = Y \int_{\mathcal{V}_m} b(v)dv.$$

Average income in occupation  $m$ ,  $i_m$ , is then obtained by dividing both sides by the mass of workers in the occupation,  $S_m$ :

$$i_m := \frac{1}{S_m} \int_{\mathcal{V}_m} \omega(\tilde{k}(v), v)e(\tilde{k}(v), v)\tilde{k}_v(v)dv = \frac{Y}{S_m} \int_{\mathcal{V}_m} b(v)dv.$$

Thus, the average value of  $b$  in occupation  $m$ ,  $b_m \Delta v := \int_{v_{m-1}}^{v_m} b(v)dv$ , is:

$$b_m = \frac{S_m i_m}{\Delta v Y}, \quad \forall m = 1, \dots, M. \quad (2.24)$$

We identify  $Y$  with per capita labor income.<sup>35</sup> A smooth estimate of the  $b$ -function is obtained by fitting a LOWESS model to  $\{v_m, \log b_m\}$  data.<sup>36</sup> Figure 2.3 displays estimates of  $b$  for the 1970's and the 2000's. The figure shows that  $b$  rises (slightly) for low and (significantly) for high  $v$ -occupations,

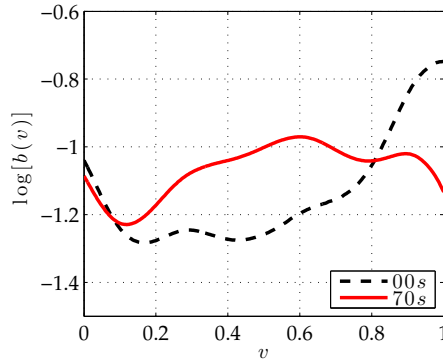


Figure 2.3: Evolution of  $\log(b(v))$  across decades.

but falls for intermediate ones. The picture is consistent with the phenomenon of *job polarization* as

<sup>35</sup>In 2005 dollars we have  $Y_{70} = \$36,998$  and  $Y_{00} = \$45,260$ .  $M$  is 302. In aggregate data using GDP deflator (table 1.1.9 in NIPA) and total non farm payroll (BLS) we get a value of real compensation per worker equal to  $Y_{70} = \$37,114$  and  $Y_{00} = \$53,304$ . However deflating using CPI we get values consistent with our sample:  $Y_{70} = \$37,966$  and  $Y_{00} = \$45,151$ .

<sup>36</sup>The LOWESS scatterplot smoothing builds up a smooth curve through a set of data points by fitting simple linear or quadratic models to localized subsets of data. We use a smoothing parameter of 0.4.

discussed in Section 2.2. This polarization feature is robust to different sample selection assumptions, see Appendix B.4.1 for details.

Figure 2.4 sharpens intuition concerning the relation of different  $v$ 's to the data. The figure overlays the values of  $b(\cdot)$  with a bar graph displaying the employment shares of occupations belonging to particular sectors. Figure 2.4a does this for services and Figure 2.4b for manufacturing.<sup>37</sup> The service sector is associated mostly with extreme and, especially, “low”  $v$  occupations (the bar on the right in Figure 2.4a refers to managers and administrative support), while manufacturing is mostly middle  $v$  occupations (although with a wider range).

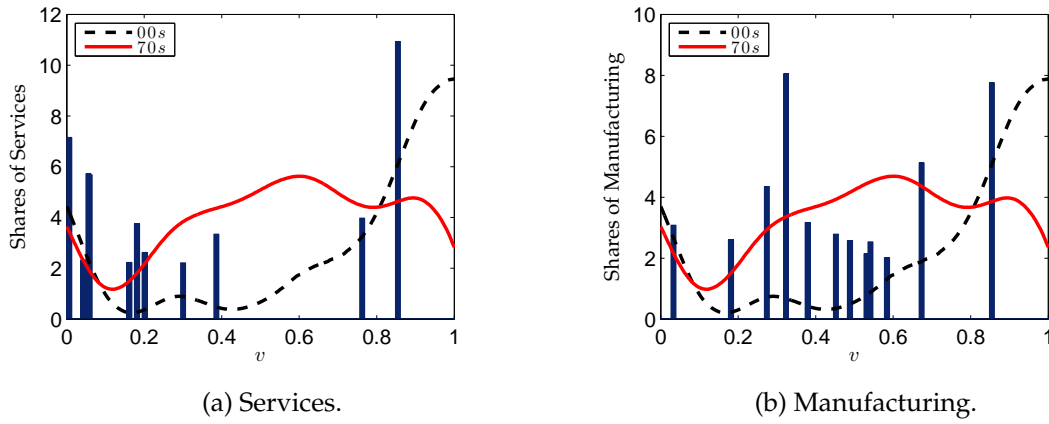


Figure 2.4: Occupations and  $v$

Histograms: shares of occupations over  $v$ . Plots: smoothed values for  $\log(b(v))$  over  $v$  and across decades.

### 2.5.3 Estimating $a$

The envelope condition from the task choice component of the worker's equilibrium problem,

$w(k) = \max_{v \in [\underline{v}, \bar{v}]} \omega(k, v)$ , implies that:

$$\frac{d \log w}{dk}(k) = \frac{\partial \log \omega(k, \tilde{v}(k))}{\partial k} = \frac{\partial \log a(k, \tilde{v}(k))}{\partial k} = \frac{\partial \alpha}{\partial k}(k, \tilde{v}(k)), \quad (2.25)$$

where  $\alpha(k, v) := \log a(k, v)$ . An empirical counterpart for  $\frac{d \log w}{dk}$  is constructed in three steps. First, information from the CPS on weeks and usual hours worked in the previous year and self reported yearly labor income is used to impute workers' average hourly wages. Second, wages are averaged over occupation to construct empirical counterparts of  $w(\tilde{k}(v))$ . Third, a LOWESS

<sup>37</sup>Not shown are occupations that constitute less than 2% of the workforce of each sector.

smoother is applied to the log of this series and to  $\tilde{k}$ , derivatives of each function are calculated and  $\frac{d \log w(\tilde{k}(v))}{dk} = \frac{\frac{d \log w(\tilde{k}(v))}{dv}}{\frac{\partial \tilde{k}(v)}{\partial v}}$  is found. Figure 2.5 displays the empirical values for (smoothed)  $\log w(k)$  for the 1970s and the 2000s. From the 20<sup>th</sup> to the 80<sup>th</sup> talent percentile, this function is roughly linear in  $k$  in the 1970s and remains so in the 2000s. In the 1970s, it steepens over the top talent decile, while in the 2000s, it steepens over the top *two* deciles. In addition, for both decades, but especially for the 1970's, the profile is steeper over the bottom two deciles.

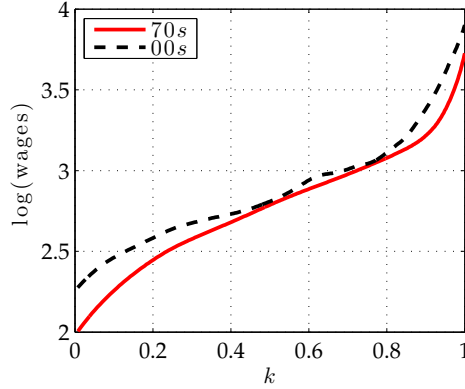


Figure 2.5: LOWESS smoothed  $\log(\text{wages})$  over talents.

The evolution of  $\log w(k)$  shown in Figure 2.5 suggests that between the 1970s and the 2000's the wages of low-ranked talents caught up with mid-ranked talents, while the wages of mid-ranked talents fell behind those at the top. These developments are qualitatively consistent with a fall in the returns to talent in simpler tasks combined with an increase in talent-complexity comparative advantage (so that talent premia rise in the most complicated tasks and occupations). This motivates us to select:

$$\frac{\partial \alpha}{\partial k}(k, v) = \alpha_1 + \alpha_2 \cdot v. \quad (2.26)$$

Here,  $\alpha_1$  captures the return to pure talent, while  $\alpha_2$  captures comparative advantage.<sup>38</sup> We recover estimates of  $\alpha_1$  and  $\alpha_2$  by regressing  $\frac{d \log w(\tilde{k}(v))}{dk}$  onto a constant and the task index  $v$ . The regression is weighted by the share of workers in each  $v$ . Results are reported in Table 2.1.

They show a significant increase in the comparative advantage parameter  $\alpha_2$  between the 1970's and 2000's. Loosely, this is driven by the increase in wage growth over high talents occurring

<sup>38</sup>Appendix B.4.5 considers a case in which comparative advantage is increasing with task complexity.

between the 1970s and the 2000s.<sup>39</sup>

Table 2.1: Estimation of Productivity Function.

Decade	Parameters	
	$\alpha_1$	$\alpha_2$
70s	1.07 (0.25)	1.71 (0.28)
00s	0.42 (0.32)	3.01 (0.22)

In Appendix B.4.4, we look outside of the CPS for corroborating evidence of increasing comparative advantage. Specifically, we use data on the change in the skill/ability content of occupations contained in different editions of the O\*NET database. We find evidence that the use and importance of skills and abilities associated with complex tasks has increased in high wage occupations relative to low.

Finally, the parameter  $A$  is given by the ratio of per capita income to the approximation of the CES aggregator  $\exp\{\int_{\underline{v}}^{\bar{v}} b(v) \log\{y(v)\} dv\}$ .

## 2.6 Quantitative Implications for Policy

In this section, we compute optimal policy responses to the technical change estimates derived in Section 2.5. Calculation of policy requires a specification of worker and societal preferences and the amount of resources devoted to public spending. We briefly turn to this and then give our quantitative results.

### 2.6.1 Selection of Other Parameters and Computational Method

We assume that worker preferences are given by:  $U(c, e) = \log c - \frac{e^{1+\gamma}}{1+\gamma}$ . Note that the choice of  $U$  has no impact on the estimation of  $b(v)$  and  $a(k, v)$ . We follow Chetty, Guren, Manoli, and Weber

<sup>39</sup>Kaplan and Rauh (2013) emphasize the rise of “superstar” pay across a variety of high income occupations. In our empirical strategy “superstar” workers belong to (measured) occupations inhabited by much lower paid workers. It is arguable that these different workers trade in distinct task-markets with distinct shadow prices. The implication of this is a downward bias in the estimate of comparative advantage ( $\alpha_2$ ). Given the evolution of inequality in the US this bias is likely to be more significant for the 2000s.

(2011) and set the Frisch labor supply elasticity to  $1/\gamma = 0.75$ . We identify the share of output allocated to public spending with the aggregate tax to income ratio in our CPS sample. On this basis,  $(G/Y)_{70} = 16.2\%$  and  $(G/Y)_{00} = 14.0\%$ ; we set the  $G/Y$  ratio to the intermediate value of 15%.<sup>40</sup> Finally, in our benchmark calculations a utilitarian government is assumed:  $g_k = \pi_k$  for all talents  $k$ .

To calculate optimal policy at our selected and estimated parameters, we first formulate the government's optimization as an optimal control problem. Details of this formulation are given in Appendix B.3. We then solve the problem numerically using the GPOPS-II software.<sup>41</sup>

## 2.6.2 Optimal Tax Results

Table 2.2 reports optimal average and marginal tax rates as a function of income percentiles for the 1970s and the 2000s. Over this time period, average rates rise at low incomes and fall at high and, especially, middle incomes. Transfers to the lowest deciles are reduced. Overall, the reform favors those in the middle. Marginal rates fall at low to mid incomes and rise at higher incomes. In the extreme tails they move in the opposite directions: rising in the very lowest and falling in the very highest (where marginal subsidies are increased ) percentiles.

Table 2.2: Optimal Tax Rates on Real Labor Income.

		Percentiles of Income					
	Decade	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	99 <sup>th</sup>
<b>Averages</b>	70s	-11.9	-7.3	6.9	22.3	26.1	22.2
	00s	-2.3	-1.1	5.6	19.9	26.1	21.9
<b>Marginals</b>	70s	20.3	34.1	44.3	40.3	23.9	-0.6
	00s	15.3	25.4	39.7	42.2	27.4	-2.2

To understand the evolution of optimal tax reported in Table 2.2, we return to the tax formula (2.4)

<sup>40</sup>NIPA data (Table 1.1.6) gives  $(G/Y)_{70} = 23.9\%$  and  $(G/Y)_{00} = 19.3\%$ . However, since we are concerned with spending financed out of income taxation (paid by our subsample of labor income earners) we use the alternative CPS-generated estimates.

<sup>41</sup>GPOPS-II is a flexible software for solving optimal control problems, see Patterson and Rao (2013).

derived earlier.

## ACCOUNTING FOR OPTIMAL TAXES

Tax formula (2.4) allows us to decompose optimal tax rates into “Mirrleesian” and “Wage Compression” components. In particular, let  $\tau_k^M$  denote the “Mirrleesian” marginal tax rate in the absence of the wage compression term:<sup>42</sup>

$$\tau_k^M = \frac{\frac{\Delta w_{k+1}^*}{w_{k+1}^*} \frac{1-\Pi_k}{\pi_k} \mathcal{H}_k^* \Psi_k^*}{1 + \frac{\Delta w_{k+1}^*}{w_{k+1}^*} \frac{1-\Pi_k}{\pi_k} \mathcal{H}_k^* \Psi_k^*}.$$

The tax rate  $\tau_k^M$  is that which an optimizing government would apply if wages were fixed at their optimal levels  $\{w_k^*\}$ . Define the wage compression component of taxes to be the residual  $\tau_k^{WC} = \tau_k^* - \tau_k^M$ . In Figure 2.6, we plot the Mirrleesian tax rate  $\tau_k^M$  and the overall optimal marginal rate  $\tau_k^*$  at each income percentile  $k$  and for each decade. Figure 2.6 shows that technical change deforms

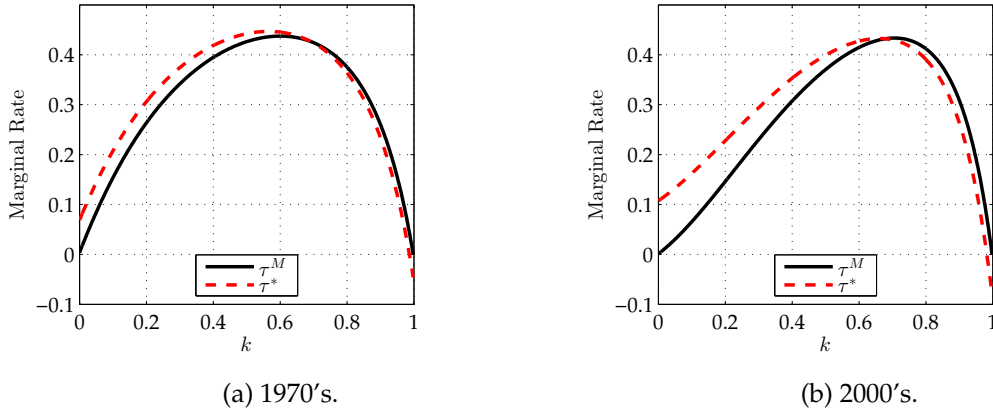


Figure 2.6: Decomposing taxes. Solid curve: Mirrleesian tax,  $\tau^M$ . Dashed curve: overall tax,  $\tau^*$ .

the Mirrleesian tax rate pushing it to the right except at the lowest and highest talent. In addition, it raises the wage compression component at lower incomes and reduces it at higher ones. Overall the wage compression component becomes quantitatively more important.

<sup>42</sup>That is set the wage compression term to zero in (2.4) and rearrange. For convenience, we continue to state tax formulas and their components in their discrete, rather than continuous forms.

EVOLUTION OF THE MIRRLEES TERM We further decompose the Mirrlees term into its redistributive  $\Psi$  and wage growth parts in Figure 2.7.<sup>43</sup> The main impact of technical change is upon wage

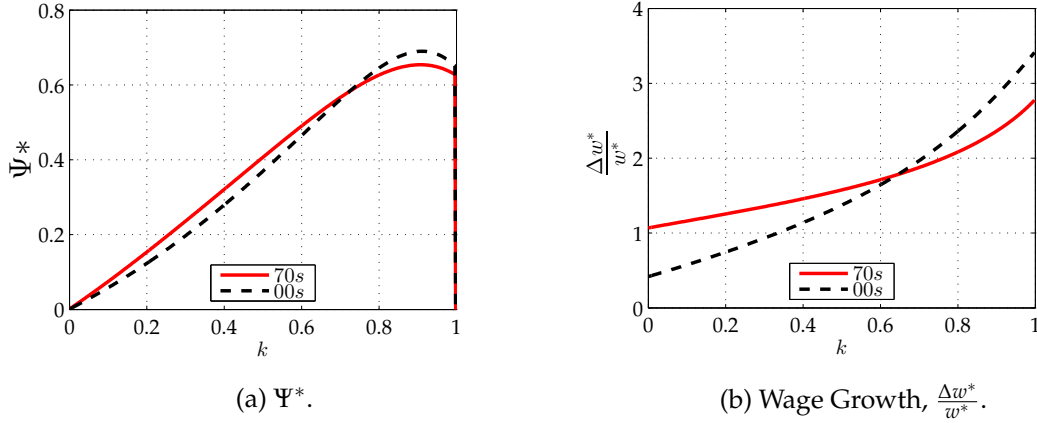


Figure 2.7: Mirrlees term decomposition.

growth (with some slight reinforcement from the redistributive term  $\Psi^*$ ). This is largely driven by shifts to the  $a$  function. As noted previously, our estimates suggest that the productivities of low talents catch up with high in less complicated tasks and fall behind in more complex ones. At any effort profile and, in particular, at the optimal one, this shift compresses wage differentials at the bottom and expands them at the top. Shifts in the  $b$  function and in task demand from the middle to the extremes slightly reinforce the effect. The impact of the latter is, however, surprisingly small. This is largely because, in relevant areas of the task space, modest adjustments in the tasks of workers  $\tilde{v}^*$  are consistent with quite large variations in the density of workers across tasks  $\tilde{k}_v^*$ . Consequently, increases in the demand for low and high tasks are met with increases in the number of workers performing these tasks, but relatively little adjustment in task assignment and, hence, relative productivities and wages. For more details see Appendix B.5. The overall effect of these  $a$  and  $b$  changes is to relax incentive constraints and reduce marginal taxes at the bottom, but to tighten them and raise marginal taxes at the top.

EVOLUTION OF THE WAGE COMPRESSION TERM Adjustment of the wage compression terms is in the opposite direction to the adjustment of the Mirrlees term previously described. Figure 2.8 displays this adjustment.

<sup>43</sup>The other components are constant over time under our assumptions.



It shows that the wage compression term *rises* at low incomes and *falls*, becoming more negative, at higher ones. These changes are largely attributable to adjustments in the relative wage

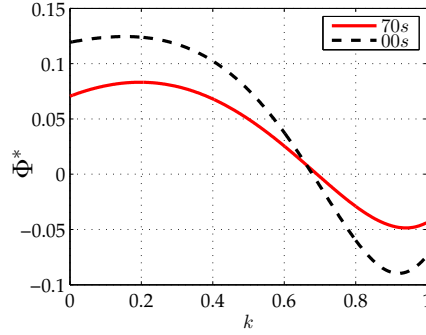


Figure 2.8: Evolution of the wage compression term.

elasticities  $\phi_{k,j}^*$ . The  $k$ -th talent's wage compression term is given by:  $\Phi_k^* = \sum_{j=1}^{K-1} \mathcal{M}_{k,j} \phi_{k,j}^*$ . This equation expresses  $\Phi_k^*$  as a weighted sum of relative wage elasticities, with the weights depending upon the marginal incentive benefit of adjusting each pair of relative wages. Mechanically,  $\phi_{k,j}^*$  is positive if  $j \geq k$  and negative otherwise, so that all  $\phi_{k,j}^*$  are positive if  $k = 1$  and all are negative if  $k = K$ . For some intermediate  $k$ , positive and negative terms cancel and the wage compression term is zero. An increase in the lowest talent's effort pushes all higher talents upwards through the task spectrum, raising the relative wages of all adjacent talents. This tightens all incentive constraints and is undesirable. Consequently, the lowest talent has the highest wage compression term and that talent's effort should be deterred at the margin. For the highest talent, this argument is reversed. An increase in the highest talent's effort pushes all lower talents downwards through the task set, compressing relative wages. This relaxes incentive constraints and should be encouraged at the margin with lower marginal taxes on high talent incomes. For intermediate talents these effects wholly or partially offset, leading to wage compression terms that are smaller in absolute value. Figure 2.9 shows the impact of technical change on relative wage elasticities (normalized by population shares)  $\phi_{k,j}^* / \pi_k$ ,  $j = 1, \dots, K$  for low, mid and high talents (labelled  $L$ ,  $M$  and  $H$ ).<sup>44</sup> It indicates that almost all  $\phi_{k,j}^*$  rise in absolute value. This is largely a consequence of the rise in the comparative adjustment parameter  $\alpha_2$  which, although it dampens the assignment response to adjustments in effort, raises the sensitivity of relative wages to any reassignment that occurs.

<sup>44</sup>Note the global impact of relative wages to an effort adjustment, an example of the ripple effect described previously.

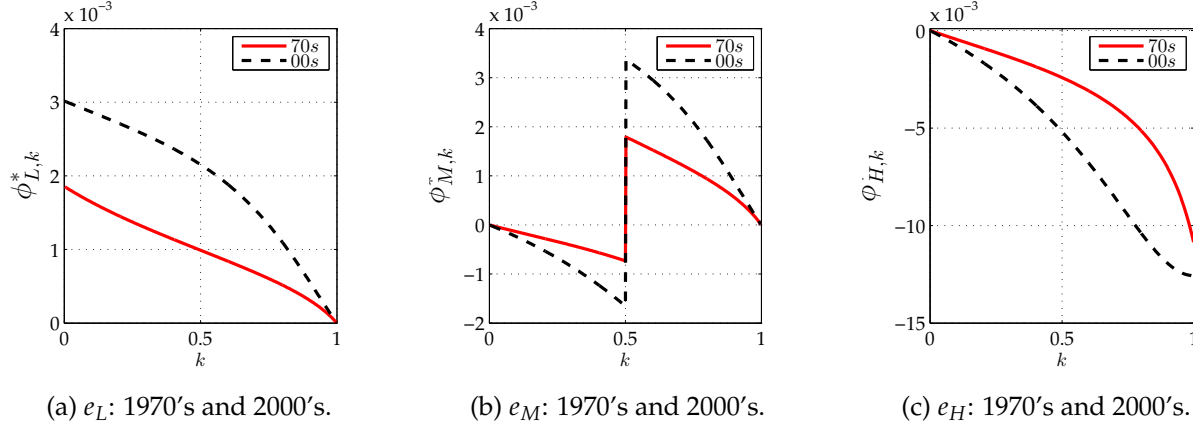


Figure 2.9: Effort elasticity of tasks.

Changes in the  $b$  function have only moderate effects on these elasticities, see Appendix B.7.

**COMBINING TERMS** The Mirrlees and wage compression terms evolve in opposite directions. Of the two, it is the adjustment to the Mirrlees term that is largest over most incomes. Consequently, marginal tax rates fall at low (but not the lowest) and rise at high (but not the highest) incomes. These adjustments are significantly muted by changes to the wage compression term and at the extremes of the wage distribution changes in this term predominate.

## 2.7 Extension: INTRA-TASK WAGE DISPERSION

This paper's appendices contain various robustness checks, extensions and optimal tax calculations under alternative parameterizations. In the remainder of this section we focus on a specific extension that can accommodate intra-occupational wage dispersion. Recall that in the equilibrium of our (benchmark) model, differently talented workers partition the task space and all workers within a task receive the same wage: there is no intra-task wage dispersion. Our empirical strategy identifies tasks with occupations and uses dispersion in occupational average wages to determine the  $a$  function. It makes no use of measured intra-occupational wage dispersion.

Simple regressions suggest that between one third and one half of wage dispersion can be attributed to occupation. [Mouw and Kalleberg \(2010\)](#) impute wages using income and hours data in the CPS and regress this on three digit occupation dummies. They obtain an  $R^2$  of 39% in the 1980s rising to 43% in 2010. [Lane, Salmon, and Spletzer \(2007\)](#) using OES microdata from 1996-1997 find

that one digit occupational dummies account for 28% of wage variation rising to 54% when three digit occupational data is used. Overall, although occupations account for an important part of wage variation, significant residual wage variation remains. However, the identification of this residual variation with intra-task wage dispersion must be qualified in two ways. First, the residual absorbs measurement error in incomes and hours (from which wages are imputed).<sup>45</sup> Second, it absorbs occupational misclassifications and, more generally, unmeasured variation in task complexity. Several occupational categories within the CPS have fairly expansive definitions (e.g. some managerial occupations include managers of small, simple organizations, as well as managers of large complex ones) and it is likely that different workers sharing such occupational classifications perform different complexity-ranked activities.<sup>46</sup> It is notable that when Lane et al. (2007) introduce establishment dummies on top of occupational ones and interact these dummies the  $R^2$  in their regressions rises to 88%. While establishment dummies may capture many things, it is plausible that they help further refine the task performed by a worker (especially when interacted with occupation). To address this issue requires further unbundling of measured occupations.<sup>47</sup>

Notwithstanding the preceding concerns, intra-task wage variation is present and does contribute to measured intra-occupational wage dispersion. We consider the extent to which it qualifies our results in Appendix B.7. We do so by extending the model and numerically parameterizing it to enhance intra-task wage dispersion. Our goal is to provide a lower bound for the responsiveness of policy to technical change. In the extended model, there are two aspects of talent: one captures comparative advantage in complex tasks, the other the ability to do all things well. Similar to our baseline model in the main text, comparative advantage types partition the ordered space of tasks amongst themselves. Wage variation within these partitions (and, hence, within tasks) is created by dispersion in the second (absolute advantage) component of talent. Such dispersion weakens the link between wages and tasks. It diffuses the impact of technical change and of taxes targeted at a particular income across the wage distribution. Thus, it dampens the responsiveness

---

<sup>45</sup>Bound and Krueger (1991) find that measurement error accounts for 27.6% of total variance of CPS earnings, while Bound, Brown, Duncan, and Rodgers (1994) find that it is more severe for hours and wages.

<sup>46</sup>Relatedly, “superstar” workers belong to (measured) occupations inhabited by much lower paid workers. It is arguable that these different workers trade in distinct task-markets with distinct shadow prices.

<sup>47</sup>We use the O\*NET to provide some very preliminary results in this direction in Appendix B.8. There we report summaries of survey results that indicate disagreement as to the knowledge requirements of occupations (amongst workers employed in or firms employing workers in these occupations). These disagreements are greatest in occupations paying higher wages.

of policy to technical change. In taking the model to the data, we assume a coarse set of comparative advantage types and attribute all measured residual wage variation (about 75% of the total in our CPS sample) to variations in absolute advantage. Since the set of comparative advantage types is coarse, the partitions of the occupation space are large. Hence, we (deliberately) attribute some measured inter-occupational wage variation to absolute advantage (and as discussed above measured inter-occupational wage variation may understate inter-task wage variation and the contribution to overall wage dispersion created by the interaction of talents and tasks). As expected, the impact of technical change on marginal taxes is smaller than in our baseline case: the largest adjustment is about 2.5 points as compared to about 8.5 points before. Again, this adjustment is the net effect of countervailing changes to the Mirrlees and wage compression terms. We interpret this number as a lower bound on the responsiveness of policy to technical change. Moreover, while the quantitative response is more muted than in the benchmark case, the broad policy prescription of modest marginal tax reductions over a band of low to mid level incomes combined with an increase over higher incomes is robust.

## 2.8 Conclusion

We relate the positive literature on technical change to normative work on optimal taxation by embedding an assignment model into an optimal tax framework. The assignment component induces an indirect production function over worker efforts enabling us to map technical parameters determining the productivity of task-talent matches and the demand for tasks to the variables and elasticities relevant for optimal tax analysis. We investigate the implications of changes in these parameters for optimal taxes, measure the extent of this change in US data and evaluate its implications for optimal policy.

The impacts of technical change on wage growth across talents and the substitutability of talents across tasks emerge as key drivers of policy. The twisting of the task-talent productivity function with low talents catching up in simple tasks and falling behind in more complex ones compresses wage differentials at the bottom, while expanding them at the top. It is a force for less redistribution and lower marginal taxes from the middle to the bottom and more redistribution and higher marginal taxes from the top to the middle. On the other hand, increased complemen-

tarity between talent and task complexity reduces the substitutability of talents. In particular, the highest talents become increasingly locked into the highest tasks. Migration to lower ranked tasks to avoid lower task shadow prices entails greater erosion of productivity. This gives the government more tax leverage over the wage distribution. It is a force for higher marginal tax rates at the bottom. A key message of this paper is that policy depends upon the balance of these forces. Models that treat wages (or even the elasticity of substitution between talents) as exogenous omit the latter. We find its impact to be moderate, but non-negligible.

Our paper takes a first step in integrating a task-based model of technical change into a normative public economics framework. We conclude by describing four extensions that we leave for future research. First, our model focuses on the intensive margin of labor supply.<sup>48</sup> It abstracts from indivisibilities in labor supply. If working at a given task requires a minimal (task specific) effort, then some workers may choose inactivity under the optimal tax code. As [Saez \(2002\)](#) shows such modeling of the worker extensive margin can significantly affect optimal tax results at the bottom. However, its implications for the impact of technical change on tax design are less clear. Second, our model assumes that the matching of talents to tasks is frictionless. Thus, our quantitative work is best viewed as capturing the long run policy response to technical change after the (possibly slow) reassignment of workers to tasks following such change. The role of income taxation in supplementing other sources of insurance during transitions is omitted. Third, our model omits accumulation of experience or skill within tasks that can impede or promote transitions to other tasks. Fourth, we abstract from the endogenous nature of technical change. Relaxing these restrictions remain important topics for further research.

---

<sup>48</sup>However, our model admits an alternative interpretation in which workers exert effort in skill accumulation rather than market work. Our theoretical insights are applicable to this interpretation.



## Chapter 3

# Discriminatory Auctions with Resale

### 3.1 Introduction

The discriminatory or “pay-your-bid” auction is a popular mechanism to sell many important goods, including treasury bills or bonds, electricity, foreign exchange, airport landing slots, and more recently carbon emissions. In this type of auction, each bidder submits a demand curve for multiple units of a good and the auctioneer acts as a perfectly discriminating monopolist by charging each bidder his or her winning bid.

In many of these applications, the bidders can freely engage in post-auction resale. That is, they can resell some or all of the items they receive to other bidders after the auction is over. Of course, when bidders anticipate a resale market, their bidding behavior in the auction might change. Nevertheless, the vast majority of the single-item or multi-item auction theory literature has neglected the effect of post-auction resale on the allocation of given auction formats.

In this paper, we consider multi-unit auctions: items sold in the discriminatory auction are identical to each other. Moreover, bidders have single-unit demands, i.e. they value only one item and their valuation for any more items is 0, and bidders’ valuation for the first item is independently and identically distributed. In other words, we consider a symmetric single-unit demand environment. We model the resale stage as a game in which the sellers–winners of the auction stage, which we call *resellers* –sell their (excess) objects optimally. If there is only one reseller, she would use the optimal auction. If there is more than one reseller, then they may compete or coop-

erate with each other in the resale stage, and our results are not affected by either kind of behavior.

In a model with symmetric single-unit demand, it is well known that discriminatory auctions have a symmetric and monotone equilibrium that results in an efficient allocation (see [Krishna \(2002\)](#), section 13.5.2). Therefore, one might think that adding a resale stage to this setup should not alter the equilibrium behavior: in this equilibrium all winners have higher valuations than all losers, so there would be no incentive for resale. We show that this intuition is wrong: it turns out that when sellers in the resale market have all the bargaining power and can design any mechanism to sell the items, then the symmetric, monotone and efficient “no resale equilibrium” is no longer an equilibrium when resale is allowed (Proposition 4). The reason for this is that auction prices may be too low to attract “speculative behavior,” i.e. buying and selling in the resale market.

In our main model, we consider an auctioneer who sells the items via discriminatory price auctions with no reserve prices. In the resale stage we assume that resellers can use optimal mechanisms. Hence they can use reserve prices. Our main results are the following. When there are two units for sale, we find an equilibrium in which resellers make zero profit (Theorem 1). When more than two units are for sale, surprisingly, it turns out that there may not be a symmetric and monotone equilibrium (Theorem 2). The main reason for this is that when two (or more) items could be sold in the resale stage, selling one or two items results in different expected revenues, which results in contradicting requirements for the bid for the first unit. To the best of our knowledge, Theorem 2 is the first result that shows non-existence of a symmetric and monotone equilibrium in a standard auction setup (with independent private values, risk-neutrality, and single-unit demands.)

We then consider some variations of the model. When the resale market has to be efficient, and hence resellers cannot use reserve prices in the resale stage (like the original seller who does not use a reserve price), “no resale equilibrium” remains an equilibrium with resale (Proposition 5). Yet there exists another “resale equilibrium” in which one bidder buys all the items and sells all but one of them in the resale market (Proposition 6). Moreover, resale equilibrium is revenue equivalent to no resale equilibrium (Proposition 7). Furthermore, as a corollary to this result, we note that when there are two units for sale and resellers can use reserve prices in the resale market,



banning the resale market strictly decreases the expected revenue in a discriminatory price auction (Corollary 1). Finally, when reserve prices can be used both in the auction stage and resale stage, no resale equilibrium remains an equilibrium (Proposition 8).

The balance of this section discusses the related literature. Section 3.2 formally introduces the model and ends with a motivating example (Example 1). In Section 3.3, we establish our main results. Section 3.4 considers the variations. Section 3.5 concludes. The Appendix contains omitted proofs.

**Related Literature** Research on the theory of multi-unit or multi-item auctions is not as large and advanced as the research on single-item auctions. [Noussair \(1995\)](#) and [Engelbrecht-Wiggans and Kahn \(1998a\)](#) work on the derivation of equilibrium bidding behavior in private value uniform-price auctions. [Engelbrecht-Wiggans and Kahn \(1998b\)](#) analyzes equilibrium strategies in discriminatory auctions. In another important paper, [Reny \(1999\)](#) shows the existence of pure strategy equilibrium for discriminatory auctions with independent private values. [Ausubel, Cramton, Pycia, Rostek, and Weretka \(2014\)](#) study the issue of demand reduction in multi-unit auctions.

There is now quite a large literature on single-unit auctions with resale. Most of this literature studied environments where resale takes place due to inefficient allocation (such as in asymmetric first price auctions) in the bidding stage. [Gupta and Lebrun \(1999\)](#), [Haile \(2000\)](#), [Haile \(2001\)](#), [Haile \(2003\)](#), [Garratt and Troger \(2006\)](#), [Pagnozzi \(2007\)](#), [Hafalir and Krishna \(2008\)](#) are earlier notable examples of this literature.<sup>1</sup> However, in this paper we look at an environment in which the equilibrium without resale is efficient, yet resale may take place. This phenomenon also occurs in the online supplement to [Garratt and Troger \(2006\)](#). When there is one speculator (who values the item at 0) and  $n$  symmetric bidders in a first-price auction, they show that—under some conditions—the speculator may play an active role (buy in the auction stage and resell in the resale stage)<sup>2,3</sup>. In our setup, it turns out that no resale equilibrium *always* gives rise to speculative behavior.

---

<sup>1</sup>This literature has grown more in the recent years. See, for instance, [Hafalir and Krishna \(2009\)](#), [Lebrun \(2010\)](#), [Pagnozzi \(2010\)](#), [Cheng and Tan \(2010\)](#), [Cheng \(2011\)](#), [Xu, Levin, and Ye \(2013\)](#), [Virag \(2013\)](#), and [Zheng \(2014\)](#).

<sup>2</sup>These conditions depend on number of regular bidders and value distribution. For instance, when value distribution is uniform, speculators do not play an active role.

<sup>3</sup>[Garratt and Troger \(2006\)](#) also show that speculators play an active role in second-price or English auctions. In English auctions, since there are many equilibria (some of which are inefficient), resale may affect equilibrium behavior more easily. See also [Garratt, Troger, and Zheng \(2009\)](#).

Finally, we discuss theoretical literature on multi-unit auctions with resale. In an earlier work, [Bukhchandani and Huang \(1989\)](#) analyze a multi-item (discriminatory or uniform price) auction with *common values*. In the resale market, bidders receive information about the bids submitted in the auction. They examine the information linkage between auction and resale stage and compare expected revenues in two auction formats. Recently, [Filiz-Ozbay, Lopez-Vargas, and Ozbay \(2015\)](#) have studied multi-unit auctions with resale where bidders have either single or multi-unit demand. More specifically, they consider environments in which there are  $k$  local markets,  $k$  local bidders, and 1 global bidder. They analyze the equilibrium of Vickrey auctions and simultaneous second-price auctions. In another recent work, [Pagnozzi and Saral \(2013\)](#) analyze different bargaining mechanisms at the resale stage following a uniform price auction when bidders are ex-ante asymmetric. In contrast, in our model, all bidders are ex-ante symmetric with private values, demand only one item, and participate in the same discriminatory price auction.

### 3.2 Model

There are  $n$  bidders and  $k$  items for sale. We assume that each bidder has single-unit demand, i.e., the values of all items except the first are zero. Bidders are risk neutral, and bidder  $i$ 's value for the first item is  $v_i$ , which is independently and identically distributed from a continuously differentiable and regular (in Myerson's sense) function  $F$  over  $[0, 1]$ .

The rule of the discriminatory auction is simple: the highest  $k$  bids are awarded the objects, and winners have to pay their bids to the auctioneer (pay-your-bid auction). Ties are broken randomly. No information is revealed after the auction stage. Moreover, there is no discounting between auction and resale stage. After the auction stage, bidders are allowed to sell item(s) in the resale stage. As discussed in the introduction, we assume that *resellers* sell their (excess) objects optimally. If there is only one reseller, she would use the optimal auction (in Section 4, we consider the case in which resellers cannot use reserve prices; hence they use optimal "efficient" auctions). If there is more than one reseller, then they may compete or cooperate with each other while selling items to the buyers at the resale stage.

We study the weak perfect Bayesian Nash equilibrium (WPBNE) of this game. That is, we

assume players are sequentially rational and they update their beliefs according to Bayes' rule and equilibrium behavior whenever possible. We restrict our attention to the symmetric and monotone WPBNE. Hence, we consider an equilibrium such that each bidder with value  $v$  bids  $(\beta_1(v), \beta_2(v), \dots, \beta_k(v))$  such that  $\beta_1(v) \geq \beta_2(v) \geq \dots \geq \beta_k(v)$  where  $\beta_l$  denotes the  $l^{th}$  highest bid of a bidder with value  $x$ . We assume that  $\beta_l$  is a nondecreasing and continuously differentiable function for each  $l = 1, \dots, k$ . It is important to note that we only require  $\beta_l$  to be nondecreasing (not strictly increasing) and hence allow  $\beta_l(v)$  to be zero. That is, bidders do not have to make more than one positive bid.

The behavior in the resale stage is straightforward. In equilibrium, any bidder who wins  $j$  items in the bidding stage sells only  $j - 1$  items in the resale stage. This is because his value for the first item is greater than the expected return of selling that item, since bidders follow the symmetric nondecreasing equilibrium. When a bidder is the only reseller, she would use the optimal auction—a uniform-price auction with the optimal reserve price—and buyers in the resale stage would bid their valuations. We do not need to specify what happens if there is more than one reseller, since in the equilibria we will find, there will always be one reseller.<sup>4</sup>

### 3.2.1 Motivating Example

In this subsection, we give an example that illustrates allowing resale changes equilibrium behavior of bidders.

**Example 1.** Consider an auctioneer who sells two identical objects to three bidders whose values (are single-unit demand and) are uniformly distributed over  $[0, 1]$ . From Section 13.5.2 of Krishna (2002), bidder with value  $x$  bidding  $\left(\frac{3x-2x^2}{6-3x}, 0\right)$  is an equilibrium when bidders cannot engage in post-auction resale.

When bidders can engage in resale, and reserve prices can be used in resale stage, bidding  $\left(\frac{3x-2x^2}{6-3x}, 0\right)$  is not an equilibrium. To see this, consider a bidder with value 1, and compare her utility when she bids  $\left(\frac{1}{3}, \frac{1}{3}\right)$  with when she bids  $\left(\frac{1}{3}, 0\right)$  (her equilibrium bid). In the latter case, her utility is  $\frac{2}{3}$ . In the former case, she wins both items with probability 1. She would use one of the items for consumption. When she

---

<sup>4</sup>The main contribution of this paper is to solve for an equilibrium when there are two items and show that there may not be a symmetric and monotone equilibrium when there are three items. Under both cases, in a symmetric and monotone equilibrium there can be at most one reseller in the resale market.

sells the second item by a second-price auction with the optimal reserve price  $\frac{1}{2}$ , we can quickly check that the revenue is  $\frac{5}{12}$ , which is strictly higher than  $\frac{1}{3}$ . The expected utility from bidding  $(\frac{1}{3}, \frac{1}{3})$  and setting the reserve price to  $\frac{1}{2}$  in the resale stage results in an expected utility of  $\frac{2}{3} + \frac{5}{12} - \frac{1}{3} = 0.75$ , which is higher than 0.66. Therefore bidding  $(\frac{1}{3}, \frac{1}{3})$  is a profitable deviation.

So, a natural question is what the equilibrium in this game is? We claim that a bidder with a value  $x$  bidding  $(\frac{5}{12}x, \frac{5}{12}x)$  is an equilibrium of this game. To see this, first note that if a bidder wins the second item at a bid of  $\frac{5}{12}z$ , this means that both of the other bidders must have values between 0 and  $z$ . By running an optimal auction for selling the second item, the expected revenue can be shown to be exactly  $\frac{5}{12}z$ .<sup>5</sup> Therefore, the expected utility from winning the second item is 0 no matter what they bid for it, and hence bidders cannot benefit from deviating in the bid for the second item. Then we just need to check that deviating for the first item is not profitable. Consider a bidder whose value is  $x$  and her bid is  $\frac{5}{12}z$  (for the first item, her bid for the second item will always bring 0 expected utility.) Her expected utility is given by

$$\Pi(x, z) = (x - \frac{5}{12}z)z^2 + 2 \int_z^{\min\{1, 2x\}} \left( \int_0^{\frac{y_1}{2}} \left(x - \frac{y_1}{2}\right) dy_2 + \int_{\frac{y_1}{2}}^{\min\{y_1, x\}} (x - y_2) dy_2 \right) dy_1$$

where the second summand represents the expected utility from buying in the resale stage. To see this, note that for this bidder to win the item in the resale stage, the highest value among two competitors should be greater than  $z$  (so that she would win the auction) but not more than  $2x$  (so that the reserve price he would charge is not more than  $x$ ). Moreover, the price she would pay in the resale market is the maximum of the reserve price  $\frac{y_1}{2}$  or the second-highest value in the resale market  $y_2$ .

Then the profit of deviation is:

$$\begin{aligned} \Pi(x, z) - \Pi(x, x) &= \\ &= \begin{cases} 0 & \text{if } z \leq x \\ -\frac{1}{3}(z - x)^3 & \text{if } z \geq x \end{cases} \end{aligned}$$

---

<sup>5</sup>The optimal auction is to run a second-price auction with a reserve price  $\frac{z}{2}$ . Then, with probability  $\frac{1}{2}$ , the object will be sold at the reserve price, and when the object is sold higher than reserve price (with probability  $\frac{1}{4}$ ) the expected selling price is  $\frac{2}{3}z$  (expectation of second highest of two random variables uniformly distributed between  $\frac{z}{2}$  and  $z$ ). Hence the optimal revenue is  $\frac{1}{2} \times \frac{z}{2} + \frac{1}{4} \times \frac{2z}{3} = \frac{5}{12}z$ .

which is  $\leq 0$  for each case. Hence,  $(\frac{5}{12}x, \frac{5}{12}x)$  is an equilibrium.

We introduce some notations before we move on to our main results. Let the random variable  $Y_k^{(n)}$  represent the  $k^{th}$  highest random value among  $n$  random variable independently distributed according to  $F$ , and let  $F_k^{(n)}$  denote the distribution function for  $Y_k^{(n)}$ . Let  $\psi(x)$  denote Myerson's virtual valuation:

$$\psi(x) = x - \frac{1 - F(x)}{f(x)}.$$

We assume  $\psi(x)$  to be increasing. Let  $F(\cdot | x)$  denote the conditional distribution: for  $y \in [0, x]$

$$F(y | x) = \frac{F(y)}{F(x)}$$

and let  $f(\cdot | x)$  denote density, and  $\psi_x(\cdot)$  denote virtual valuation of  $F(\cdot | x)$ . We also assume that  $\psi_x(y)$  is increasing in  $y$  for all  $x$ .<sup>6</sup> Finally, let us denote

$$t(x) = \begin{cases} y & \text{if } x < \psi^{-1}(0), \text{ where } \psi_y^{-1}(0) = x \text{ or } 0 = \psi_y(x) \\ 1 & \text{if } x \geq \psi^{-1}(0) \end{cases}$$

The relevance of this definition can be seen by noting that any bidder with value greater than  $t(x)$  would charge the optimal reserve price greater than  $x$  in the resale market (so a bidder with value  $x$  will not be able to buy from that bidder).

### 3.3 Main Results

Let us first formally define "no resale equilibrium." From Section 13.5.2 of Krishna (2002), we know that

$$\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < x \right] \equiv \beta_1^N(x), \beta_2(x) = \dots = \beta_k(x) = 0$$

is an equilibrium of discriminatory auction without resale when bidders have single-unit demand and their valuations are i.i.d. We call  $\beta^N(x) = (\beta_1^N(x), 0, \dots, 0)$  the *no resale equilibrium strategy*.  $\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < x \right]$  is obviously increasing. Hence, in this equilibrium,  $k$  bidders with highest

---

<sup>6</sup>Myerson's regularity assumption is satisfied by many distributions and it is commonly made in auction theory and mechanism design literature.

valuations will be awarded the items and all losers would have valuations smaller than all winners. Therefore, if this was the equilibrium of the auction stage, there would be no transactions in the resale stage. Yet, as Example 1 illustrates, no resale equilibrium bid functions do not constitute an equilibrium when  $n = 3$ ,  $k = 2$  and  $F$  is uniform. We first generalize this for arbitrary  $n$ ,  $k$ , and  $F$ .

**Proposition 4.** *The bidding strategy of the game with no resale is not part of an equilibrium of discriminatory auctions with resale when resellers can use optimal auction in the resale market.*

The proof follows from an argument similar to that in Example 1 and is relegated to the Appendix.

Next, we focus on the case  $k = 2$  (and any  $n \geq 3$ ) and find an equilibrium of this game.

### 3.3.1 When there are 2 units for sale

Before establishing equilibrium characterization for  $k = 2$ , we introduce some convenient notation and prove some lemmas that are necessary for the main result. Let  $\gamma(x)$  denote the expected revenue of the optimal auction when there are  $n - 1$  buyers who all have values smaller than  $x$ . We know that the optimal auction allocates the object to the bidder with the highest virtual value if this highest virtual value is greater than zero. Since our setup is symmetric and virtual values are assumed to be increasing, the optimal auction will allocate the item to highest valued bidder if his virtual value is greater than 0. Moreover, the contribution to the revenue is exactly equal to the virtual value. Hence,

$$\gamma(x) = \int_{\psi_x^{-1}(0)}^x \psi_x(z) dF_1^{(n-1)}(z | x)$$

We first establish the following lemma regarding  $\gamma$  and  $\gamma'$ , which will be useful for our results.

**Lemma 4.** *We have*

$$\gamma(x) = \frac{1}{F(x)^{n-1}} \int_{\psi_x^{-1}(0)}^x \left( z - \frac{F(x) - F(z)}{f(z)} \right) dF(z)^{n-1} \quad (3.1)$$

and

$$\gamma'(x) = (n-1) \left[ \frac{f(x)}{F(x)} (x - \gamma(x)) - \frac{f(x)}{F(x)^{n-1}} \int_{\psi_x^{-1}(0)}^x F(y)^{n-1} dy \right] \quad (3.2)$$

Now, we establish one of our main results.

**Theorem 1.** *Bidding  $\beta_1(x) = \beta_2(x) = \gamma(x)$  is an equilibrium.*

*Proof.* Let us consider a bidder with value  $x$ . First of all, as in the above example, the expected utility from winning the second item is 0 no matter what she bids for the second item. This is because, when she wins the item with a bid of  $b$ , then she knows that all  $n - 1$  bidders have values smaller than  $\gamma^{-1}(b)$ . By running an optimal auction to sell the second item her expected revenue is  $\gamma(\gamma^{-1}(b)) = b$ . Hence this bidder cannot benefit from deviating in the bid for the second item. Then we just need to check that deviating for the first item is not profitable.

Let us denote the interim expected utility of a bidder with value  $x$  and a bid  $\gamma(z)$  for the first item (and an arbitrary bid for the second item that is not greater than  $\gamma(z)$ ) by  $\Pi(x, z)$ . We have

$$\begin{aligned} \Pi(x, z) &= F(z)^{n-1} (x - \gamma(z)) \\ &+ \int_z^{t(x)} \int_0^{\min\{x, y_1\}} \int_0^{y_2} \dots \int_0^{y_{n-2}} \left( x - \max\{\psi_{y_1}^{-1}(0), y_2\} \right) f(y_1, \dots, y_{n-1}) dy_{n-1} dy_{n-2} \dots dy_1 \end{aligned}$$

The first summand equals to the expected utility of winning two items. The bidder wins two items with probability  $F(z)^{n-1}$ , and her utility of winning two items is  $x - \gamma(z)$ . The second summation represents her expected utility from resale when she is a buyer: the highest value among competitors,  $y_1$ , has to be between  $z$  and  $t(x)$  so that she will be a buyer in the resale stage and the reserve price in the resale market is not greater than her value; the second highest value among competitors,  $y_2$ , should be smaller than her value so that she would win the item in the resale market; finally, she would pay  $\max\{\psi_{y_1}^{-1}(0), y_2\}$  when she receives the item in the resale stage (since winner of the auction would be using the optimal auction: a second price auction with a reserve price). Note that we have

$$f(y_1, \dots, y_{n-1}) = (n-1) \prod_{l=1}^{n-1} f(y_l)$$

Hence, we can rewrite  $\Pi(x, z)$  as

$$\begin{aligned}
\Pi(x, z) &= F(z)^{n-1} (x - \gamma(z)) + \int_z^{t(x)} \left( \int_0^{\min\{x, y_1\}} (x - \max\{\psi_{y_1}^{-1}(0), y_2\}) dF(y_2)^{n-1} \right) f(y_1) dy_1 \\
&= F(z)^{n-1} (x - \gamma(z)) + \underbrace{\int_z^{t(x)} \left( \int_0^{\psi_{y_1}^{-1}(0)} (x - \psi_{y_1}^{-1}(0)) dF(y_2)^{n-1} \right) f(y_1) dy_1}_{:=\Omega_1} \\
&\quad + \underbrace{\int_z^{t(x)} \left( \int_{\psi_{y_1}^{-1}(0)}^{\min\{x, y_1\}} (x - y_2) dF(y_2)^{n-1} \right) f(y_1) dy_1}_{:=\Omega_2}.
\end{aligned}$$

Note that

$$\Omega_1 = \int_z^{t(x)} (x - \psi_{y_1}^{-1}(0)) F(\psi_{y_1}^{-1}(0))^{n-1} f(y_1) dy_1$$

and

$$\begin{aligned}
\Omega_2 &= \int_z^{t(x)} x \left( F(\min\{x, y_1\})^{n-1} - F(\psi_{y_1}^{-1}(0))^{n-1} \right) f(y_1) dy_1 \\
&\quad - \int_z^{t(x)} \left( u F(u)^{n-1} - \psi_{y_1}^{-1}(0) F(\psi_{y_1}^{-1}(0))^{n-1} - \int_{\psi_{y_1}^{-1}(0)}^u F(y_2)^{n-1} dy_2 \right) f(y_1) dy_1
\end{aligned}$$

where  $w := \min\{x, y_1\}$ .

Hence, we can simplify  $\Pi(x, z)$  as

$$\Pi(x, z) = F(z)^{n-1} (x - \gamma(z)) + \int_z^{t(x)} \left( F(w)^{n-1} (x - w) + \int_{\psi_{y_1}^{-1}(0)}^w F(y_2)^{n-1} dy_2 \right) f(y_1) dy_1.$$

Now, let us consider the difference  $\Pi(x, z) - \Pi(x, x) = D(x, z)$ . We will show that  $D(x, z) \leq 0$  to complete the proof. We have

$$\begin{aligned}
D(x, z) &= F(z)^{n-1} (x - \gamma(z)) - F(x)^{n-1} (x - \gamma(x)) \\
&\quad + \int_z^{t(x)} \left( F(w)^{n-1} (x - w) + \int_{\psi_{y_1}^{-1}(0)}^w F(y_2)^{n-1} dy_2 \right) f(y_1) dy_1.
\end{aligned}$$



By definition, we have  $D(x, x) = 0$ . Hence the following is sufficient to finish proof:

$$\frac{\partial}{\partial z} D(x, z) \begin{cases} \leq 0 & \text{if } z > x \\ \geq 0 & \text{if } z < x \end{cases}$$

First, consider the case where  $z > x$ . We have

$$\begin{aligned} \frac{\partial}{\partial z} D(x, z) &= (n-1) F(z)^{n-2} f(z) (x - \gamma(z)) - F(z)^{n-1} \gamma'(z) - \left( \int_{\psi_z^{-1}(0)}^x F(y_2)^{n-1} dy_2 \right) f(z) \\ &= f(z) \left( (n-1) F(z)^{n-2} (x - \gamma(z)) - \int_{\psi_z^{-1}(0)}^x F(y_2)^{n-1} dy_2 \right) - F(z)^{n-1} \gamma'(z) \\ &< f(z) \left( (n-1) F(z)^{n-2} (z - \gamma(z)) - \int_{\psi_z^{-1}(0)}^z F(y_2)^{n-1} dy_2 \right) - F(z)^{n-1} \gamma'(z) \end{aligned}$$

where the third line is obtained by noting that  $(n-1) F(z)^{n-2} (x - \gamma(z)) - \int_{\psi_z^{-1}(0)}^x F(y_2)^{n-1} dy_2$  is increasing in  $x$  for all  $x \leq z$ .

Now, let us substitute Equation 3.2 in Lemma 4 for  $\gamma'$ . We have

$$\begin{aligned} \frac{\partial}{\partial z} D(x, z) &< f(z) \left( (n-1) F(z)^{n-2} (z - \gamma(z)) - \int_{\psi_z^{-1}(0)}^z F(y_2)^{n-1} dy_2 \right) \\ &\quad - F(z)^{n-1} \left( (n-1) \left[ \frac{f(z)}{F(z)} (z - \gamma(z)) - \frac{f(z)}{F(z)^{n-1}} \int_{\psi_z^{-1}(0)}^z F(y)^{n-1} dy \right] \right) \\ &= 0 \end{aligned}$$

Next, we consider the latter case  $z < x$ . The marginal benefit is:

$$\begin{aligned} \frac{\partial}{\partial z} D(x, z) &= (n-1) F(z)^{n-2} f(z) (x - \gamma(z)) - F(z)^{n-1} \gamma'(z) \\ &\quad - \left( F(z)^{n-1} (x - z) + \int_{\psi_z^{-1}(0)}^z F(y_2)^{n-1} dy_2 \right) f(z) \end{aligned}$$

Use Equation 3.2 in Lemma 4 for  $\gamma'$  and some algebra:

$$\begin{aligned} \frac{\partial}{\partial z} D(x, z) &= f(z) \left( (n-1) F(z)^{n-2} (x - z) - F(z)^{n-1} (x - z) \right) \\ &= f(z) \left( (x - z) F(z)^{n-2} (n - 1 - F(z)) \right) \geq 0 \end{aligned}$$

Thus, bidding  $\beta_1(x) = \beta_2(x) = \gamma(x)$  is an equilibrium.  $\square$

### 3.3.2 When there are 3 or more units for sale

In this subsection, we show that, interestingly, when there are 3 or more units for sale, there may not be a symmetric and monotone equilibrium. We show this by considering a specific example with 3 units and 4 bidders and showing that there is no symmetric and monotone equilibrium for that case.

**Theorem 2.** *When there are 3 or more units for sale, there may not be any symmetric and monotone equilibrium.*

*Proof.* Consider 3 items for sale and 4 bidders who have single-unit demands that are distributed according to a *uniform* distribution on unit interval. We consider a symmetric and monotone equilibrium. That is, each bidder with value  $x$  submits three bids,  $\beta(x), \delta(x), \theta(x)$ , where  $\beta, \delta, \theta$  are nondecreasing, continuously differentiable and satisfy  $\beta(x) \geq \delta(x) \geq \theta(x) \geq 0$ . First of all, it is not difficult to see that  $\beta(0) = \delta(0) = \theta(0)$  and  $\beta(\cdot)$  is strictly increasing. We first establish the following two lemmas (all proofs of the lemmas are relegated to the Appendix).

**Lemma 5.** *Consider a bidder with value  $x$ . If he receives three items in the auction, his optimal revenue from resale is  $\frac{23}{32} \beta^{-1}(\theta(x))$ .*

**Lemma 6.** *Consider a bidder with value  $x$  and  $\delta(x) = \theta(x)$ , if he receives two items in the auction, his optimal revenue from resale is  $\frac{5}{12} \beta^{-1}(\delta(x))$ .*

Next, suppose that we have an equilibrium in which  $\theta(1) > 0$ . Let us denote  $\beta^{-1}(\theta(1))$  by  $c$ . The following four lemmas provide us the conditions to prove that we cannot have an equilibrium of this kind.

**Lemma 7.** *For all  $x \in [0, c]$ , we have  $\beta(x) \leq \frac{23}{64}x$ .*

**Lemma 8.** (i) *For all  $t \in [0, \theta(1)]$ , if  $(\beta^{-1}(t))' > \frac{32}{23}$ , then  $\delta^{-1}(t) = \theta^{-1}(t)$ , and (ii) there exists  $d \in (0, 1]$  such that for all  $x \in [0, d]$ , then we have  $\beta(x) \in [0, \frac{23}{64}x]$  and  $\delta(x) = \theta(x)$ .*

**Lemma 9.** *There exists  $e \in (0, 1]$  such that for all  $x \in [0, e]$ , we have  $\beta(x) \in [0, \frac{23}{64}x]$  and  $\beta(x) = \delta(x) = \theta(x)$ .*

**Lemma 10.** *Suppose that we have an equilibrium such that  $\beta(x) = \delta(x) = \theta(x)$  for all  $x \in [0, e]$  for some  $e \in (0, 1]$ . Then, we have to have  $\beta(x) = \frac{3}{8}x$  for all  $x \in [0, e]$*

Finally, since  $\frac{3}{8} > \frac{23}{64}$ , Lemma 9 and Lemma 10 give us a contradiction. There cannot be an equilibrium in which  $\theta(1) > 0$ .

Next, we check the case where  $\theta(x) = 0$  for all  $x \in [0, 1]$ .

Let us consider an equilibrium in which  $\theta(x) = 0$ , but  $\delta(1) > 0$ . Similar to the above lemmas, we can first argue that we have to have

$$\beta(x) \leq \frac{5}{12}x.$$

Then we can argue that  $\beta(x) = \delta(x)$  for a neighborhood around zero (with arguments similar to Lemma 9).

Hence, let us consider an equilibrium that satisfies  $\beta(x) \in [0, \frac{5}{12}x]$  and  $\beta(x) = \delta(x)$  for all  $x \in [0, e]$ . We can then get a contradiction by arguments similar to Lemma 10 as follows.

Consider a bidder with value  $x \in (0, e)$  who bids as if his value is  $z$  very close to  $x$ . His expected utility is given by

$$u(x, z) = z^3 \left( x - 2\beta(z) + \frac{5}{12}z \right) + 3z^2(1 - z)(x - \beta(z)) + R(x, z)$$

where  $R(x, z)$  is his expected utility from the resale stage when he is a buyer and is given by:<sup>7</sup>

$$R(x, z) = 6 \int_z^{\max\{1, 2x\}} \int_l^{\max\{1, 2x\}} \int_{\frac{k}{2}}^x (x - m) dm dk dl + 6 \int_z^{\max\{1, 2x\}} \int_l^{\max\{1, 2x\}} \int_0^{\frac{k}{2}} \left( x - \frac{k}{2} \right) dm dk dl.$$

A necessary condition  $(\beta(x), \gamma(x))$  to be an equilibrium is

$$\left. \frac{\partial u(x, z)}{\partial z} \right|_{z=x} = 0.$$

---

<sup>7</sup>The variables in integrals  $k, l, m$  denote the realizations for highest, the second highest, and the third highest values among the competitors, the first term in the summation represents the case in which the bidder with value  $x$  pays the third highest value, and the last term represents the case in which the bidder with value  $x$  pays the reserve price.

The partial of the expected utility from the resale stage is:

$$\begin{aligned}\frac{\partial R(x, z)}{\partial z} \Big|_{z=x} &= -6 \left( \int_z^{\max\{1, 2x\}} \int_{\frac{k}{2}}^x (x - m) dm dk + \int_z^{\max\{1, 2x\}} \int_0^{\frac{k}{2}} \left(x - \frac{k}{2}\right) dm dk \right) \\ &= \frac{1}{4} \max\{1, 2x\}^3 - 3 \max\{1, 2x\} x^2 + \frac{11}{4} x^3 \\ &= \begin{cases} -\frac{5}{4} x^3 & \text{if } 0 \leq x \leq \frac{1}{2} \\ \frac{1}{4} - 3x^2 + \frac{11}{4} x^3 & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases}\end{aligned}$$

Hence, for  $x \leq \frac{1}{2}$  we have to have

$$\frac{\partial u(x, z)}{\partial z} \Big|_{z=x} = \frac{\partial}{\partial z} \left( z^3 \left( x - 2\beta(z) + \frac{5}{12} z \right) + 3z^2 (1 - z) (x - \beta(z)) \right) \Big|_{z=x} = \frac{5}{4} x^3$$

with the boundary condition  $\beta(x) = 0$ . The solution to this differential equation is given by

$$\beta(x) = \frac{35x^2 - 64x}{32x - 96}$$

which is greater than  $\frac{5}{12}x$  for all  $x \in [0, \frac{1}{2}]$ .<sup>8</sup>

Hence, there cannot be any equilibrium in which  $\beta(x) \in [0, \frac{5}{12}x]$  and  $\beta(x) = \delta(x)$  for all  $x \in [0, e]$  for some  $e \in [0, 1]$ .

To finish proof, we need to show that we cannot have an equilibrium where  $\delta(1) = 0$ . In this case, a bidder with value  $x$  mimicking a bidder with value  $z$  receives a payoff of

$$u(x, z) = \left(1 - (1 - z)^3\right) (x - \beta(z)) = z(z^2 - 3z + 3) (x - \beta(z)).$$

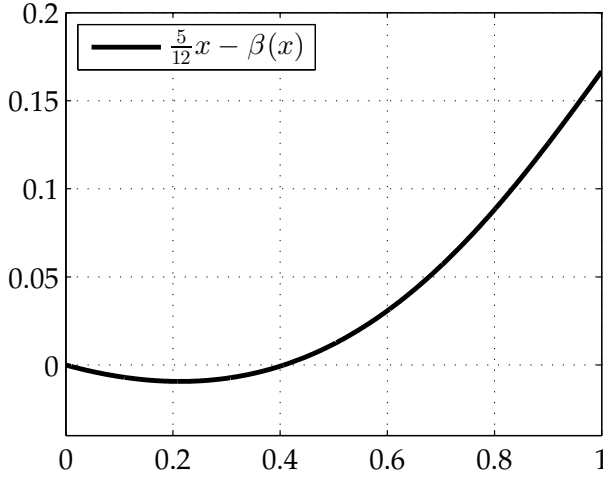
After solving for the necessary first-order conditions, the optimal bidding function can be found as:

$$\beta(x) = \frac{\frac{3}{4}x^4 - 2x^3 + \frac{3}{2}x^2}{x^3 - 3x^2 + 3x}.$$

But this solution cannot be an equilibrium since it is less than  $\frac{5}{12}x$  for  $x > 0.41$ :

---

<sup>8</sup>The difference is  $x \frac{72-65x}{96(3-x)}$ .



The bidders with values greater than 0.41 can benefit by bidding 0.17 or more for the second item and can make a positive profit from resale.

Hence, there is no monotone equilibrium for this example . □

We now recap the reasons behind the interesting finding of there is no symmetric and monotone equilibrium in this setup. This nonexistence result follows from the following observations: (i) consider a seller in a resale market in which all buyers' values are smaller than  $v$ ; if this seller sells two items, per unit optimal revenue is  $\frac{23}{64}v$ , whereas if the seller sells one item, per unit optimal revenue is  $\frac{5}{12}v$ , (ii) since  $\frac{5}{12} > \frac{23}{64}$ , in a neighborhood around 0 bid for the third item has to be the same as the bid for the second item. (iii) We can then argue that we have  $\beta(x) \leq \frac{23}{64}x$  (because otherwise the seller makes a loss from the resale market), and when that is the case bids for all three items have to be the same. Finally, (iv) when all three bids are the same, the differential equation around 0 gives us  $\beta(x) = \frac{3}{8}x$ , and since  $\frac{3}{8} > \frac{23}{64}$ , we have a contradiction.

Next, we discuss why the “monotone equilibrium existence results” of [Athey \(2001\)](#), [McAdams \(2003\)](#), and [Reny \(2011\)](#) do not contradict our “no symmetric monotone equilibrium finding.” First of all, while these three papers consider simultaneous move games, our game is a two stage game. Yet, since our equilibrium concept is WPBE, we can incorporate resale stage payoffs into auction stage and consider our game as a simultaneous move Bayesian game. Hence, their results may be applicable in our setup. However, the results in [Athey \(2001\)](#) and [McAdams \(2003\)](#) only concern the existence of a monotone equilibrium, not a symmetric monotone equilibrium. The only

papers that give existence results for a symmetric equilibrium in a symmetric game are those of [Reny \(1999\)](#) and [Reny \(2011\)](#). [Reny \(2011\)](#) has shown that (i) if the game satisfies 6 assumptions, G.1-G.6, then a monotone equilibrium exists, and (ii) if, in addition, the game is symmetric, then a symmetric monotone equilibrium exists. In our game, the assumptions G.1-G.5 are satisfied, but G.6—the continuity assumption—is not. Hence, the main result in [Reny \(2011\)](#) does not directly apply to our setup. In its applications section, [Reny \(2011\)](#) also shows that some Bayesian games that are not continuous (most relevantly discriminatory multi-unit auctions with CARA bidders) also have a monotone equilibrium. However, [Reny \(2011\)](#) does not establish existence of a symmetric monotone equilibrium in these applications.<sup>9</sup> Hence these methods are not applicable to our setup.

Next, we consider some variations of the model.

### 3.4 Variations

In this section, we consider some variations of the model where we allow for arbitrary  $k$  and  $n$  with  $n > k$ . First of all, we consider a case in which the sellers in the resale market cannot use reserve prices (for instance, because of commitment problems). In this variation, if there is one seller in the resale market he would use the “optimal efficient mechanism” which is a uniform price auction with no reserve price. For this case, we find two equilibria. One of them is the “no resale equilibrium”  $(\beta^N(x), 0, 0, \dots, 0)$  where  $\beta^N(x) = \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x]$ . The second one is an equilibrium in which all bidders bid the same amount. In particular, a bidder with value  $x$  bids  $(\beta^R(x), \beta^R(x), \dots, \beta^R(x))$  where

$$\beta^R(x) = \mathbb{E}[Y_k^{(n-1)} \mid Y_1^{(n-1)} < x] \quad (3.3)$$

---

<sup>9</sup>The proof in [Reny \(2011\)](#) is done by appealing to Remark 3.1 in [Reny \(1999\)](#) and showing that this game is “better-reply secure.” From [Reny \(2011\)](#)’s extensions, one may conjecture that if a game (i) is symmetric, (ii) is better-reply secure, and (iii) satisfies G.1-G.5, then there exists a symmetric monotone equilibrium. However, this conjecture is wrong as our game can be shown to be better-reply secure.

**Proposition 5.** *When sellers in the resale stage cannot use reserve prices, no resale equilibrium  $(\beta^N(x), 0, 0, \dots, 0)$  remains an equilibrium of the discriminatory auction with resale.*

The proof is relegated to the Appendix. The idea is straightforward. Consider a bidder who wins one additional unit for a bid  $b$ , then (in a second-price auction with no reserve price) expects to sell it for  $\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_{k-1}^{(n-1)} < \beta^{-1}(b) \right]$ . It is easy to see that we have  $\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_{k-1}^{(n-1)} < \beta^{-1}(b) \right] < \mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < \beta^{-1}(b) \right] = b$ . The same logic applies for winning more than one extra unit.

In the second equilibrium, bidders bid the same for each item. In this equilibrium there will be one bidder (with the highest value) who will win all  $k$  items, and he will sell  $k - 1$  items in the resale stage using a uniform-price auction.

**Proposition 6.** *When sellers in the resale stage cannot use reserve prices, the  $k$ -tuple  $(\beta^R(x), \beta^R(x), \dots, \beta^R(x))$  is an equilibrium of the discriminatory auction with resale where  $\beta^R(x)$  is given by Equation 3.3.*

The proof is relegated to the Appendix. In the proof, we carefully check for each deviation and show that no deviation can make a bidder better off. Note that the above two equilibria results in very different allocations after the auction stage. In the first equilibrium,  $k$  bidders with the highest values obtain the units, whereas in the second equilibrium the highest-valued bidder obtains all the units. Yet, after the resale stage they result in the same allocation:  $k$  bidders with the highest values obtain the units. The auctioneer's revenues in these two equilibria also seem quite different. In the first equilibrium, the revenue is given by

$$\mathbb{E} \left[ \sum_{l=1}^k \beta^N \left[ Y_l^{(n-1)} \right] \right]$$

whereas in the second one, it is given by

$$k \times \mathbb{E} \left[ \beta^R \left( Y_1^{(n-1)} \right) \right]$$

However, they are equal to each other. We establish that in the following Proposition.

**Proposition 7.** *The two equilibria  $((\beta^N(x), 0, 0, \dots, 0)$  and  $(\beta^R(x), \beta^R(x), \dots, \beta^R(x))$  are revenue equivalent.*

The proof is relegated to the Appendix. In the proof, we directly show that two revenue expressions turn out to be identical to each other. We also would like to note that this result can be also obtained by appealing to the revenue equivalence principle. This can be argued by noting that (i) the two equilibria result in the same (efficient allocation), (ii) the expected payment of the bidder with value 0 is 0 under both equilibria, and (iii) the transfers between the bidders in the resale stage aggregate to 0.

As a corollary to Proposition 7, we establish the following result.

**Corollary 1.** *When there are two units for sale and resellers can use reserve prices in the resale market, banning the resale market strictly decreases the expected revenue in a discriminatory price auction.<sup>10</sup>*

This corollary can be simply obtained by the following observations: (i) if there is no resale market, the revenue is equal to the revenue from the symmetric strategies  $(\beta^R(x), \beta^R(x))$  (Proposition 7 and the fact that  $(\beta^N(x), 0)$  is an equilibrium of discriminatory auctions with no resale); (ii) with the resale market, the revenue is equal to the revenue from symmetric strategies  $(\gamma(x), \gamma(x))$  (Theorem 1); and (iii) we have  $\gamma(x) > \beta^R(x)$  for all  $x \in (0, 1]$  since  $\gamma(x)$  is the revenue from the optimal auction and  $\beta^R(x)$  is the revenue from a second-price auction (when there are  $k - 1$  buyers who have values smaller than  $x$ ).

Next, we consider the case where reserve prices are allowed for both the auction stage and the resale stage. More specifically, consider a discriminatory price auction with a reserve price  $r_1^* \equiv \psi^{-1}(0)$  and where a seller in the resale market uses an optimal auction (and hence can use reserve prices). We show the following.

**Proposition 8.** *The standard equilibrium  $(\beta^{RR}(x), 0, \dots, 0)$  where  $\beta^{RR}(x) = \mathbb{E}[\max\{Y_k^{n-1}, r_1^*\} \mid Y_k^{n-1} < x]$ , is also the equilibrium of the bidding stage of a game where a reserve price is allowed in both the bidding and resale stage.*

---

<sup>10</sup>More specifically, the revenue in the symmetric and monotone equilibria we have found in the model with resale is higher than that of in the model without resale.



The proof is relegated to the Appendix. Also, it turns out that we cannot have an equilibrium in which bidders bid the same for all items (as in Proposition 6).

### 3.5 Conclusion

In this paper, we consider an environment where ex-ante symmetric bidders who have private single-unit demands can engage in post-auction resale after participating in a discriminatory (pay-your-bid) auction. This environment without resale opportunities result in an efficient allocation of items. Hence one might expect that adding resale opportunities will not change the equilibrium behavior. We prove this intuition wrong by observing that this auction results in low prices to attract speculative behavior (buying and then selling in the resale stage). We find an equilibrium when there are 2 units for sale, and show that there may be no symmetric and monotone equilibrium when there are 3 or more units for sale. We then consider some variations of the model. If reserve prices cannot be used in the resale market, then there are two revenue equivalent equilibria. If the auction also includes a reserve price, then no resale equilibrium is an equilibrium of the game.

Overall, we establish that the possibility of resale—even when the equilibrium without resale is efficient—may have significant effects on the auction outcome: equilibrium without resale is not an equilibrium, and for some cases there may not be any symmetric and monotone equilibrium. Finding a (non symmetric monotone, or arbitrary non monotone) equilibrium when there is no symmetric and monotone equilibrium is left as an open question.



# Appendix A

## Appendix for Chapter 1

### A.0.1 Mechanism Design: Two-Dimensional Private Information

In this section, I show the implementability conditions for a two-dimensional private information problem. I approach it similarly to [Jacquet et al. \(2013\)](#) and [Kleven et al. \(2009\)](#). I differ from these works in two ways. First, both of these papers consider two groups of households. Yet, the families can have an arbitrary number of children in my paper. So I have a more general model. Second, the previous works do not consider the time effect of secondary shock. However, in this work, any existing child requires parental time, which is perfectly substitutable with market labor.

Let  $\gamma = (\beta, \theta) \in \mathcal{B} \times \Theta = \Gamma$  be the private information of a family. If the family reports  $\gamma$  as their type, the government chooses optimal  $c(\gamma), n(\gamma), z(\gamma)$ . This mechanism should satisfy the revelation principle, by which any government mechanism can be decentralized by a truthful mechanism  $(z(\gamma), n(\gamma), c(\gamma))_{\gamma \in \Gamma}$  such that

$$U(\gamma, \gamma) \geq U(\gamma, \gamma').$$

In this setup, a strategy has two dimensions, and hence a possible mimicking strategy has two dimensions. However, the possibility of double deviation in the mimicking strategy can be eliminated and double deviation can be reduced to single deviation by two constraints: indirect utility of  $n$  child families (1.4) and threshold tastes for children (1.5) for each  $n$ .

From the classical mechanism design problem to a pseudo-mechanism design problem, I first show that the solution to the classical problem can be replaced by a pseudo-problem solution in

the next Lemma.

**Lemma 11.** Any truthful mechanism  $(z(\gamma), n(\gamma), c(\gamma))_{\gamma \in \Gamma}$  can be replaced by a new mechanism  $(c_n(\theta), z_n(\theta))_{n \in \{0, 1, \dots, N\}, \theta}$  such that

- for each  $\theta$  and for each  $n$ , there is a  $\beta_n(\theta)$  such that if  $\beta \in (\beta_n(\theta), \beta_{n+1}(\theta))$ ,<sup>1</sup> then  $U(z_n(\theta), n, c_n(\theta), \gamma) \geq \max_{\gamma'} U(\gamma, \gamma')$ , and
- the new mechanism is truthful and provides as much as taxes collected by the original mechanism.

*Proof.* For each  $\theta$ , partition the set  $\mathcal{B}$  into  $N + 1$  sets such that if  $\beta \in \mathcal{B}_j$  then  $n(\beta, \theta) = j$  for  $j = \{0, 1, \dots, N\}$ . If the family is indifferent between having  $k$  children and  $k + 1$  children I assume that  $n(\beta, \theta) = k + 1$ .

For a given  $\theta$  and  $\beta, \beta' \in \mathcal{B}_j$ , the truthfulness of the original mechanism implies:

$$\begin{aligned} u(c(\beta, \theta)) - h \left( \frac{z(\beta, \theta)}{\theta} + b_j \right) \theta + m(j, \beta) &\geq u(c(\beta', \theta)) - h \left( \frac{z(\beta', \theta)}{\theta} + b_j \right) \theta + m(j, \beta) \\ u(c(\beta', \theta)) - h \left( \frac{z(\beta', \theta)}{\theta} + b_j \right) \theta + m(j, \beta') &\geq u(c(\beta, \theta)) - h \left( \frac{z(\beta, \theta)}{\theta} + b_j \right) \theta + m(j, \beta'). \end{aligned}$$

The first inequality is  $U((\beta, \theta), (\beta, \theta)) \geq U((\beta, \theta), (\beta', \theta))$  and the second inequality is  $U((\beta', \theta), (\beta', \theta)) \geq U((\beta', \theta), (\beta, \theta))$ . It is easy to see  $U((\beta, \theta), (\beta, \theta)) = U((\beta', \theta), (\beta', \theta))$ , which implies  $u(c(\beta, \theta)) - h \left( \frac{z(\beta, \theta)}{\theta} + b_j \right) \theta$  is constant for all  $\beta \in \mathcal{B}_j$ , and let  $V_j(\theta)$  be its value.

Note that at least as much taxes should be collected with the new mechanism. Let  $Z_j(\theta) = \{z(\beta, \theta) | \beta \in \mathcal{B}_j(\theta)\}$ . Define  $t = \sup_{z \in Z_j(\theta)} z - u^{-1}(V_0(\theta) + h(\frac{z}{\theta} + b_j)\theta)$ . Note that  $z - u^{-1}(V_j(\theta) + h(\frac{z}{\theta} + b_j)\theta)$  is a weakly concave function in  $z$  and reaches maximum for a  $\tilde{z}$  value and goes to  $-\infty$  when  $z \rightarrow \infty$ . So there is a  $z_j(\theta) \in \overline{Z_j(\theta)}$  such that  $t = z_j(\theta) - u^{-1}(V_j(\theta) + h(\frac{z_j(\theta)}{\theta} + b_j)\theta)$ .<sup>2</sup> Define  $c_j(\theta) := u^{-1}(V_j(\theta) - h(\frac{z_j(\theta)}{\theta} + b_j)\theta)$ . Note that  $(c_j(\theta), z_j(\theta))$  maximizes the taxes over the closure of the set  $(c(\beta, \theta), z(\beta, \theta))_{\beta \in \mathcal{B}_j(\theta)}$ . These procedures can be followed for all  $j = 0, 1, \dots, N$ .

Finally, I define  $\beta_n(\theta) := M^{-1}(V_n(\theta) - V_{n+1}(\theta))$  where  $M(\beta) := m(n + 1, \beta) - m(n, \beta)$  for all  $n = 0, 1, \dots, N$ .<sup>3</sup>  $\beta_n(\theta)$  are the threshold tastes for children for each  $\theta$  and for each  $n$ . Note that truthfulness of original mechanism implies: for all  $\beta \in \mathcal{B}_j(\theta)$  the family chooses  $n = j$  and

<sup>1</sup>I let  $\beta_0 = \underline{\beta}$  and  $\beta_{N+1} = \overline{\beta}$ .

<sup>2</sup> $\overline{Z_j(\theta)}$  is the closure of the  $Z_j(\theta)$

<sup>3</sup>Let  $\beta_0 = \underline{\beta}$  and  $\beta_{N+1} = \overline{\beta}$ .

$(z_j(\theta), c_j(\theta))$ , i.e.  $V_j(\theta) + m(j, \beta) \geq V_{j'}(\theta) - m(j', \beta)$  for all  $j' = 0, 1, \dots, N$ . Pick  $j' = j - 1$  and  $j' = j + 1$ . Then it is easy to see that  $\underbrace{M(V_j(\theta) - V_{j+1}(\theta))}_{\beta_j(\theta)} \geq \beta \geq \underbrace{M(V_{j-1}(\theta) - V_j(\theta))}_{\beta_{j-1}(\theta)}$ .<sup>4</sup> Therefore  $\mathcal{B}_j(\theta) = (\beta_{j-1}(\theta), \beta_j(\theta))$ .

All is left to show the new mechanism  $(c_n(\theta), z_n(\theta))_{n \in \{0, 1, \dots, N\}, \theta \in \Theta}$  is truthful. First I show it is truthful within families with the same number of children: For all  $\theta, \theta', \beta \in \mathcal{B}_j(\theta), \beta' \in \mathcal{B}_j(\theta')$ :

$$U(c_j(\theta), z_j(\theta), (\beta, \theta)) = V_j(\theta) + m(j, \beta) \geq u(c(\beta', \theta')) - h\left(\frac{z(\beta', \theta')}{\theta}\right) - m(j, \beta)$$

where the inequality is from the truthfulness of the initial mechanism.<sup>5</sup> As a result,

$$U(c_j(\theta), z_j(\theta), (\beta, \theta)) \geq U(c_j(\theta'), z_j(\theta'), (\beta, \theta)).$$

I also show the mechanism is truthful cross-sectionally: for all  $\theta, \theta', \beta \in \mathcal{B}_j(\theta), \beta' \in \mathcal{B}_{j'}(\theta')$ :

$$\begin{aligned} U(c_j(\theta), z_j(\theta), (\beta, \theta)) &= V_j(\theta) + m(j, \beta) \geq V_{j'}(\theta) + m(j', \beta) \\ &\geq u(c(\beta', \theta')) - h\left(\frac{z(\beta', \theta')}{\theta}\right) + m(j', \beta) \end{aligned}$$

where the first inequality comes from the definition of  $\beta_n$  and the second inequality is satisfied by the truthfulness of the original truthful mechanism. Hence:

$$U(c_j(\theta), z_j(\theta), (\beta, \theta)) \geq U(c_{j'}(\theta'), z_{j'}(\theta'), (\beta, \theta)).$$

This procedure can be followed for any  $j = 0, 1, \dots, N$ . As a result, the proof is completed.  $\square$

This lemma reduces the two-dimensional schedule to a one-dimensional schedule, from  $c(\beta, \theta), c(\beta, \theta), z(\beta, \theta)$  to  $\{c_n(\theta), z_n(\theta)\}_{n=0, 1, \dots, N}$ . As a result, I can directly use the one-dimensional implementation requirement as long as the single-crossing condition is satisfied.

**Definition 1.**  $z_n(\theta)_{n \in \{0, 1, \dots, N\}}$  is implementable if and only if there exist transfer functions  $c_n(\theta)_{n \in \{0, 1, \dots, N\}}$  such that  $(c_n(\theta), z_n(\theta))_{n \in \{0, 1, \dots, N\}, \theta \in \Theta}$  is a truthful mechanism.

<sup>4</sup>Note that I let  $m$  to be concave in its first dimension and therefore  $\beta_{j-1}(\theta) < \beta_j(\theta)$ .

<sup>5</sup>Note that  $(c_j(\theta'), z_j(\theta'))$  is in the closure of the set  $(c(\beta, \theta), z(\beta, \theta))_{\beta \in \mathcal{B}_j(\theta)}$ .

In the following lemma, I prove that a one-dimensional requirement is sufficient for the two-dimensional problem in this framework:

**Lemma 12.** *The income profile  $z_n(\theta)_{n \in \{0,1,\dots,N\}}$  for all  $\theta \in \Theta$  is implementable if and only if  $\dot{z}_n \geq 0$ .*

*Proof.* Note that  $u(c) - h\left(\frac{z}{\theta} + b_n\right)\theta$  satisfies the classic single crossing condition. The one-dimensional implementability condition is that:  $\dot{z} \geq 0$  if and only if there is  $c(\theta)$  such that  $u(c(\theta')) - h\left(\frac{z}{\theta} + b_n\right)\theta \geq u(c(\theta)) - h\left(\frac{z(\theta')}{\theta} + b_n\right)\theta$  for all  $\theta, \theta'$ .

For the "if" side of the lemma, I directly apply the one-dimensional implementability condition: for all  $n = 0, 1, \dots, N$ , let  $z_n(\theta)_{n \in \{0,1,\dots,N\}}$  is implementable. Then for a particular  $n$ , truthfulness implies  $u(c_n(\theta)) - h\left(\frac{z_n(\theta)}{\theta} + b_n\right) \geq u(c_n(\theta')) - h\left(\frac{z_n(\theta')}{\theta} + b_n\right)$  for all  $\theta, \theta'$ . As a result, the one-dimensional result suggests that for each  $n = 0, 1, \dots, N$  income is non-decreasing:  $\dot{z}_n \geq 0$ .

Now let  $\dot{z}_n \geq 0$ . Similarly, using the one-dimensional result, there is  $c_n(\theta)$  such that  $u(c_n(\theta)) - h\left(\frac{z_n(\theta)}{\theta} + b_n\right) \geq u(c_n(\theta')) - h\left(\frac{z_n(\theta')}{\theta} + b_n\right)$  for all  $\theta, \theta'$ .

Within sections, the one-dimensional condition is directly applicable, as shown above. All that is need to be shown is that cross-sectional truth-telling is satisfied. Note that the steps are similar in the proof of previous lemma where I show that cross-sectional deviation is not profitable. Hence I skip it here.  $\square$

### A.0.2 Proof of Proposition 1

*Proof.* The Hamiltonian of the problem is:

$$\begin{aligned} \mathcal{H} = & \sum_{n=0}^{n=N} \int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \left( \Psi(V_n(\theta) + m(n, \beta)) + \lambda[z_n(\theta) - c_n(\theta)] \right) p(\beta|\theta) f(\theta) d\beta \\ & + \sum_{n=0}^{n=N} \mu_n(\theta) \left( -h\left(\frac{z_n}{\theta} + b_n\right) + h'\left(\frac{z_n}{\theta} + b_n\right) \frac{z_n}{\theta} \right) \end{aligned} \quad (\text{Hamiltonian})$$

where  $\mu_n(\underline{\theta}) = \mu_n(\bar{\theta}) = 0$  and  $\beta_0 = \underline{\beta}$ , and  $\beta_{N+1} = \bar{\beta}$ . I assume that the inequality of Equation (1.7) never binds for no bunching. Therefore the implementability condition holds,  $\dot{z}_n \geq 0$ .<sup>6</sup>

The first-order conditions for  $z_n$  are:<sup>7</sup>

<sup>6</sup>Numeric exercises show that there is no bunching.

<sup>7</sup>I assume that the implementability constraint does not bind and show ex-post is the case.

$$z_n(\theta) : \quad \lambda \left( 1 - \frac{\partial c_n}{\partial z_n} \right) (P(\beta_{n+1}|\theta) - P(\beta_n|\theta)) f(\theta) = -\frac{\mu_n}{\theta} h'' \left( \frac{z_n(\theta)}{\theta} + b_n \right) \frac{z_n(\theta)}{\theta}$$

for all  $n = 0, 1, \dots, N$ . Also, the co-states of the system are:<sup>8</sup>

$$\begin{aligned} \frac{-\dot{\mu}_n}{\lambda f(\theta)} &= \int_{\beta_n}^{\beta_{n+1}} \left( \frac{\Psi'(V_n - m((N-n), \beta))}{\lambda} - \frac{\partial c_n}{\partial V_n} \right) p(\beta|\theta) d\beta \\ &\quad + \Delta T_{n-1}(\theta) p(\beta_n|\theta) \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta) p(\beta_{n+1}|\theta) \frac{\partial \beta_{n+1}}{\partial V_n} \end{aligned}$$

for  $n = 0, \dots, N$ , where  $T_n(\theta) = z_n(\theta) - c_n(\theta)$  are the optimal income taxes collected from the  $\theta$ -ability families with  $n$  children and for  $n = 0, \dots, N-1$ ,  $\Delta T_n(\theta) = T_n(\theta) - T_{n+1}(\theta)$  is the tax credit for an extra child for  $\theta$ -ability families.<sup>9</sup>

Using the terminal conditions, I derive the co-states:

$$\begin{aligned} -\frac{\mu_n(\theta)}{\lambda} &= \int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} (P(\beta_{n+1}|\theta') - P(\beta_n|\theta')) \right. \\ &\quad \left. + \Delta T_{n-1}(\theta) p(\beta_n|\theta) \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta) p(\beta_{n+1}|\theta) \frac{\partial \beta_{n+1}}{\partial V_n} \right] f(\theta') d\theta' \end{aligned}$$

where

$$\begin{aligned} g_n(\theta) &= \mathbb{E}_{\beta} \left[ \frac{\Psi'(V_n - m((N-n), \beta)) u'(c_n)}{\lambda} | \beta_n < \beta < \beta_{n+1} \right] \\ &= \frac{\int_{\beta_n}^{\beta_{n+1}} \Psi'(V_n - (N-n)\beta) u'(c_n) p(\beta|\theta) f(\theta) d\beta}{\lambda (P(\beta_{n+1}|\theta) - P(\beta_n|\theta)) f(\theta)} \end{aligned} \tag{A.1}$$

is the marginal weight associated by the government to the family  $\theta$  with  $n$  children, which is the cost of giving an extra dollar of consumption to the family in terms of public goods. Note that  $g_n(\theta)$  is shaped by the government preference. If, for example, the government is a Benthamite government, i.e.  $\Psi(x) = x$  then the government weights can be further simplified:  $g_n(\theta) = \frac{u'(c_n)}{\lambda}$ . If the government is Rawlsian, i.e.  $\Psi(V_n(\theta)) = 0$  for all  $\theta > \underline{\theta}$  and  $\Psi(V_n(\theta_L)) > 0$ , then the

<sup>8</sup>I derive co-state for  $V_0(\theta)$  and  $V_N(\theta)$  separately.

<sup>9</sup>When  $n = 0$ , let  $\Delta T_{-1}(\theta) = 0$  and similarly when  $n = N$ , let  $\Delta T_N(\theta) = 0$ .

government only values the lowest-ability families' consumption and income, hence  $g_n(\theta) = 0$  for all  $\theta > \underline{\theta}$ .

Combining the of previous terms shows that the optimal tax function should satisfy:

$$\begin{aligned} \frac{T'_n}{1 - T'_n} = \frac{1}{\varepsilon_n} \times \frac{1}{\theta f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta))} \times \\ \int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} (P(\beta_{n+1}|\theta') - P(\beta_n|\theta')) \right. \\ \left. + \Delta T_{n-1}(\theta) p(\beta_n|\theta) \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta) p(\beta_{n+1}|\theta) \frac{\partial \beta_{n+1}}{\partial V_n} \right] u'(c_n(\theta)) f(\theta') d\theta' \end{aligned} \quad (\text{A.2})$$

for  $n = 0, 1, \dots, N$ .<sup>10</sup> □

### A.0.3 Family Income Elasticity

Let  $\varepsilon_m := \frac{\partial \log z_m}{\partial \log(1-\tau)}$  be the elasticity of male income with respect to net marginal tax rates. Similarly, let  $\varepsilon_f$  represents the female income elasticity. In this work, I focus on married households who file tax returns jointly. According to the US tax code, the next dollar earned by a family member is marginally taxed unconditional on gender. So if the family income is the sum of earnings of couples, i.e.  $z = z_m + z_f$ , the family income elasticity is:

$$\varepsilon := \frac{\partial \log z}{\partial \log(1-\tau)} = \frac{(1-\tau)}{z} \frac{\partial z}{\partial (1-\tau)} = \frac{(1-\tau)}{z_f + z_m} \frac{\partial (z_f + z_m)}{\partial (1-\tau)} = \frac{z_f}{z_f + z_m} \varepsilon_f + \frac{z_m}{z_f + z_m} \varepsilon_m.$$

This means that the family elasticity is a convex combination of elasticities.

To figure out family income elasticity, I need to find  $\varepsilon_f$ ,  $\varepsilon_m$ , and the share of female earnings of family income. Note that the utility function is quasi-linear in consumption and hence elasticity of income with respect to net marginal tax rates is equal to the Frisch elasticity of labor supply. Therefore I look at the literature on Frisch elasticity.

There is a voluminous literature on elasticity of labor supply. [Pencavel \(1986\)](#) and [Keane \(2011\)](#) give an excellent survey of labor responses and taxes. They state that the median value is 0.2 for Frisch elasticity of men although the former gives a range from zero to 0.5 and the latter gives a range from zero to 0.7. Some of the works in these surveys use non-US data. Hence, I look particularly at [French \(2005\)](#) and [Ziliak and Kniesner \(2005\)](#) who use Panel Study of Income Dynamics

---

<sup>10</sup>I let  $\Delta T_{-1}(\theta) = 0$  when  $n = 0$  and  $\Delta T_N(\theta) = 0$  when  $n = N$ .



(PSID) data. The former estimates the Frisch elasticity of men at 0.3 and the latter estimates around 0.5. I take the average value  $\varepsilon_m = 0.4$  in my setup.<sup>11</sup>

The research on Frisch elasticity of females is not as large as on male elasticities. [Blundell et al. \(2012\)](#) estimate that the elasticity of married women lies between 0.8 to 1.1. When the utility is additive separable, the estimate is 0.8, and I pick  $\varepsilon_f = 0.8$ . Note that they use dummies for existing children, and hence I can use these values immediately.

Note the convex combination coefficient is the fraction of female (male) earnings. In my sample, females earn around 39% of the family income (see Figure A.1). Hence,  $\varepsilon = 0.61 \times 0.4 + 0.39 \times 0.8 \simeq 0.56$ .

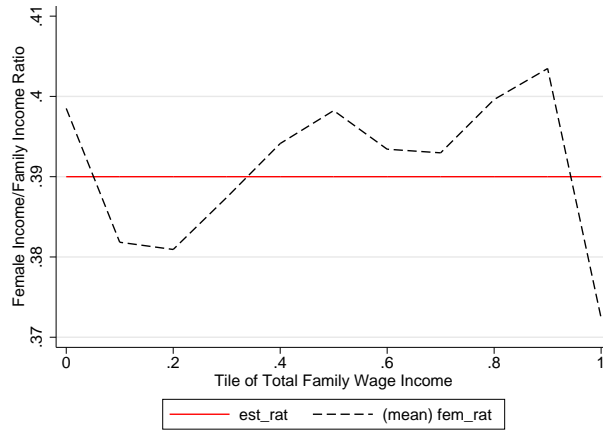


Figure A.1: Evolution of the ratio of female income for different income groups

The ratio is very close to 0.39. Note that this graph suggests that the gender gap for this sample is 0.64, which is quite close to the actual gender gap in the US (0.7).

#### A.0.4 Evolution of $b_n$

$b_n$  is the ratio of time devoted child care to the total time devoted to market and child care. I use ATUS 2003 to find  $b_n$  for different income groups. The data set contains individual time devoted to many different categories.<sup>12</sup> The data set also contains weekly earnings of individuals. Hence, I am able to derive  $b_n$  for different income groups:

<sup>11</sup>[Blundell et al. \(2012\)](#) finds that the Frisch elasticity of married men is 0.4. For different models the value goes up to 0.6.

<sup>12</sup>Refer to [Aguilar and Hurst \(2007\)](#) for further details.

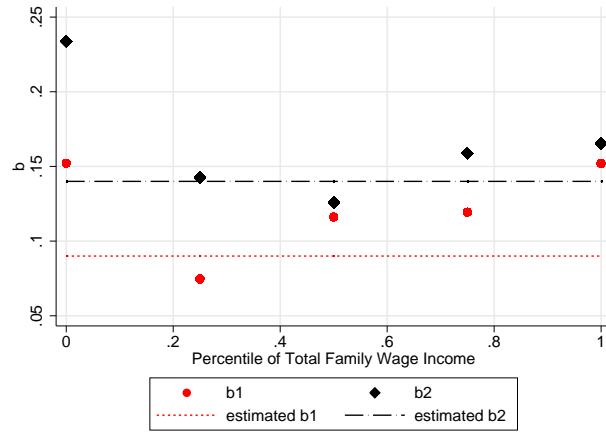


Figure A.2: Evolution of  $b_n$  for different income groups

The numbers represent the individual levels. If the sample contains only males (females), the fraction equals 0.06 (0.12) and 0.10 (0.18) for 1-child and 2-children families, respectively. As a result, since the time endowment is normalized, I use the average values.



## A.0.5 Figures and Tables

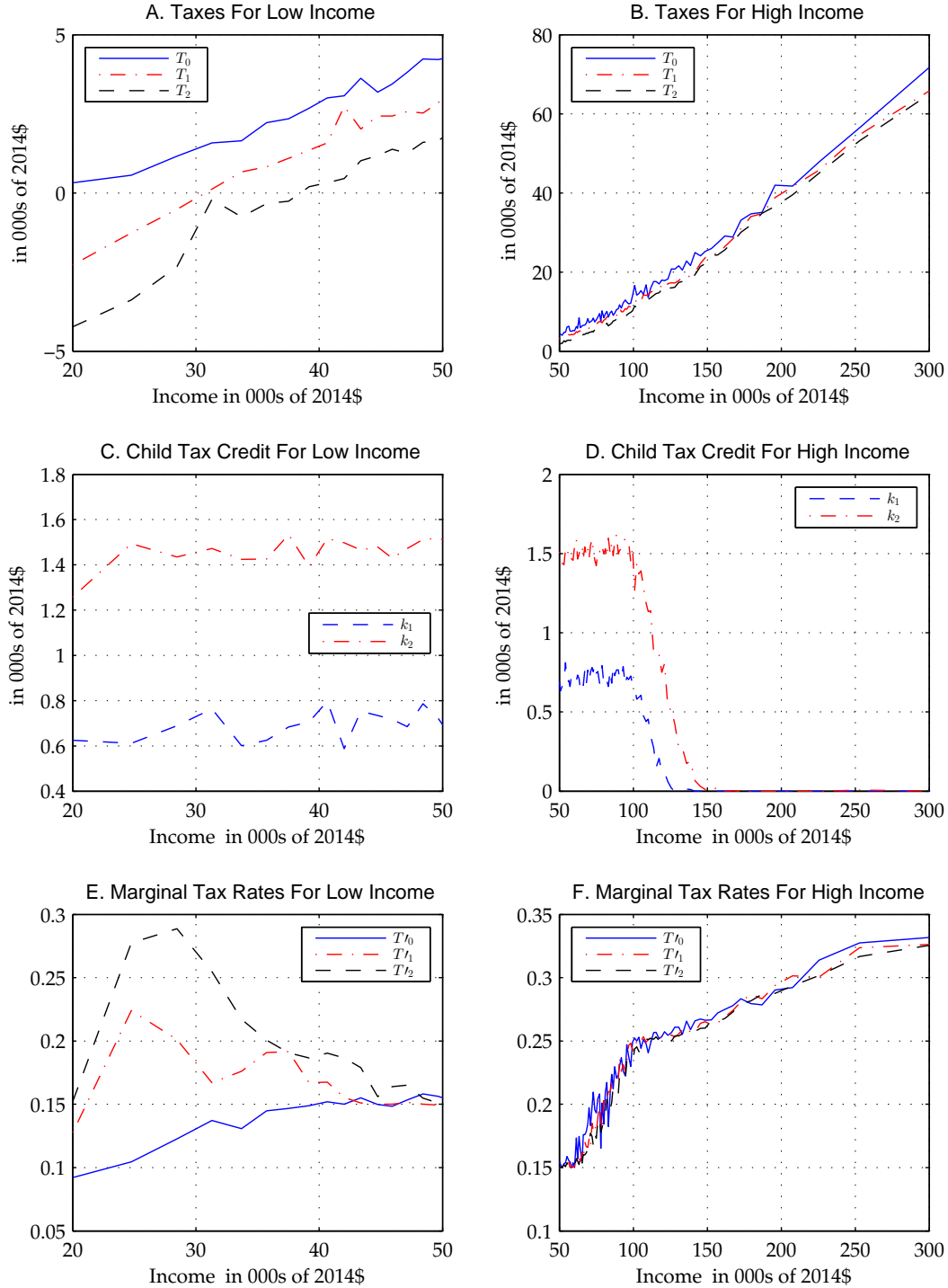


Figure A.3: Current Tax System

Left (right) panel is for families with less (more) than \$50,000. The first row shows the actual taxes paid by families. The second row shows how much tax credits they get. Note that  $k_1$  ( $k_2$ ) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families.

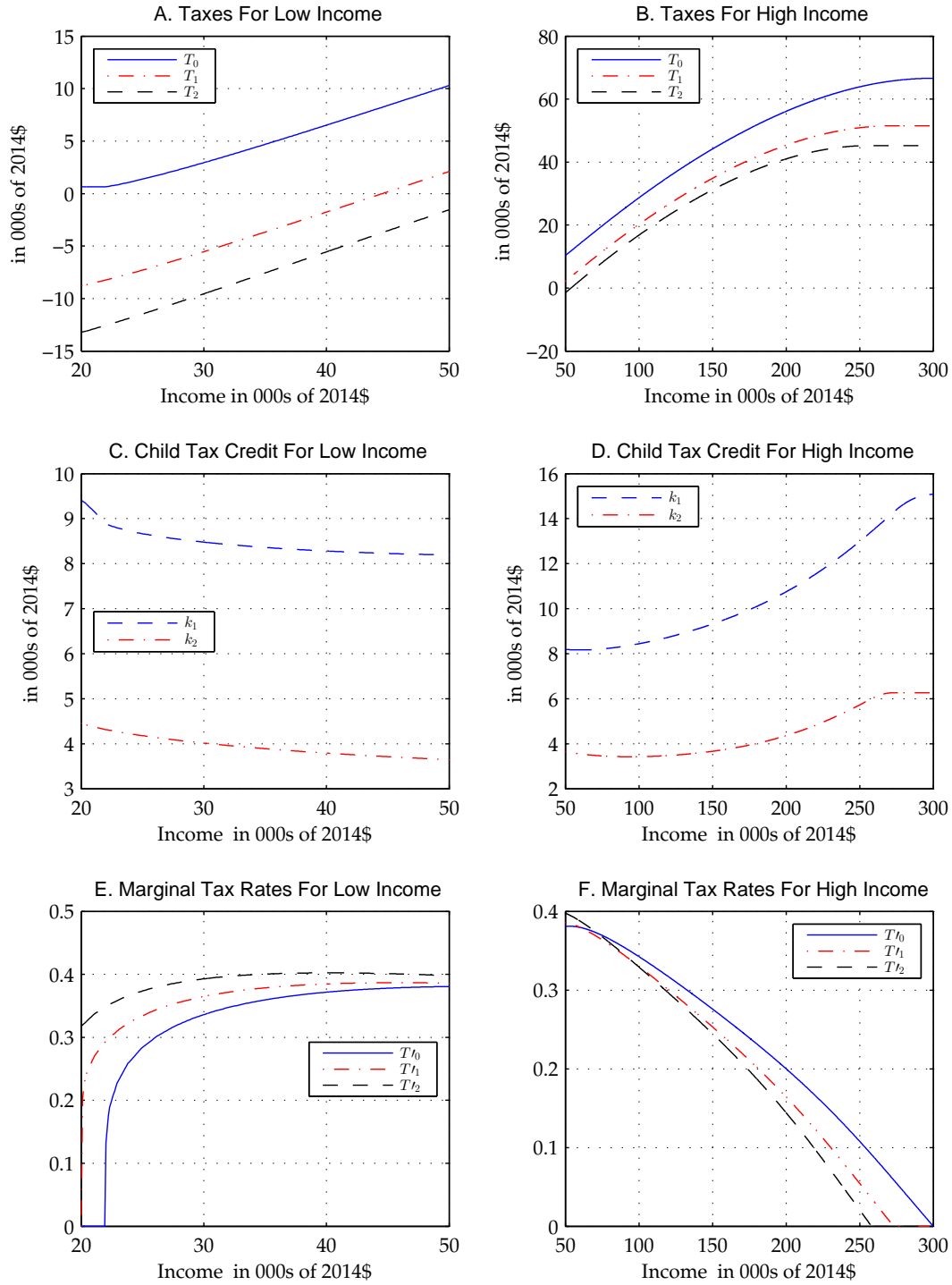


Figure A.4: Optimal Tax System

Left (right) panel is for families with less (more) than \$50,000. The first row shows the actual taxes paid by families. The second row shows how much tax credits they get. Note that  $k_1$  ( $k_2$ ) represents the tax credits for two (one) child families. The last row shows the federal marginal tax rates of families.

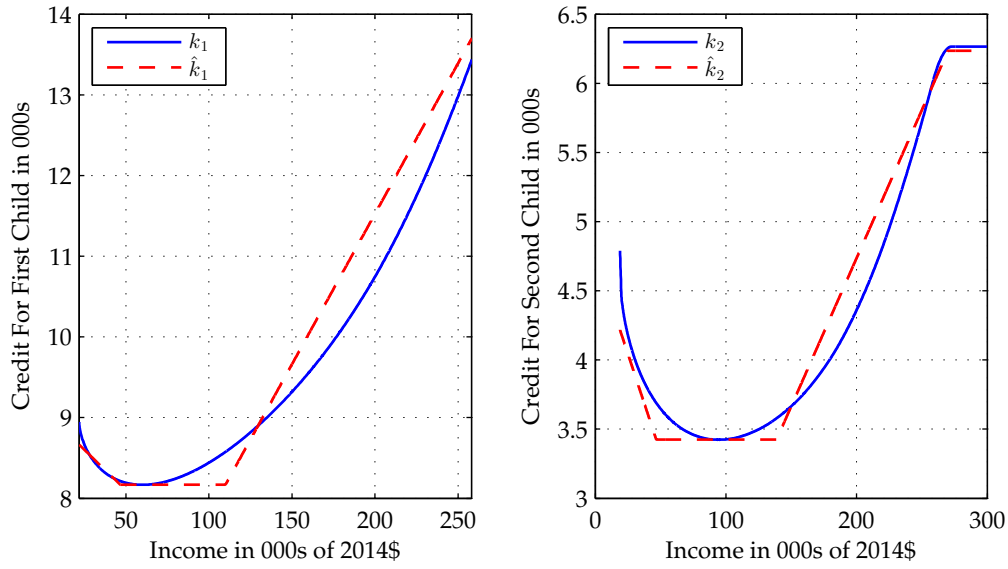


Figure A.5: Proposed Tax Credits

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$.  $k_n$  is the optimal tax credits and  $\hat{k}_n$  is the proposed tax credit for  $n = 1, 2$ .

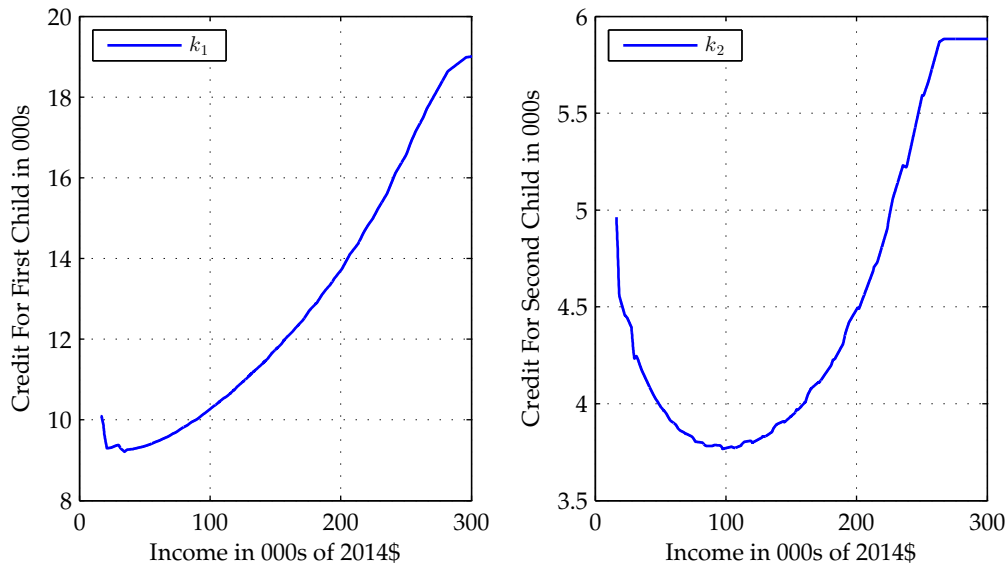


Figure A.6: Tax Credits: Age Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$.  $k_n$  is the optimal tax credits for  $n = 1, 2$ .

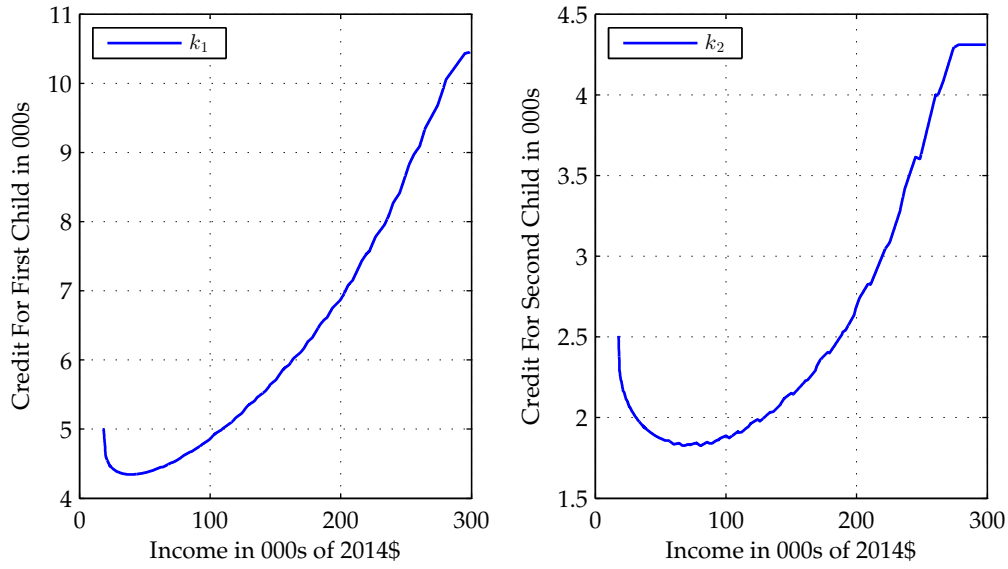


Figure A.7: Tax Credits: Costs Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$.  $k_n$  is the optimal tax credits for  $n = 1, 2$ .

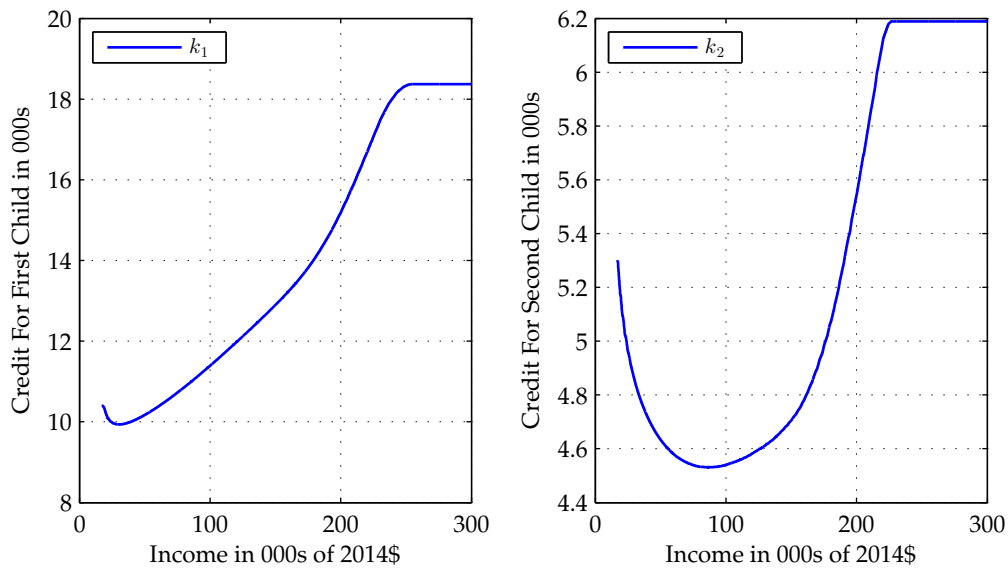


Figure A.8: Tax Credits: Marital Status Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014\$.  $k_n$  is the optimal tax credits for  $n = 1, 2$ .

# of children	earnings $\leq$	credit rate	max credit	phase-out begins	phase-out rate
0	6,480	0.08	496	13,540	0.08
1	9,720	0.34	3,305	23,260	0.16
2	13,650	0.40	5,460	23,260	0.21
3	13,650	0.45	6,143	23,260	0.21

Table A.1: Earned Income Tax Credit Phase in and Phase out regions for 2014

Credits are in terms of \$. The numbers in cells present values for married couples who fill taxes jointly. First column is for number of children. Second column shows maximum earnings to be eligible for the credit.



## Appendix B

# Appendix for Chapter 2

### B.1 Mechanism Design Formulation

It is straightforward to verify that any allocation that solves the mechanism design problem (MDP) is implementable as part of a tax equilibrium. On the other hand, the allocation from a tax equilibrium is feasible for (MDP). Consequently, an optimal tax equilibrium may be constructed from a solution to (MDP)  $\{c_k^*, e_k^*\}_{k=1}^K$  by associating with it the wages and taxes needed to implement this solution.

We make two preliminary observations on solutions to (MDP). First, given a solution  $\{c_k^*, e_k^*\}_{k=1}^K$ , worker types may be ordered according to their optimal shadow wages  $\{w_k^*\}_{k=1}^K$ . Types whose wages are tied may be further ordered by their pre-tax incomes  $q_k^* = w_k^* e_k^*$ . Types may then be relabeled accordingly (with ties between both wage and income ordered arbitrarily). Thus, the  $k$ -th worker type has a wage that is weakly greater than the wages of types 1 to  $k - 1$  and if the  $k$ -th type's wage ties with the  $k - 1$ -th type, then its income is weakly greater. We impose this labeling below. Second, only a subset of incentive constraints (2.2) bind. Recall that a  $(k, j)$ -th incentive constraint is local if  $j \in \{k - 1, k + 1\} \cap \{1, \dots, K\}$ ; otherwise it is non-local. A well known consequence of the Spence-Mirrlees single crossing property and the structure of the incentive constraints in settings with exogenous wages is that non-local incentive constraints do not bind at an optimum. This result continues to hold in the present setting under our ordering.<sup>1</sup> We record this

---

<sup>1</sup>We omit the proof. It follows from a slight modification of Theorems 3 and 4 in [Milgrom and Shannon \(1994\)](#). In our setting it is possible for two worker types  $k$  and  $k + 1$  to have the same wage, but different efforts, incomes and consumptions at the optimum. If the  $k$ -th type has a higher income than the  $k + 1$ -th type, then it is possible that the

fact in Lemma A.

**Lemma 13.** Let  $\{c_k^*, e_k^*\}_{k=1}^K$  denote a solution to (MDP) with corresponding shadow wages  $\{w_k^*\}_{k=1}^K$ ,  $w_k^* = F_k(e_1^* \pi_1, \dots, e_K^* \pi_K)$  and incomes  $q_k^* = w_k^* e_k^*$  and types labeled consistently with their ranking in the wage (and when wages are tied income) distribution, then (i)  $c_{k+1}^* \geq c_k^*$  and  $q_{k+1}^* \geq q_k^*$ , and (ii) non-local incentive constraints do not bind.

As in the main text, we gather the constraint functions from the incentive and resource constraints into the single function  $\mathcal{G} : \mathbb{R}_{++}^{2K} \rightarrow \mathbb{R}^{K(K-1)+1}$  and say that  $\{c_k, e_k\}_{k=1}^K \in \mathbb{R}_{++}^{2K}$  satisfies the constraint qualification if there is an  $x \in \mathbb{R}^{2K}$  such that  $\nabla \mathcal{G}(\{c_k, e_k\}_{k=1}^K) x < 0$ .

**Proposition 9.** Let  $T^*$  and  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  denote an optimal tax equilibrium with worker types indexed so that  $w_k^* = F_k(e_1^* \pi_1, \dots, e_K^* \pi_K)$  is non-decreasing in  $k$ . Assume that  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$  is interior (i.e. in  $\mathbb{R}_{++}^{2K}$ ) and that  $\mathcal{G}$  satisfies the constraint qualification at  $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$ , then optimal tax rates satisfy:

$$\frac{\tau_k^*}{1 - \tau_k^*} = \underbrace{\frac{1 - \Pi_k}{\pi_k} \left\{ \frac{\Delta w_{k+1}^*}{w_{k+1}^*} \Psi_{k,k+1}^* \mathcal{H}_{k,k+1}^* - \frac{\Delta w_k^*}{w_{k-1}^*} \Psi_{k,k-1}^* \mathcal{H}_{k,k-1}^* \right\}}_{\text{Mirrlees}} + \underbrace{\sum_{j=1}^{K-1} \mathcal{M}_{k,j}^* \phi_{k,j}^*}_{\text{Wage compression}}, \quad (\text{B.1})$$

where  $\Delta w_k^* := w_k^* - w_{k-1}^*$ ,  $\Psi_{k,k+1}^* := \frac{U_c(c_k^*, e_k^*)}{1 - \Pi_k} \frac{\eta_{k+1,k}^*}{\chi^*}$  and  $\Psi_{k,k-1}^* := \frac{U_c(c_k^*, e_k^*)}{1 - \Pi_k} \frac{\eta_{k-1,k}^*}{\chi^*}$  are normalized optimal multipliers on the  $(k+1, k)$ -th and  $(k-1, k)$ -th incentive constraints,

$$\mathcal{H}_{k,j}^* := -\frac{\Delta_e U_c(c_k^*, e_k^*; q_k^*/w_j^* - e_k^*)}{U_c(c_k^*, e_k^*)} e_k^* + \frac{\Delta_e U_e(c_k^*, e_k^*; q_k^*/w_j^* - e_k^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_j^*} e_k^* + 1,$$

with  $q_k^* = w_k^* e_k^*$ ,

$$\mathcal{M}_{k,j}^* := \frac{U_c(c_k^*, e_k^*)}{U_e(c_k^*, e_k^*) e_k^*} \left[ \frac{\eta_{j+1,j}^*}{\chi^*} U_e \left( c_j^*, \frac{q_j^*}{w_{j+1}^*} \right) \frac{q_j^*}{w_{j+1}^*} - \frac{\eta_{j,j+1}^*}{\chi^*} U_e \left( c_{j+1}^*, \frac{q_{j+1}^*}{w_j^*} \right) \frac{q_{j+1}^*}{w_j^*} \right] \frac{1}{\pi_k},$$

and  $\phi_{k,j}^* = \frac{e_k^*}{w_{j+1}^*/w_j^*} \frac{\partial w_{j+1}^*/w_j^*}{\partial e_k} (e_1^*, \dots, e_K^*)$ .

*Proof of Proposition 9.* By the preceding discussion if  $\{T^*, \{c_k^*, e_k^*, w_k^*\}_{k=1}^K\}$  is an optimal tax equilibrium, then  $\{c_k^*, e_k^*\}_{k=1}^K$  solves (MDP). Since  $\mathcal{G}$  satisfies the constraint qualification at  $\{c_k^*, e_k^*\}_{k=1}^K$  and  $\{c_k^*, e_k^*\}_{k=1}^K$  is interior to  $\mathbb{R}_{++}^{2K}$ , then  $\{c_k^*, e_k^*\}_{k=1}^K$  satisfies Karush-Kuhn-Tucker conditions with multipliers  $\chi^*$  and  $\eta_{k,j}^*$  on the resource and incentive constraints (and zero multipliers on the non-

---

(non-local)  $(k-1, k+1)$  and  $(k, k+2)$  incentive constraints bind. Thus, ordering of worker types with tied wages by income is necessary to ensure only local incentive constraints bind.

negativity conditions  $c_k^*, e_k^* \geq 0$ ). Also, since worker types are indexed so that shadow wages  $w_k^* = F_k(e_1^* \pi_1, \dots, e_K^* \pi_K)$  are non-decreasing in  $k$  and if wages of different types are tied so that incomes are non-decreasing in  $k$ , then by Lemma A, only local incentive constraints are potentially binding and, hence, only the  $\eta_{k,k-1}^*$  and  $\eta_{k,k+1}^*$  multipliers are potentially non-zero. The first order condition for  $e_k^*$  reduces to:

$$-U_e(c_k^*, e_k^*) = \frac{\lambda^* w_k^* \pi_k}{D_k^*},$$

where:  $D_k^* := g_k + \eta_{k,k-1}^* - \eta_{k+1,k}^* \frac{U_e(c_k^*, q_k^*/w_{k+1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k+1}^*} + \eta_{k,k+1}^* - \eta_{k-1,k}^* \frac{U_e(c_k^*, q_k^*/w_{k-1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k-1}^*} + \frac{\lambda^* \pi_k}{U_e(c_k^*, e_k^*)} \Phi_k^* + \frac{\lambda^* \pi_k}{U_e(c_k^*, e_k^*)} Y_k^*$  and

$$\Phi_k^* := \frac{U_c(c_k^*, e_k^*)}{\pi_k} \sum_{j=1}^{K-1} \frac{\eta_{j+1,j}^*}{\lambda^*} \frac{U_e(c_j^*, q_j^*/w_{j+1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_j^* e_j^*}{w_{j+1}^* e_k^*} \phi_{k,j}^*,$$

and

$$Y_k^* := \frac{U_c(c_k^*, e_k^*)}{\pi_k} \sum_{j=1}^{K-1} \frac{\eta_{j-1,j}^*}{\lambda^*} \frac{U_e(c_j^*, q_j^*/w_{j-1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_j^* e_j^*}{w_{j+1}^* e_k^*} \phi_{k,j-1}^*.$$

The first order condition for  $c_k^*$  reduces to:

$$U_c(c_k^*, e_k^*) = \frac{\lambda^* \pi_k}{g_k + \eta_{k,k-1}^* - \eta_{k+1,k}^* \frac{U_e(c_k^*, q_k^*/w_{k+1}^*)}{U_e(c_k^*, e_k^*)} + \eta_{k,k+1}^* - \eta_{k-1,k}^* \frac{U_e(c_k^*, q_k^*/w_{k-1}^*)}{U_e(c_k^*, e_k^*)}}.$$

Define the consumption-effort wedge:  $\frac{\tau_k^*}{1-\tau_k^*} = -\frac{w_k^* U_c(c_k^*, e_k^*)}{U_e(c_k^*, e_k^*)} - 1$ . Combining expressions gives:

$$\begin{aligned} \frac{\tau_k^*}{1-\tau_k^*} = \frac{U_c(c_k^*, e_k^*)}{\pi_k} \left\{ \frac{\eta_{k+1,k}^*}{\lambda^*} \left\{ \frac{U_c(c_k^*, q_k^*/w_{k+1}^*)}{U_e(c_k^*, e_k^*)} - \frac{U_e(c_k^*, q_k^*/w_{k+1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k+1}^*} \right\} \right. \\ \left. + \frac{\eta_{k-1,k}^*}{\lambda^*} \left\{ \frac{U_c(c_k^*, q_k^*/w_{k-1}^*)}{U_e(c_k^*, e_k^*)} - \frac{U_e(c_k^*, q_k^*/w_{k-1}^*)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k-1}^*} \right\} \right\} + \Phi_k^* + Y_k^*. \end{aligned}$$

The formulas in Proposition 9 then follow immediately from the definitions of  $\Psi_{k,k+1}^*$ ,  $\Psi_{k,k-1}^*$ ,  $\mathcal{H}_{k,j}^*$ ,  $\mathcal{M}_{k,j}^*$  and  $\phi_{k,j}^*$  after substitution into and rearrangement of the preceding expression.  $\square$

The terms on the right hand side of the optimal tax formula (B.1) are generalizations of the “Mirrlees” and “Wage Compression” terms obtained in the main text. These terms incorporate the impact of binding (local) upwards constraints as well as downwards constraints. In standard

models upwards constraints bind when it is optimal to pool agents with *distinct wages* at a *common consumption-effort allocation*. In the more general problem (MDP), they may also bind when it is optimal to pool distinct types with *distinct allocations* at a *common wage*. Rothschild and Chen (2014) provide an example in which such wage pooling occurs.<sup>2</sup> Our later assignment model micro-founds the production function  $F$ . In that setting, the induced production function does not feature wage pooling. This motivates us to consider situations in which the local upwards incentive constraints  $(k, k+1)$  are strictly non-binding at the optimum:<sup>3</sup>

$$U(c_k^*, e_k^*) > U(c_{k+1}^*, q_{k+1}^*/w_k^*). \quad (\text{NUIC})$$

In such cases, the optimal tax formula (B.1) reduces to that given in Proposition 2. The latter is obtained as a simple corollary of Proposition 9.

*Proof of Proposition 2.* The optimal tax formula (2.4) in Proposition 2 follows directly from that in Proposition 9 after setting all  $\eta_{k-1,k}^*$  equal to 0, using the modified definitions in Proposition 2 and expanding the recursion for  $\eta_{k+1,k}^*$  implied by the first order condition for  $c_{k+1}^*$ :

$$\frac{\eta_{k+1,k}^*}{\chi^*} = \left(1 - \frac{g_{k+1} U_c(c_{k+1}^*, e_{k+1}^*)}{\chi^* \pi_{k+1}}\right) \frac{\pi_{k+1}}{U_c(c_{k+1}^*, e_{k+1}^*)} + \frac{\eta_{k+2,k+1}^*}{\chi^*} \frac{U_c(c_{k+1}^*, q_{k+1}^*/w_{k+2}^*)}{U_c(c_{k+1}^*, e_{k+1}^*)},$$

with  $\eta_{K+1,K}^* = 0$ . □

## B.2 Proofs from Section 2.4

*Proof of Proposition 3.* Let  $\{T, l, \{c_k, e_k, \lambda_k\}_{k=1}^K\}$  and  $\{\omega_k\}_{k=1}^K$  denote a tax equilibrium at spending level  $G$ . Since workers of a given type select the highest possible wage, it follows that for each  $k$  there is a  $w_k < \infty$  such that for every  $v \in \text{Supp } \Lambda_k$ ,  $\omega(v) a_k(v) = w_k$  and for  $v \notin \text{Supp } \Lambda_k$ ,  $\omega(v) a_k(v) \leq w_k$ . Firm optimality implies that for almost every  $v \in [\underline{v}, \bar{v}]$ :  $\omega(v) = b(v) \left(\frac{Y}{l(v)}\right)^{\frac{1}{\varepsilon}}$ . If  $v \in [\underline{v}, \bar{v}] \setminus \bigcup_{k=1}^K \text{Supp}(\Lambda_k)$ , then in equilibrium  $l(v) = \sum_{k=1}^K \lambda_k(v) a_k(v) e_k = 0$ . Since  $\omega$  is finite, almost all tasks must be performed. Without loss of generality, we select versions of tax equilibria

<sup>2</sup>The effort of two worker types  $k$  and  $k'$  sharing a common wage may interact differently with that of a third worker type  $\hat{k}$ . Thus, it may be desirable to give  $k$  and  $k'$  distinct allocations (lying upon the same indifference curve).

<sup>3</sup>Standard pooling of consumption-effort allocations across types with distinct wages is still, in principle, possible. Inequality (NUIC) excludes this.

in which all tasks are performed. For all  $v$  and  $v'$  in  $\text{Supp } \Lambda_k$  with  $v > v'$ ,

$$1 = \frac{\omega(v)a_k(v)}{\omega(v')a_k(v')} = \frac{b(v) \left(\frac{Y}{l(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)}{b(v') \left(\frac{Y}{l(v')}\right)^{\frac{1}{\varepsilon}} a_k(v')} < \frac{b(v) \left(\frac{Y}{l(v)}\right)^{\frac{1}{\varepsilon}} a_{k+j}(v)}{b(v') \left(\frac{Y}{l(v')}\right)^{\frac{1}{\varepsilon}} a_{k+j}(v')} < \frac{\omega(v)a_{k+j}(v)}{\omega(v')a_{k+j}(v')}.$$

It follows that  $v' \notin \Lambda_{k+j}$  and so  $\sup \Lambda_k \leq \inf \Lambda_{k+j}$ . Since the supports  $\Lambda_k$  cover  $[\underline{v}, \bar{v}]$ , it follows that they partition  $[\underline{v}, \bar{v}]$  into sub-intervals  $[\tilde{v}_0, \tilde{v}_1], [\tilde{v}_1, \tilde{v}_2], \dots, [\tilde{v}_{K-1}, \tilde{v}_K]$ , with  $\tilde{v}_0 = \underline{v}$ ,  $\tilde{v}_K = \bar{v}$  and  $\text{cl } \Lambda_k = [\tilde{v}_{k-1}, \tilde{v}_k]$ . By assumption each  $\Lambda_k$  has a density  $\lambda_k$  (concentrated on  $[\tilde{v}_{k-1}, \tilde{v}_k]$ ). Since  $w_k = \omega(v)a_k(v)$ ,  $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$ , we have for all such  $v$ :

$$w_k = b(v) \left( \frac{Y}{a_k(v)e_k \lambda_k(v)} \right)^{\frac{1}{\varepsilon}} a_k(v). \quad (\text{B.2})$$

Hence, from the labor market clearing condition:

$$\pi_k = \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} \lambda_k(v) dv = \frac{Y}{w_k^\varepsilon e_k} \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv$$

And so:

$$w_k = B_k(\tilde{v}_{k-1}, \tilde{v}_k) \left( \frac{Y}{\pi_k e_k} \right)^{\frac{1}{\varepsilon}}, \quad (\text{B.3})$$

where  $B_k(\tilde{v}_{k-1}, \tilde{v}_k) := \left[ \int_{\underline{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv \right]^{\frac{1}{\varepsilon}}$ . Substituting (B.2) into (B.3) gives for  $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$ ,  $\lambda_k(v) = \frac{b(v)^\varepsilon a_k(v)^{\varepsilon-1}}{B_k(\tilde{v}_{k-1}, \tilde{v}_k)^\varepsilon} \pi_k$ . In addition, for  $v < \tilde{v}_{k-1}$  and  $v > \tilde{v}_k$ ,  $\lambda_k(v) = 0$ . Now for  $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$ ,  $w_{k+1} > \omega(v)a_{k+1}(v) = b(v) \left(\frac{Y}{l(v)}\right)^{\frac{1}{\varepsilon}} a_{k+1}(v)$  and  $w_k = \omega(v)a_k(v) = b(v) \left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)$ . Hence:  $\frac{w_{k+1}}{w_k} > \frac{a_{k+1}(v)}{a_k(v)}$ . Conversely, for  $v \in (\tilde{v}_k, \tilde{v}_{k+1})$ ,  $w_{k+1} = \omega(v)a_{k+1}(v) = b(v) \left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_{k+1}(v)$  and  $w_k > \omega(v)a_k(v) = b(v) \left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)$ . Consequently,  $\frac{w_{k+1}}{w_k} < \frac{a_{k+1}(v)}{a_k(v)}$ . Then, by continuity of  $a_k$  and  $a_{k+1}$ ,  $\frac{w_{k+1}}{w_k} = \frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)}$ . Combining the last equality with (B.3) gives the desired expression in the proposition.

Finally, given the effort allocation  $\{e_k\}_{k=1}^K$  consider assigning workers so as to maximize output, i.e. solving:

$$\max_{\{\lambda_k\}} \left[ \int_{\underline{v}}^{\bar{v}} b(v) \{ \lambda_k(v) e_k a_k(v) \}^{\frac{\varepsilon-1}{\varepsilon}} dv \right]^{\frac{\varepsilon}{\varepsilon-1}}.$$

subject to for each  $k$ ,  $\pi_k = \int_{\underline{v}}^{\bar{v}} \lambda_k(v) dv$ . This is a strictly concave maximization whose unique solu-

tion is determined by the first order conditions. Straightforward manipulation of these conditions establishes that the  $\lambda_k$  solved for above attains the solution to this problem.  $\square$

*Proof of Lemma 3.* Totally differentiating:  $\frac{a_{j+1}}{a_j}(\tilde{v}_j) = \frac{B_{j+1}(\tilde{v}_j, \tilde{v}_{j+1})}{B_j(\tilde{v}_{j-1}, \tilde{v}_j)} \left( \frac{e_j \pi_j}{e_{j+1} \pi_{j+1}} \right)^{\frac{1}{\varepsilon}}$ , with respect to  $\tilde{v}_{j-1}$  and  $\tilde{v}_j$  holding  $e_j/e_{j+1}$  fixed gives:

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}} = \frac{-\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} \frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j}}.$$

Let  $L_{j-1} = -\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} - \frac{\partial \log B_j}{\partial \log \tilde{v}_j} \frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}}$ . It follows that:

$$L_{j-1} = -\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} \left\{ 1 - \frac{\frac{\partial \log B_j}{\partial \log \tilde{v}_j}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} + L_j} \right\}.$$

Thus, if  $L_j > 0$ , then  $L_{j-1} > 0$ . For  $j = K-1$ ,  $\tilde{v}_{j+1} = \tilde{v}_K = \bar{v}$  and  $\frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j} = \frac{\partial \log \tilde{v}_K}{\partial \log \tilde{v}_{K-1}} = 0$ . Hence,  $L_{K-1} = -\frac{\partial \log B_K}{\partial \log \tilde{v}_{K-1}} > 0$ . It follows by induction that for all  $j \in \{k, \dots, K-2\}$ ,  $L_j > 0$  and, hence,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}} = \frac{-\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} + L_j} > 0.$$

Similarly, for all  $j \in \{1, \dots, k-1\}$ ,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}} = \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} \frac{\partial \log \tilde{v}_{j-1}}{\partial \log \tilde{v}_j}}.$$

Let  $M_{j+1} = \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} \frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}}$ . It follows that:

$$M_{j+1} = \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} \left\{ 1 - \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + M_j} \right\}.$$

Thus, if  $M_j > 0$ , then  $M_{j+1} > 0$ . For  $j = 1$ ,  $\tilde{v}_{j-1} = \tilde{v}_0 = \underline{v}$  and  $\frac{\partial \log \tilde{v}_{j-1}}{\partial \log \tilde{v}_j} = \frac{\partial \log \tilde{v}_0}{\partial \log \tilde{v}_1} = 0$ . Hence,

$M_1 = \frac{\partial \log B_1}{\partial \log \tilde{v}_1} > 0$ . It follows by induction that for all  $j \in \{1, \dots, k-1\}$ ,  $M_j > 0$  and, hence,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}} = \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}}}{\frac{\partial \log \frac{a_{j+1}}{a_j}}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + M_j} > 0.$$

Next, taking logs and totally differentiating  $\frac{a_{k+1}}{a_k}(\tilde{v}_k) = \frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{B_j(\tilde{v}_{k-1}, \tilde{v}_k)} \left( \frac{e_k \pi_k}{e_{k+1} \pi_{k+1}} \right)^{\frac{1}{\varepsilon}}$  with respect to  $\log e_k$  gives:

$$\frac{\partial \log \tilde{v}_k}{\partial \log e_k} = \frac{1}{\varepsilon} \left\{ \frac{1}{\frac{\partial \log \frac{a_{k+1}}{a_k}}{\partial \log \tilde{v}_k} + M_k + L_k} \right\} > 0.$$

Similarly, taking logs and totally differentiating  $\frac{a_k}{a_{k-1}}(\tilde{v}_{k-1}) = \frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{B_j(\tilde{v}_{k-2}, \tilde{v}_{k-1})} \left( \frac{e_{k-1} \pi_{k-1}}{e_k \pi_k} \right)^{\frac{1}{\varepsilon}}$  with respect to  $\log e_k$  gives:

$$\frac{\partial \log \tilde{v}_{k-1}}{\partial \log e_k} = -\frac{1}{\varepsilon} \left\{ \frac{1}{\frac{\partial \log \frac{a_k}{a_{k-1}}}{\partial \log \tilde{v}_{k-1}} + M_{k-1} + L_{k-1}} \right\} < 0.$$

The implications for the elasticities  $\phi_{k,j}$  described in the lemma then follow immediately from (2.17).  $\square$

Complete characterization of the sensitivity of task thresholds to the perturbation of a given talent's effort is provided in the next lemma.

**Lemma 14.** *The threshold sensitivities satisfy:*

$$\frac{\partial \log \tilde{v}_j}{\partial \log e_k} = (\delta_{j,k-1} - \delta_{j,k}) \frac{1}{\varepsilon},$$

where:

$$\delta_{j,k} = \begin{cases} (-1)^{j+k} \Pi_{l=j}^{k-1} \left( -\frac{\tilde{v}_l}{B_l} \frac{\partial B_l}{\partial \tilde{v}_l} \right) n_{j-1} m_{k+1} / n_{K-1} & 1 \leq j \leq k \leq K-1 \\ (-1)^{j+k} \Pi_{l=k}^{j-1} \left( \frac{\tilde{v}_l}{B_{l+1}} \frac{\partial B_{l+1}}{\partial \tilde{v}_l} \right) n_{k-1} m_{j+1} / n_{K-1} & K-1 \geq j > k \geq 1, \end{cases}$$

for  $j = 1, \dots, K-1$ ,  $\delta_{j,K} = \delta_{j,0} = 0$ , the  $n_i$  satisfy the recursion,  $i = 2, \dots, K-1$ ,

$$n_i = \left\{ \frac{\tilde{v}_i}{a_{i+1}/a_i} \frac{\partial(a_{i+1}/a_i)}{\partial \tilde{v}_i} - \frac{\tilde{v}_i}{B_{i+1}} \frac{\partial B_{i+1}}{\partial \tilde{v}_i} + \frac{\tilde{v}_i}{B_i} \frac{\partial B_i}{\partial \tilde{v}_i} \right\} n_{i-1} + \left( \frac{\tilde{v}_i}{B_i} \frac{\partial B_i}{\partial \tilde{v}_i} \right) \left( \frac{\tilde{v}_{i-1}}{B_i} \frac{\partial B_i}{\partial \tilde{v}_{i-1}} \right) n_{i-2}$$

with  $n_0 = 1$  and  $n_1 = \frac{\tilde{v}_1}{a_2/a_1} \frac{\partial(a_2/a_1)}{\partial \tilde{v}_1} - \frac{\tilde{v}_1}{B_2} \frac{\partial B_2}{\partial \tilde{v}_1} + \frac{\tilde{v}_1}{B_1} \frac{\partial B_1}{\partial \tilde{v}_1}$  and the  $m_i$  satisfy the recursion,  $i = K-2, \dots, 1$ ,

$$m_i = \left\{ \frac{\tilde{v}_i}{a_{i+1}/a_i} \frac{\partial(a_{i+1}/a_i)}{\partial \tilde{v}_i} - \frac{\tilde{v}_i}{B_{i+1}} \frac{\partial B_{i+1}}{\partial \tilde{v}_i} + \frac{\tilde{v}_i}{B_i} \frac{\partial B_i}{\partial \tilde{v}_i} \right\} m_{i+1} + \left( \frac{\tilde{v}_{i+1}}{B_{i+1}} \frac{\partial B_{i+1}}{\partial \tilde{v}_{i+1}} \right) \left( \frac{\tilde{v}_i}{B_{i+1}} \frac{\partial B_{i+1}}{\partial \tilde{v}_i} \right) m_{i+2}.$$

with  $m_{K-1} = \frac{\tilde{v}_{K-1}}{a_K/a_{K-1}} \frac{\partial(a_K/a_{K-1})}{\partial \tilde{v}_{K-1}} - \frac{\tilde{v}_{K-1}}{B_K} \frac{\partial B_K}{\partial \tilde{v}_{K-1}} + \frac{\tilde{v}_{K-1}}{B_{K-1}} \frac{\partial B_{K-1}}{\partial \tilde{v}_{K-1}}$  and  $m_K = 1$ .

*Proof of Lemma 14.* Given an (equilibrium) effort profile  $\{e_k\}_{k=1}^K$ , (equilibrium) production maximizing task thresholds  $\{\tilde{v}_k\}$  are determined by the conditions,  $k = 1, \dots, K-1$ ,

$$\frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{\{\pi_k e_k\}^{\frac{1}{\varepsilon}}} = \frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{\{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}} \frac{a_k(\tilde{v}_k)}{a_{k+1}(\tilde{v}_k)}.$$

with  $\tilde{v}_0 = \underline{v}$  and  $\tilde{v}_K = \bar{v}$ . Hence, there are  $K-1$  unknowns (and  $K-1$  equations). The threshold sensitivities may be computed by taking logs in the preceding equations and totally differentiating with respect to  $\log e_k$ . This leads to the equations:

$$\Gamma \Delta v_k = E_k,$$

where:

$$\Gamma := \begin{pmatrix} \alpha_1 \tilde{v}_1 - \frac{\tilde{v}_1}{B_2} \frac{\partial B_2}{\partial \tilde{v}_1} + \frac{\tilde{v}_1}{B_1} \frac{\partial B_1}{\partial \tilde{v}_1} & -\frac{\tilde{v}_2}{B_2} \frac{\partial B_2}{\partial \tilde{v}_2} & 0 & 0 & \dots & 0 \\ \frac{\tilde{v}_1}{B_2} \frac{\partial B_2}{\partial \tilde{v}_1} & \alpha_2 \tilde{v}_2 - \frac{\tilde{v}_2}{B_3} \frac{\partial B_3}{\partial \tilde{v}_2} + \frac{\tilde{v}_2}{B_2} \frac{\partial B_2}{\partial \tilde{v}_2} & -\frac{\tilde{v}_3}{B_3} \frac{\partial B_3}{\partial \tilde{v}_3} & 0 & \dots & 0 \\ 0 & \frac{\tilde{v}_2}{B_3} \frac{\partial B_3}{\partial \tilde{v}_2} & \alpha_3 \tilde{v}_3 - \frac{\tilde{v}_3}{B_4} \frac{\partial B_4}{\partial \tilde{v}_3} - \frac{\tilde{v}_3}{B_3} \frac{\partial B_3}{\partial \tilde{v}_3} & -\frac{\tilde{v}_4}{B_4} \frac{\partial B_4}{\partial \tilde{v}_4} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$\Delta v_k = (\frac{\partial \log v_1}{\partial \log e_k} \dots \frac{\partial \log v_k}{\partial \log e_k} \dots \frac{\partial \log v_{K-1}}{\partial \log e_k})'$  and  $E_k = (0 \dots -\frac{1}{\varepsilon} \frac{1}{\varepsilon} \dots 0)'$  with non-zero elements in the  $k-1$  and  $k$ -th rows. Thus,

$$\Delta v_k = \Gamma_k^{-1} E_k.$$

and in fact:

$$\frac{\partial \log v_j}{\partial \log e_k} = (\delta_{j,k-1} - \delta_{j,k}) \frac{1}{\varepsilon},$$

where  $\delta_{j,k}$  is the  $(j, k)$ -th element of  $\Gamma_k^{-1}$ . Since  $\Gamma_k$  is a tridiagonal matrix, explicit formulas for its inverse are available. Applying these formulas gives the expression in the text.  $\square$



### B.3 Continuous Talent-Continuous Task Model

In this appendix, we briefly describe the continuous talent (and continuous task) assignment model and its optimal control formulation. In our quantitative work, we treat the data as a discrete approximation to this model and solve it using the open-source numerical optimal control software GPOPS-II. Code is available on request. Workers are now distributed across an interval of talents  $k \in [\underline{k}, \bar{k}]$  according to a distribution function  $\Pi : [\underline{k}, \bar{k}] \rightarrow [0, 1]$  with strictly positive and continuously differentiable density  $\pi$ . As before there is a continuum of tasks ranked by complexity  $v \in [\underline{v}, \bar{v}]$ . The productivity of talent-task combinations is given by a function  $a : [\underline{k}, \bar{k}] \times [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_{++}$  satisfying the following assumption.

**Assumption 2.** (i)  $a$  is twice continuously differentiable on the interior of  $[\underline{k}, \bar{k}] \times [\underline{v}, \bar{v}]$  with first derivatives  $a_i$ ,  $i \in \{k, v\}$  and second derivatives  $a_{ij}$ ,  $i, j \in \{k, v\}$ . (ii) (strict absolute advantage)  $a_k > 0$ , (iii) (strict comparative advantage, log supermodularity)  $\frac{\partial^2 \log a}{\partial k \partial v} > 0$ .

Otherwise technologies and preferences are as in the main text. An allocation is a triple of measurable functions  $c : [\underline{k}, \bar{k}] \rightarrow \mathbb{R}_+$ ,  $v : [\underline{k}, \bar{k}] \rightarrow [\underline{v}, \bar{v}]$  and  $e : [\underline{k}, \bar{k}] \rightarrow \mathbb{R}_+$  describing the consumption, task and effort assignments of each talent type.<sup>4</sup> As before, task output is linear in labor input. The task output density  $y : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$  satisfies for all  $k$ ,

$$\int_{\underline{v}}^{v(k)} y(v) dv = \int_{\underline{k}}^k a[k', v(k')] e(k') \pi(k') dk'. \quad (\text{B.4})$$

If  $v$  is differentiable with derivative  $v_k$ , then (B.4) can be re-expressed as, for all  $k$ :

$$y(v(k)) = a[k, v(k)] e(k) \frac{\pi(k)}{v_k(k)}, \quad (\text{B.5})$$

Heuristically, the numerator is total output of type  $k$ , while the denominator gives the tasks over which the type  $k$  workers are "spread". The shadow wage is given by:

$$w[k, v] = b(v) \left( \frac{y(v)}{Y} \right)^{-\frac{1}{\varepsilon}} a[k, v].$$

---

<sup>4</sup>The implicit assumption that all talents are assigned to a specific consumption, task and effort is without loss of generality. It may be shown, along the lines of Proposition 3, that assignment of talents to tasks is strictly increasing in talent given strict comparative advantage.

We restrict planners and policymakers to smooth allocations and mechanisms. This permits the application of optimal control techniques.

**OPTIMAL CONTROL FORMULATION OF GOVERNMENT'S PROBLEM** We formulate the government's problem as a mechanism design problem and recover optimal taxes from this. Mechanisms are analogous to those considered previously. Each worker reports its talent  $k$  and, conditional on this, is assigned a consumption  $c$ , task  $v$  and effort  $e$ . The combination of mechanism and truthfully reported talent imply utility and normalized shadow wage and income levels for each type:

$$\begin{aligned}\psi(k) &= U(c(k), e(k)) \\ \varphi(k) &= w[k, v(k)] / Y^{\frac{1}{\varepsilon}} \\ \rho(k) &= \varphi(k) e(k).\end{aligned}$$

In addition, let  $\omega(k, v) = w[k, v] / Y^{\varepsilon}$ . A worker claiming to be type  $k'$  must reproduce the observable income level  $\rho(k')$ . Incentive-compatibility thus requires for all  $k, k'$  and  $v'$ :

$$U(c(k), e(k)) \geq U\left(c(k'), \frac{\rho(k')}{\omega(k, v')}\right). \quad (\text{B.6})$$

Let  $\mathcal{U} = \{(u, e) \in \mathbb{R} \times [0, \bar{e}] : u = U(c, e) \text{ for some } c \in \mathbb{R}_+\}$  and let  $C : \mathcal{U} \rightarrow \mathbb{R}_+$  be defined according to  $u = U(C[u, e], e)$ . The next proposition gives simpler necessary and sufficient conditions for incentive-compatibility.

**Proposition 10.** *Let  $(v, e, c)$  be a smooth mechanism that induces a smooth task output function  $y$ . The mechanism is incentive-compatible if and only if: (i) (Monotonicity)  $v_k \geq 0$  and  $\rho_k \geq 0$  hold and (ii) (Envelope) the envelope conditions for utility and shadow wages hold:*

$$\psi_k(k) = -U_e(C[\psi(k), e(k)], e(k)) e(k) \frac{a_k[k, v(k)]}{a[k, v(k)]} \quad (\text{B.7})$$

$$\varphi_k(k) = \varphi(k) \frac{a_k[k, v(k)]}{a[k, v(k)]}. \quad (\text{B.8})$$

*Proof. (Necessity).* Let  $(v, e, c)$  be a smooth incentive-compatible mechanism. Incentive compatibility implies that  $w[k, v(k)] \geq w[k, v(k')]$  and  $w[k', v(k')] \geq w[k', v(k)]$ . Since  $w[k, v] = b(v) \left(\frac{y(v)}{Y}\right)^{-\frac{1}{\varepsilon}}$

$a[k, v]$  and  $a$  is strictly log supermodular, it follows that if  $k > k'$ , then  $v(k') > v(k)$ . Hence,  $v$  is increasing. To verify (B.8), we apply (envelope) Theorem 4.3 of [Bonnans and Shapiro \(2000\)](#) to:  $\max_{[v, \bar{v}]} U \left( c(k'), \frac{\rho(k')}{\omega(k, v')} \right)$ . This requires  $U$  to be continuously differentiable,  $\omega(\cdot, v)$  to be continuously differentiable and  $\omega(k, \cdot)$  to be continuous. The first two properties hold by assumption (and the definition of  $\omega$  and  $w$ , see B.3), the latter holds if  $y$  is continuous. Suppose that  $y$  is discontinuous at  $v$  and, without loss of generality assume  $y(v) > y(v_n)$  for some sequence  $v_n \rightarrow v$ . Let  $k_n = v^{-1}(v_n)$ , then for  $n$  large enough,  $w[k_n, v_n] < w[k_n, v]$ , which is a contradiction. Then Theorem 4.3 and Remark 4.14, p.273-4 in [Bonnans and Shapiro \(2000\)](#) and the definition of  $w$  imply that the function  $\varphi$ ,  $\varphi(k) = \max_{v \in [v, \bar{v}]} w[k, v]$ , is differentiable with  $\varphi_k = \varphi \frac{a_k}{a} > 0$ .

Let  $\rho(k) = \varphi(k)e(k)$  and:

$$Y[k, k'] = U \left( c(k'), \frac{\rho(k')}{\varphi(k)} \right).$$

Incentive-compatibility requires that:  $Y[k, k] \geq Y[k, k']$  and  $Y[k', k'] \geq Y[k', k]$ . Hence,  $U \left( c(k), \frac{\rho(k)}{\varphi(k)} \right) - U \left( c(k'), \frac{\rho(k')}{\varphi(k)} \right) \geq U \left( c(k), \frac{\rho(k)}{\varphi(k')} \right) - U \left( c(k'), \frac{\rho(k')}{\varphi(k')} \right)$ . The assumed Spence-Mirrlees condition and the increasingness of  $\varphi$ , then imply that  $\rho$  and  $c$  are increasing also. Additionally, since  $(v, e, c)$  is continuous by assumption and  $w$  is continuous, Theorem 4.3 in [Bonnans and Shapiro \(2000\)](#) can again be applied to show that:  $\psi(k) = \max_{k' \in [k, \bar{k}]} Y(k, k')$  is differentiable with:

$$\psi_k(k) = -U_e \left( C[\psi(k), e(k)], e(k) \right) e(k) \frac{\varphi_k(k)}{\varphi(k)} = -U_e \left( C[\psi(k), e(k)], e(k) \right) e(k) \frac{a_k[k, v(k)]}{a[k, v(k)]}.$$

*Sufficiency.* Let  $(v, e, c)$  be a smooth mechanism satisfying the conditions in the proposition. The definition of  $\varphi$ , the envelope condition for wages (B.8) and the smoothness of  $v$  imply the first order condition:  $w_v[k, v(k)]v_k = 0$ . The smoothness of the various functions also implies that  $w_v$  exists and is given by:

$$w_v[k, v] = \left\{ \frac{b_v(v)}{b(v)} - \frac{1}{\varepsilon} \frac{y_v(v)}{y(v)} + \frac{a_v[k, v]}{a[k, v']} \right\} w[k, v] \quad (\text{B.9})$$

An worker's optimization over  $v$  and  $k'$  is separable: regardless of the report choice of  $k'$ , it is optimal for the worker to select a task  $v$  that maximizes its wage  $w[k, v]$ . Let  $k^*$  denote a non-decreasing measurable selection from  $v^{-1}$ . Then, using (B.9), the first order condition  $w_v[k, v(k)]v_k =$

0 and log supermodularity, for  $\hat{v} > v(k)$ ,

$$\begin{aligned} w[k, v] - w[k, v(k)] &= \int_{v(k)}^v w_v[k, v'] dv' \\ &= \int_{v(k)}^{\hat{v}} \left\{ \frac{b_v(v')}{b(v')} - \frac{1}{\varepsilon} \frac{y_v(v')}{y(v')} + \frac{a_v[k, v']}{a[k, v']} \right\} w[k, v'] dv' \\ &= \int_{v(k)}^{\hat{v}} \left\{ -\frac{a_v[k^*(v'), v']}{a[k^*(v'), v']} + \frac{a_v[k, v']}{a[k, v']} \right\} w[k, v'] dv' < 0. \end{aligned}$$

and similarly for  $\hat{v} < v(k)$ . Consequently, the mechanism induces a  $k$ -worker to choose the task assignment  $v(k)$ .

Let  $k_2 > k_1$ , then by the envelope condition for wages, for  $k' \in [k_1, k_2]$ ,  $\varphi(k') = w[k', v(k')] \geq w[k_1, v(k_1)] = \varphi(k_1)$ . Combined with the monotonicity and concavity of  $U$ , this implies  $-U_e(c(k'), e(k')) e(k') + U_e\left(c(k'), \frac{\varphi(k')}{\varphi(k_1)} e(k')\right) \frac{\varphi(k')}{\varphi(k_1)} e(k') < 0$ . The envelope condition for reports and the smoothness of the mechanisms imply:

$$Y_{\hat{k}}[k, k] \hat{k}_k = \left\{ U_c(c(k), e(k)) c_k(k) + U_e(c(k), e(k)) e(k) \frac{\rho_k(k)}{\rho(k)} \right\} \hat{k}_k = 0.$$

The definitions of  $Y$  and  $\rho$  and the preceding discussion then imply:

$$\begin{aligned} Y[k_1, k_2] - Y[k_1, k_1] &= \int_{k_1}^{k_2} Y_{\hat{k}}[k_1, k'] dk' \\ &= \int_{k_1}^{k_2} \left\{ U_c(c(k'), e(k')) c_k(k') + U_e\left(c(k'), \frac{\rho(k')}{\varphi(k_1)}\right) \frac{\rho_k(k')}{\varphi(k_1)} \right\} dk' \\ &= \int_{k_1}^{k_2} \left\{ -U_e(c(k'), e(k')) e(k') + U_e\left(c(k'), \frac{\rho(k')}{\varphi(k_1)}\right) \frac{w[k', v(k')]}{w[k_1, v(k_1)]} e(k') \right\} \frac{\rho_k(k)}{\rho(k)} dk' \leq 0. \end{aligned}$$

A similar inequality obtains for  $k_1 > k_2$  and so the mechanism induces a  $k$ -worker to make a truthful report  $k$ .  $\square$

It is convenient to define:

$$\zeta(k) := \int_{\underline{k}}^k \left( \frac{\pi(k') a[k', v(k')] e(k')}{v_k(k')} \right)^{\frac{\varepsilon-1}{\varepsilon}} b(v(k')) v_k(k') dk' \quad (\text{B.10})$$

and

$$\zeta(k) = \int_{\underline{k}}^k C[\psi(k'), e(k')] \pi(k') dk'. \quad (\text{B.11})$$

Together  $\psi$ ,  $\varphi$ ,  $\xi$  and  $\zeta$  along with  $\nu$  form a set of state variables for the optimal control formulation of the planning problem with private information. The envelope conditions (B.7) and (B.8) supply laws of motion for  $\psi$  and  $\varphi$ . Equations (B.12) and (B.13) give laws of motion for  $\xi$  and  $\zeta$ :

$$\xi_k(k) = e(k)\varphi(k)\pi(k) \quad (\text{B.12})$$

$$\zeta_k(k) = C[\psi(k), e(k)]\pi(k). \quad (\text{B.13})$$

Finally, the definition of  $\varphi(k)$  implies a law of motion for  $\nu$ :

$$\nu_k(k) = \left( \frac{\varphi(k)}{b(\nu(k))a[k, \nu(k)]} \right)^\varepsilon \pi(k)a[k, \nu(k)]e(k), \quad (\text{B.14})$$

The monotonicity conditions on mechanisms needed to ensure incentive-compatibility are omitted and checked ex post. The effort function  $e$  is the control. The government's problem becomes:

$$\max_{\psi, \varphi, \xi, \zeta, \nu, e} \int_{\underline{k}}^{\bar{k}} \psi(k)g(k)dk \quad (\text{B.15})$$

subject to the laws of motion (B.7), (B.8) and (B.12) to (B.14) and the boundary constraints:

$$\begin{aligned} \zeta(\bar{k}) &\leq \xi(\bar{k})^{\frac{\varepsilon}{\varepsilon-1}} \\ 0 &= \xi(\underline{k}) \quad 0 = \zeta(\underline{k}) \quad \underline{\nu} = \nu(\underline{k}) \quad \nu(\bar{k}) = \bar{\nu}. \end{aligned}$$

In this problem there is one control ( $e$ ) and five states ( $\psi, \varphi, \xi, \zeta, \nu$ ). Routine manipulation of the first order and co-state equations yields the following expression for the optimal effort-consumption wedge:

$$\begin{aligned} -w[k, \nu^*(k)] \frac{U_c^*(k)}{U_e^*(k)} - 1 = & \underbrace{\mathcal{H}^*(k) \frac{1 - \Pi(k)}{\pi(k)} \frac{\varphi_k^*(k)}{\varphi^*(k)} \int_k^\infty \left( 1 - \frac{g(t)U_c^*(t)}{p^{\zeta^*} \pi(t)} \right) \frac{U_c^*(k)}{U_c^*(t)} \mathcal{N}^*(k, t) \frac{\pi(t)}{1 - \Pi(k)} dt}_{\text{Mirrlees}} \\ & - \underbrace{\mathcal{I}^*(k) \left[ \frac{p_k^{\varphi^*}(k)}{p^{\varphi^*}(k)} + \frac{a_k[k, \nu^*(k)]}{a[k, \nu^*(k)]} \right] \frac{p^{\varphi^*}(k)}{p^{\zeta^*} \pi(k)} \left( -\frac{U_c^*(k)}{U_e^*(k)} \right)}_{\text{Wage Compression}}, \end{aligned} \quad (\text{B.16})$$

where  $U_x^*(k) := U_x(c^*(k), e^*(k))$   $x \in \{c, e\}$  and similarly for  $U_{xz}^*(k)$ ,  $\mathcal{I}^*(k) := -\frac{1}{\varepsilon} \frac{w[k, \nu^*(k)]}{e^*(k)}$  is the elasticity of the  $k$ -talent wage with respect to effort holding the task allocation fixed,  $\mathcal{H}^* := \{-\frac{U_{ee}^*}{U_e^*} + \frac{U_{ee}^*}{U_e^*}\}e^* + 1$  is  $(1 + \mathcal{E}_u)/\mathcal{E}_c$ , where  $\mathcal{E}_u$  and  $\mathcal{E}_c$  are, respectively, the uncompensated and compensated labor supply elasticities,  $\mathcal{N}^*(k, t) = \exp\left\{-\int_k^t \frac{e^*(s)U_{ce}^*(s)}{\varphi^*(s)U_c^*(s)}\right\}$ ,  $p^{\zeta^*} = E\left[\int_k^{\bar{k}} \frac{1}{U_c^*(t'')} \pi(t'') dt''\right]^{-1}$  is the optimal shadow resource multiplier and  $p^{\varphi^*}$  is the optimal co-state on the shadow wage  $\varphi$ . The Mirrlees and wage compression components are labeled.

## B.4 Empirical Implementation

In this appendix, we discuss details of the empirical implementation. We describe the data set used and illustrate robustness of our estimated parameters to alternative sample selection criteria. We discuss the issues raised by dropping the Cobb-Douglas production assumption used in the main text and show how our results are modified by alternative values for the elasticity of final output with respect to occupational output. We show that our empirical results are robust to reordering occupations in different decades according to the average wage within the decade and provide supporting evidence for our functional form restrictions on  $a$ .

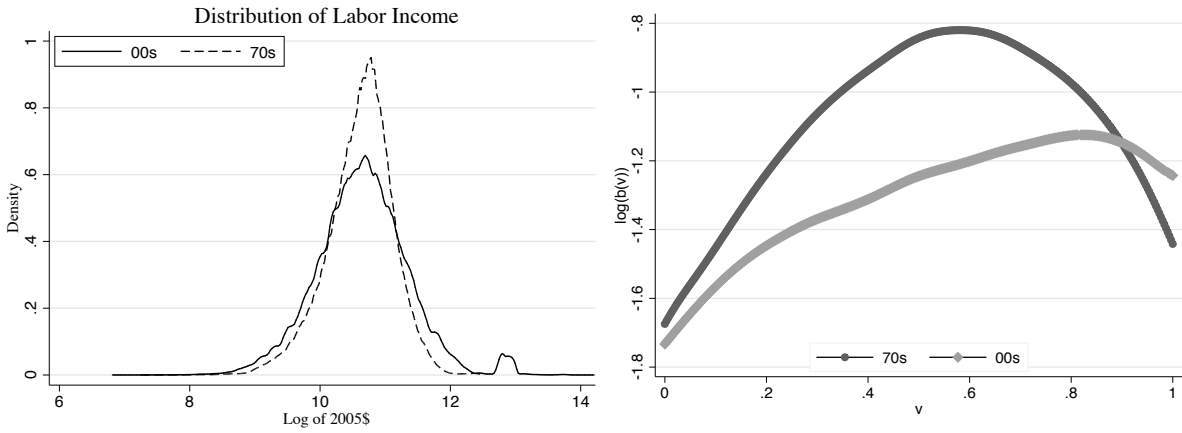
### B.4.1 Data Set and Sample Selection

Our main data source is the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics. We focus on the March release of the survey.<sup>5</sup> Data is available continuously from 1968 to 2012. On average each year of data contains about 150,000 observations, from 2001 the sample size has increased to approximately 200,000. The CPS contains detailed information on the demographic and work characteristics of each individual. For additional details on the CPS refer to [Heathcote, Perri, and Violante \(2010a\)](#) and [Acemoglu and Autor \(2011\)](#). The CPS data includes a self-reported estimate of hours worked from 1976 onwards. This question as well as questions on income are for the previous calendar year. Hence our sample covers the years 1975 to 2011 (interviews from 1976 to 2012). In the body of the paper we group observations in two groups. We call “the 70s” observations relating to years 1975-1979 (i.e inter-

<sup>5</sup>Data is taken from: Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

viewed in years 1976-1980), we call the “00s” observation relating to years 2000-2011 (interviews in 2001-2012).

The model analyzed is highly stylized. In order to make data and model compatible (and to reduce the likelihood of measurement error) we further restrict our sample. We drop individuals for whom income, age, sex, education, sector, occupation is not reported. We consider individuals of working age, i.e. between the ages of 25 and 65. We drop individuals with no formal education and the unemployed. Following [Heathcote et al. \(2010a\)](#), we also drop underemployed individuals: those working less than 250 hours per year or earning less than \$100 per year (dropping an additional 196,684 observations). Our final sample comprises of 2,039,123 individual/year observations. All variables are weighted with the provided weights and dollar denominated variables are deflated using CPI to 2005 dollars.<sup>6</sup> In Figure B.1a we display the evolution of the distribution of log labor income between the “70s” and the “00s”. The main feature that emerges is the widening of the distribution in the “00s” relative to the “70s”.



(a) Distribution of log-labor income.

(b) Values of  $\log(b(v))$  over time.

Figure B.1: Estimates on entire sample.

We briefly explore the impact of our sample selection on the estimated values of  $b$ . Figure B.1b shows  $b$  estimates for both decades obtained from the CPS sample before applying our sample selection (but after removal of individuals with missing information or an unclassifiable occupation). As can be seen polarization is still apparent. However for low  $v$  occupations we observe

<sup>6</sup>CPI for all urban consumers, all goods.

little change between the two decades. Note that a similar result would appear using the sample selection of [Acemoglu and Autor \(2011\)](#). This is because the authors only remove individuals who worked less than one week in the previous year or are less than 16 years of age.

#### B.4.2 Beyond Cobb-Douglas

In this section we extend our empirical strategy beyond the Cobb-Douglas assumption ( $\varepsilon = 1$ ) adopted in the main body. In this direction, it is important to recognize that the estimation of  $a$  is independent of  $\varepsilon$  and  $b(\cdot)$ . Estimation of  $b(\cdot)$  does, however, depend on  $\varepsilon$  and, outside of the Cobb-Douglas case,  $a$  as well. The firm's first order condition is:

$$\omega(k(v), v) = b(v) A^{\frac{\varepsilon-1}{\varepsilon}} \left( \frac{Y}{y(v)} \right)^{\frac{1}{\varepsilon}} a(k(v), v). \quad (\text{B.17})$$

Let  $b^*(v)$  denote the share of output paid to occupation  $v$  and, hence, the estimate of  $b(v)$  in the Cobb-Douglas case. Then from (B.17):

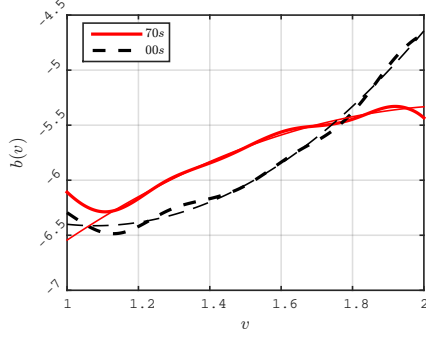
$$b(v) = A^{\frac{\varepsilon-1}{\varepsilon}} \left( \frac{\omega(k(v), v)}{a(k(v), v)} \right)^{\frac{\varepsilon-1}{\varepsilon}} b^*(v)^{\frac{1}{\varepsilon}}. \quad (\text{B.18})$$

Thus, to determine  $b$  outside of the Cobb-Douglas case values for  $a(\cdot)$  and  $\varepsilon$  are needed. Given values for these (B.18) determines  $b(\cdot)$  up to the constant  $A$ . The latter is pinned down by the restriction:  $\int_0^1 b(v) dv = 1$ .

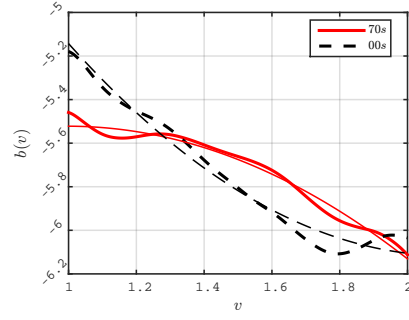
It is well known that the elasticity of substitution between goods and factor augmenting technical progress cannot be separately identified from data on outputs, inputs and marginal products - an observation that goes back to [McFadden et al. \(1978\)](#). The same logic implies that  $\varepsilon$  and  $b$  are not separately identified from this data. A typical response is to restrict the elasticity of substitution or the bias of factor augmenting technical change. In the main text, we proceed similarly by allowing the  $b$  parameter to be arbitrary and the production function to be Cobb-Douglas in occupational output. To assess the implications of this identifying assumption, we perform sensitivity analysis with respect to  $\varepsilon$ : we consider a range of values for  $\varepsilon$  and then re-compute  $b$ 's (and taxes) for each value.

Figure B.2 displays the  $b$  function for  $\varepsilon = 0.8$  and  $\varepsilon = 1.3$  (thicker lines show estimated  $b$ 's,





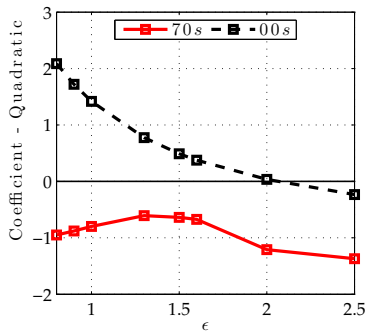
(a)  $\varepsilon = 0.8$ .



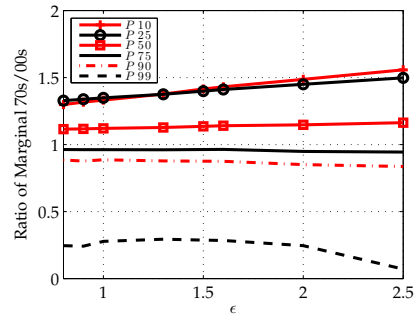
(b)  $\varepsilon = 1.3$ .

Figure B.2: Estimates of  $b(v)$  for different  $\varepsilon$  values.

thinner lines quadratic approximations to these estimates). The slope of the  $b$  function estimate is significantly impacted by variations in  $\varepsilon$ . However, polarization remains a consistent feature: between the 1970s and the 2000s, the  $b$  function rose in high  $v$  occupations, fell in mid ones and rose or only very marginally fell in low ones. We confirm and re-express this observation by taking a quadratic approximation to the  $b$  function (overlaid in Figure B.2 with thinner lines) for a variety of values of  $\varepsilon$  and plotting the quadratic coefficient across decade and  $\varepsilon$  value in Figure B.3a. The figure shows that over all the  $\varepsilon$  values considered the quadratic coefficient increases between the 1970s and the 2000s and over most it changes sign (as in the benchmark environment considered in the body of the paper).



(a) Values of quadratic coefficient for approximation of  $b(\cdot)$  for different values of  $\varepsilon$ .



(b) Ratio of marginal tax rates 70s over 00s for various percentile of income and over different values of  $\varepsilon$ .

Figure B.3: Impact of changes in  $\varepsilon$  on  $b$  estimates and optimal taxes.

We recompute the optimal tax functions for different values of  $\varepsilon$ . A summary of the impact of these values on optimal taxes is provided in Figure B.3b. Key patterns found in our benchmark

case in the main text re-emerge. In particular, optimal marginal tax rates fall at low to mid incomes between the 1970s and 2000s and rise at higher ones with this effect becoming more pronounced as  $\varepsilon$  rises. Roughly speaking as  $\varepsilon$  rises occupational outputs and, hence, workers become more substitutable. This diminishes the government's ability to influence and, hence, redistribute via relative wages. Consequently, the "wage compression" force is weakened relative to the "Mirrleesian": the government is less motivated to compress wage differentials by moderating marginal tax reductions at the bottom and marginal tax increases at the top.

### B.4.3 Occupational Ordering

In the main body of the paper, we restrict attention to 302 occupations present in both 1970s and 2000s data and order them using wage information from the 1970s. In doing so we follow the approach of [Acemoglu and Autor \(2011\)](#). This approach supposes that the complexity ordering of occupations is time invariant (and is captured by the 1970's wage ordering). To the extent that the occupational wage ordering changes and these changes reflect changes to the relative complexity of occupations estimates of the  $a$  and  $b$  functions are modified. Figure B.4 provides a scatterplot

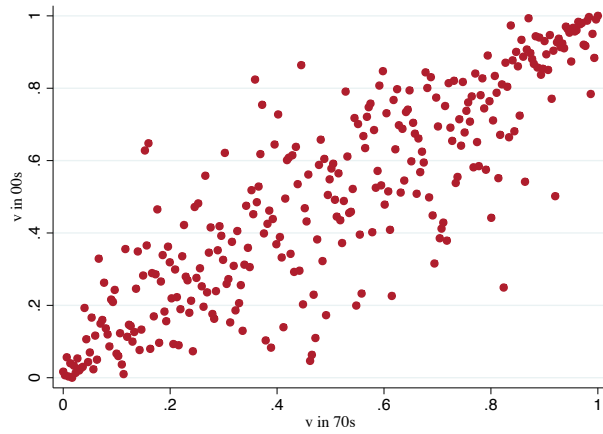


Figure B.4: Rank order over decade.

of occupations by their average wage rankings in the 1970s and 2000s. The plot shows that while these rankings are not time invariant, they do exhibit stability especially at the bottom and the top. The overall correlation of these rankings over time is 0.8623.

Re-estimating the parameters of the  $a$  function for the 2000s using the 2000s wage ordering gives values of:  $\alpha_1 = 0.64$  and  $\alpha_2 = 2.90$  (compared to our previous estimates of  $\alpha_1 = 0.42$  and  $\alpha_2 =$

3.01.) Thus, the critical comparative advantage parameter  $\alpha_2$  is only moderately changed under this alternative ordering and the overall pattern of increasing competitive advantage is preserved. New estimates of the  $b$  function using the 2000s ordering are also only moderately changed and, in particular, they preserve the increase in weights on complex (high wage) occupations found previously.

We recompute optimal taxes using the estimates of  $a$  and  $b$  function parameters after reordering and recoding. Although marginal taxes for the 2000's are slightly higher under this alternative parameterization, the overall impact of technical change upon them is little altered. As before, these taxes fall on low (but not the lowest) and rise on high (but not the highest) incomes. The numerical results are available on request.

#### **B.4.4 The demand for complex skills across occupations**

In the main text we assume a functional form for  $a$  that attributes the steepening of the profile of average wages across ranked occupations to increases in the comparative advantage of high talents in more complex tasks. However, an alternative scenario is also possible. Under this alternative, highly talented workers accumulated large stocks of general skills between the 1970s and the 2000s. This minority concentrates in high wage occupations where they have a slight comparative advantage, but talent-complexity comparative advantage has not greatly increased. Our account and this second one are difficult to distinguish using CPS data which, of course, does not give (imputed) wages for occupations other than those chosen by a worker. Greater acquisition of general skills by top talents could, in principal, be detected by adding non-linear terms in  $k$  (e.g.  $k^2$ ) to our empirical specification of the  $\log a$  function. This, however, is problematic as these terms are highly collinear with the comparative advantage term  $k\alpha$  in this function. Consequently, we look outside of the CPS for evidence of increasing talent-complexity comparative advantage. In particular, we use the O\*NET database to assess whether the intensity of complex skills and abilities has increased in high wage occupations relative to low. We treat such evidence as suggestive of increasing comparative advantage: increasing returns to talent in complex occupations.

Our empirical approach is similar to that in [Autor et al. \(2003\)](#), though our focus is distinct.<sup>7</sup>

---

<sup>7</sup>This paper documents the mix of routine and non routine tasks performed across occupations and its evolution. It uses the predecessor of the O\*NET database, the Dictionary of Occupational Titles.

We pair two editions of the O\*NET database with the CPS, merging by occupation. For each of the Census-defined occupations we record the level and the importance of each of the 52 abilities and 35 skills reported by O\*NET. This enables us to create two snapshots of the ability and skill requirement of occupations across time. We use the first (beta) release of the O\*NET database from 1998 and the latest available release (version number 19.0) from July 2014.<sup>8</sup> We multiply the two reported dimensions of skill/ability within an occupation (importance and level) to create a single skill/ability “intensity” index. The numerical range over which these dimensions are measured has not changed across the editions of O\*NET. However, the meaning associated with each score has. To overcome this limitation and allow for a consistent time comparison we follow [Autor et al. \(2003\)](#) and look at the percentile-rank of each occupation by each skill/ability index. In Figure B.5 we display the (smoothed) change of the intensity index across occupations (ranked as in the paper by  $v$ ) for three distinct skill/abilities: *complex problem solving*, *deductive reasoning* and *mathematical reasoning*. The first of these is the skill and the second the ability that most correlate with the

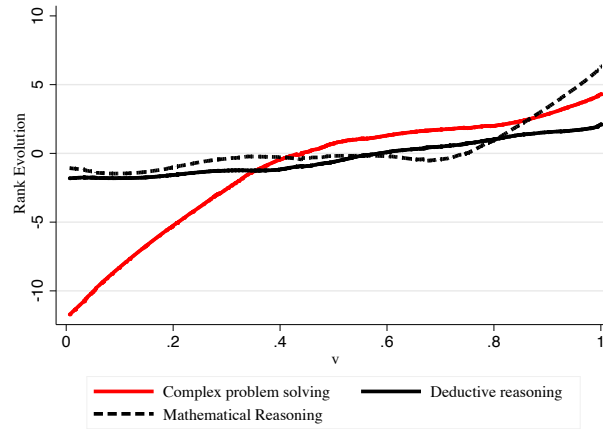


Figure B.5: Rank evolution of selected skills and abilities intensity across occupations. LOWESS smoothed.

task rank  $v$ . Thus, these intensity indices provide measures of task complexity that are consistent with the model. The third index (*mathematical reasoning*) is an ability that is commonly used to describe the complexity of an occupation. From the figure, each of these intensity index changes is negative for lower ranked occupations ( $v \leq 0.5$ ) and positive for higher ranked ones ( $v \geq 0.8$ ). The former is especially marked for complex problems solving and the latter for mathematical

<sup>8</sup>Admittedly the time span covered by these two datasets is much shorter than the one covered by our two benchmark time periods in the body of the paper. However we conjecture that the patterns uncovered in our comparison of these two dataset would be amplified with a longer timespan.

reasoning. Thus, consistent with our specification of  $a$ , skills and abilities that are particularly associated with occupational complexity have grown only in more complex occupations.

Table B.1 displays coefficients from regressions of changes in various skill/ability intensity indices on  $v$ . A positive estimate denotes a relative increase in the intensity of a skill or ability in more complex occupations. All of the point estimates in Table B.1 are positive. The majority of

Skill/Ability	Coefficient
<i>Complex problem solving</i>	12.51*** (2.34)
<i>Critical Thinking</i>	3.51** (1.91)
<i>Deductive reasoning</i>	4.70*** (2.15)
<i>Inductive reasoning</i>	9.10*** (2.14)
<i>Mathematical Reasoning</i>	4.02* (2.52)

Table B.1: Slope of the change of skill/ability intensity index over occupation

Slope of the change of skill/ability intensity index over occupation. Standard Errors in Parenthesis. \*\*\*= 95% confidence, \*\*= 90% confidence, \*= 85% confidence.

them are so with a high degree of confidence.

We conclude this section by taking a broader look at all 87 skills and abilities. To do this we perform a principal component analysis of the change in intensity indices across occupations. In Figure B.6a we display the first two principal components. The second principal component displays an increasing pattern of variation similar to the profiles plotted in Figure B.5. In Figure B.6b we display the loading factors. We label each skill/ability as being either associated with a high degree of complexity (the ones associated with information processing, problem solving, analytical thinking, managerial abilities) or not (mostly physical and interpersonal abilities).<sup>9</sup> We see that high complexity skill/abilities (in red in the graph) are on average associated with high loading on the second principal component and in some instances with a low loading on the first principal component. This implies that overall the high complexity skill/abilities display a positive sloping profile over the space of occupations. The opposite is true for low complexity skill/abilities.

<sup>9</sup>Detailed listing available is upon request.

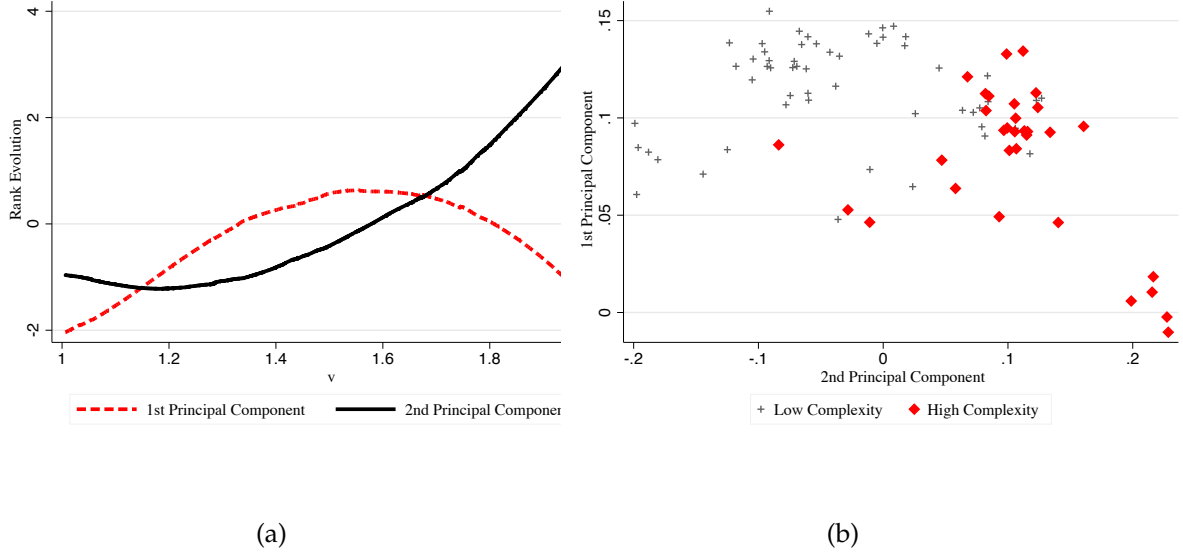


Figure B.6: Principal component analysis of the rank evolution of skill and ability intensity across occupations.

Overall, we view data on the evolution of the skill and ability composition of tasks as broadly consistent with increasing talent-task complexity comparative advantage and the functional form restrictions that we place on the  $a$  function in our analysis.

### B.4.5 Alternative Productivity Function

As discussed in Subsection 2.5.3, in our benchmark estimation the increase in the growth rate for log wages at higher talents leads to the identification of a growing comparative advantage over time. In this section we explore an alternative formulation of the productivity function aimed at fitting more closely the high and increasing growth rate of wages for high talents. Specifically, we set  $\frac{\partial a}{\partial k}(k, v) = \alpha_3 \cdot v^2$ . Proceeding as in Subsection 2.5.3 we find a value of  $\alpha_3 = 0.79$  (0.03) for the 1970's and a value of  $\alpha_3 = 0.91$  (0.04) for the 2000's.<sup>10</sup> As in our benchmark, there is an increase in the degree of comparative advantage over time. Given the quadratic nature of the productivity function the change in overall top to bottom talent inequality in wages is greater than in our benchmark setting. In Table B.2 we display the resulting optimal behavior of average and marginal tax rates over percentiles of the income distribution.

<sup>10</sup>In addition, to emphasize the behavior of wages for higher talents we estimate  $\alpha_3$  without weighting by the shares of talent in each occupation.

Table B.2: Optimal Tax Rates on Real Labor Income, Alternate Case

	Decade	Percentiles of Income					
		10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	99 <sup>th</sup>
<b>Averages</b>	70s	-3.3	0.9	7.8	20.3	24.7	20.8
	00s	-12.5	-8.7	4.0	21.6	28.3	23.7
<b>Marginals</b>	70s	17.2	28.8	40.0	38.9	24.3	-1.8
	00s	20.0	32.5	44.5	44.8	29.5	-1.5

Note: Estimates of tax rates determined using  $\frac{\partial a}{\partial k} = \alpha_3 \cdot v^2$ .

Relative to the benchmark, the striking difference is the sharp fall in average and increase in marginal rates over the time period. Now, the rise in comparative advantage dominates. It strengthens the wage compression channel and increases wage growth across talents. This increases the motive for redistribution towards the bottom.

## B.5 Counterfactuals

In this appendix we separately evaluate the impact of change in the  $a$  and  $b$  functions on optimal policy. To do so, we first hold the parameters of  $a$  fixed at their 1970s values, while allowing those of  $b$  to change to their 2000s values; we compute the corresponding optimal tax equilibrium. We then repeat the exercise holding the parameters of  $b$  fixed, while allowing those of  $a$  to change. We compare the resulting tax equilibria to those in which both functions are at their 1970s or 2000s levels.

**ASSIGNMENT** Figure B.7 shows the impact of the empirical  $a$  and  $b$  changes together and in isolation on the density of workers across tasks. Changes in  $b$  alone lead to quite large changes in the relative “number” of workers performing tasks. In particular, the polarizing adjustments in task demand (growth at the extremes relative to the middle) occurring between the 1970s and the 2000s induce growth in the density of workers at the extremes and, hence, job polarization in the associated optimal tax equilibrium. Changes in  $a$  alone have an opposite (if more modest) effect: the number of workers performing mid-level tasks grows relative to the extremes. This reflects

productivity growth in low tasks by low talents and in high tasks by high talents inducing reductions in shadow task prices at the top and the bottom and movements of some lower and higher talents into mid-level tasks. However, when changes in the  $b$  and  $a$  parameters are combined, it is the former that dominates.

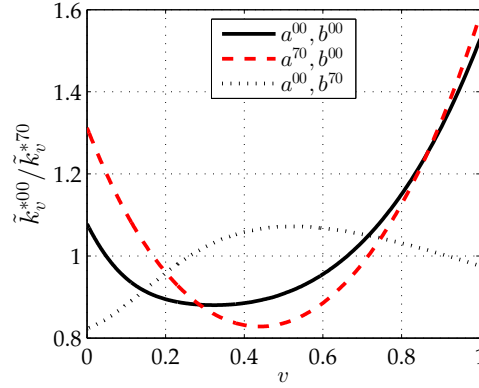


Figure B.7: Relative changes in  $\tilde{k}_v^*$  from the 1970s to the 2000s

Allowing  $b$  to change,  $a$  to change and both  $a$  and  $b$  to change.

Although, the  $b$  parameter change induces quite large changes in the numbers of workers performing particular tasks, this is achieved with only modest occupational reassignments of given workers. As shown in Figure B.10, low-mid level talents reduce their task assignment, but by no more than 2%, high-mid level talents increase their task assignment, but by no more than 3%.

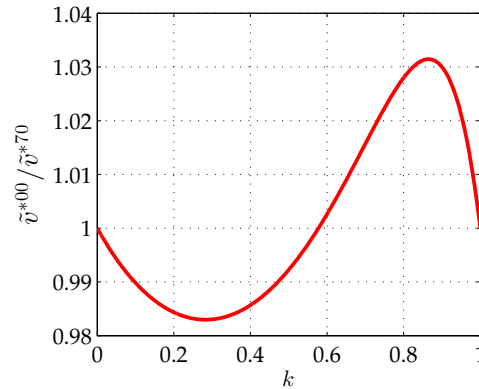


Figure B.8: Relative changes in task assignment  $\tilde{v}^*$  induced by the shift in  $b$  from 1970 to 2000.



**WAGE CHANGES** An implication of the modest change in task assignment induced by the shift in the  $b$  function is that equilibrium wage growth over talents is also only modestly altered by this shift. As shown in Figure B.9,  $b$  changes alone induce very slight compression in wage differentials across low-to-mid talents and very slight expansion across mid-to-high talents. Changes in the  $a$  function also depress wage growth across talents at the bottom and raise it at the top, but the effect is much more pronounced.

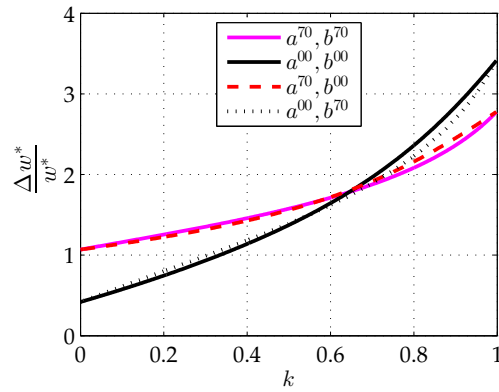
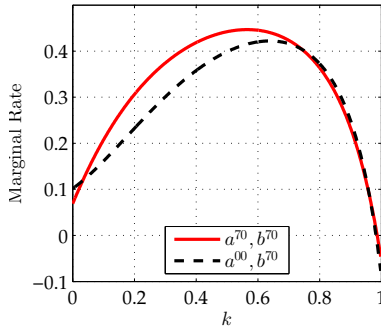
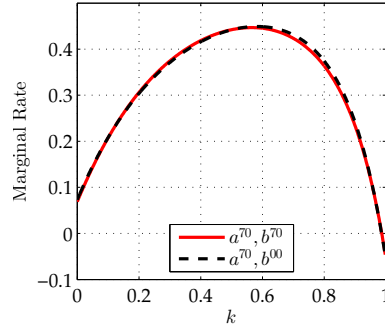


Figure B.9: Equilibrium wage growth  $\Delta w^*/w^*$  for the parameter combinations  $(a_{70}, b_{70})$ ,  $(a_{70}, b_{00})$ ,  $(a_{00}, b_{70})$  and  $(a_{00}, b_{00})$ .

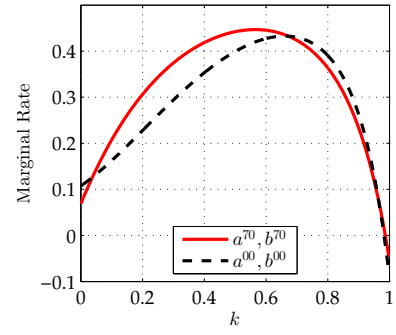
**MARGINAL TAX CHANGES** The shift in the  $b$  function alone has limited impact on the relative wage-effort elasticities and on the wage compression term. Combined with its small impact on wage growth over talents, it has a correspondingly modest effect on marginal taxes, see Figure B.10a. In contrast, the shift in the  $a$  function has a much more significant impact on wage growth and on the relative wage-effort elasticities. It has a much more significant effect on optimal marginal taxes and accounts for most of the adjustment between the 1970s and the 2000s.



(a) Effect of changes in  $a$ .



(b) Effect of changes in  $b$ .



(c) Effect of changes in  $a, b$ .

Figure B.10: Marginal tax effects.

## B.6 Optimal policy under a Rawlsian objective

In this subsection, we recompute optimal taxes under a Rawlsian societal objective that attaches positive weight only to the utility of the lowest talent. The results are given in Table B.3. Relative to the benchmark case, the government's enhanced concern for redistribution translates into higher marginal income tax rates at nearly all income levels, larger subsidies to low and middle income earners and larger average taxes on those in the top quartile. However, the impact of technical change remains unaltered. Marginal taxes fall on low to middle income quantiles (but not at the very lowest), rise on high income quantiles (but not the very highest). Average taxes rise sharply at the bottom of the income distribution and fall at mid to higher incomes. The largest beneficiary are those at the 75 income percentile who see the largest reduction in average taxes (from 34% to 28.2%).

Table B.3: Optimal Tax Rates on Real Labor Income: Rawlsian Case.

	Decade	Percentiles of Income					
		10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	99 <sup>th</sup>
<b>Averages</b>	70s	-144.1	-79.4	- 6.1	34.0	41.3	34.5
	00s	-54.1	-39.4	-6.3	28.2	39.6	32.4
<b>Marginals</b>	70s	80.6	78.9	74.4	62.2	38.0	-0.4
	00s	70.3	71.1	70.3	63.0	41.4	-3.9

## B.7 A Model with Intra-occupational Wage Dispersion and a Heavy-tailed Talent Distribution

In the main text, we consider a baseline model in which worker talent and task complexity are compressed to single dimensional variables. This formulation simplifies the theoretical analysis, facilitates empirical identification and connects models of technical change, assignment and taxation in a very direct way. However, the model's equilibrium does not permit intra-task wage variation. In this appendix we explore the extent to which our baseline results are qualified by its omission. In particular, to the extent that such wage variation is underpinned by variation in talent that is unrelated to tasks, the link between wages and tasks is weakened. A natural conjecture is that the responsiveness of policy to (task-level) technical change is similarly weakened. As described in Section 2.7, we enhance the contribution of talent unrelated to tasks and seek a lower bound for the responsiveness of policy to technical change.

In the remainder of the appendix, we proceed as follows. First, we detail a general framework that accommodates high dimensional talent and high dimensional tasks. We show how this framework can be specialized to yield the model of [Rothschild and Scheuer \(2013\)](#) and our (baseline) model in the main text and, in so doing, relate the two. We then develop a specialization that is intermediate between these cases. In this specialization, there are two aspects of worker talent: one captures comparative advantage in complex tasks, the other the ability to do all things well. The latter creates intra-task wage variation. Similar to our baseline model, comparative advan-

tag types partition the ordered space of occupations amongst themselves. We attribute all wage variation within these partitions to variations in absolute advantage uncorrelated with task. Such variation diffuses the effect of task-based technological change across the wage distribution and the impact of changes in taxes directed at a particular income across the set of tasks, task shadow prices and wages. These effects dampen the impact of technical change on policy design. In taking the model to the data we assume a coarse set of comparative advantage types and attribute all residual wage variation within the comparative advantage partitions (about 75% of total wage dispersion in our CPS sample) to variations in absolute advantage. We attribute none to measurement error in incomes, hours or tasks and absorb some measured inter-occupational wage variation into the residual by keeping the set of comparative advantage types and number of partitions small. These assumptions enlarge the dampening effect of absolute advantage variation. We find that the impact of technical change on policy is smaller, but the direction is unchanged. The qualitative conclusion that marginal taxes should be reduced on low to middle incomes, but raised on higher ones (with opposite adjustments in the extreme tails) remains intact. We interpret these results as lower bounds for the responsiveness of policy to technical change.

### B.7.1 A general framework

We first develop a general framework in which, as in Section 2.3, (consumption and effort) allocations are defined as functions of workers' types and the domain of the production function is the space of effort allocations. Subsequently, we introduce assignment. Relative to the main text the generality lies in our treatment of the talent space.

Assume that the workers are partitioned across talents according to a probability space  $(\Theta, \mathcal{F}, P)$ , where, to begin with,  $\Theta$  is an arbitrary set. Let  $\mathcal{A} = \mathcal{C} \times \mathcal{E}$  denote a set of allocations with each allocation a pair of measurable functions  $c \in \mathcal{C}$ ,  $c : \Theta \rightarrow \mathbb{R}_+$ , and  $e \in \mathcal{E}$ ,  $e : \Theta \rightarrow \mathbb{R}_+$ , describing the consumption and effort of differently talented workers. The set  $\mathcal{E}$  is further restricted to be a Banach space. Let  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  denote a production function defined directly on the space of effort allocation functions, with  $F$  concave and (Fréchet) differentiable. The government's problem is:

$$\sup_{\mathcal{A}} \int U(c(\theta), e(\theta)) P(d\theta) \tag{B.19}$$

subject to  $\forall \theta, \theta'$ ,

$$U(c(\theta), e(\theta)) \geq U\left(c(\theta'), \frac{\omega(\theta', e)e(\theta')}{\omega(\theta, e)}\right) \quad (\text{B.20})$$

and

$$\int \{\omega(\theta, e)e(\theta) - c(\theta)\}P(d\theta) \geq G, \quad (\text{B.21})$$

where (B.20) and (B.21) are, respectively, the incentive-compatibility and resource constraints and wages  $\omega$  are given by the Fréchet derivative of  $F$ .<sup>11</sup>  $G \in \mathbb{R}_+$  is government spending.

This framework can interpreted as the reduced form of an economy with assignment. Let  $(V, \mathcal{V})$  denote a measurable space of tasks and  $\mathcal{M}$  the set of finite measures on  $V$ . Interpret such measures as allocations of effective labor across tasks. Let  $H : \mathcal{M} \rightarrow \mathbb{R}_+$  be a production function (now defined on the space of effective labor allocations) and  $\mu : \Theta \times V \rightarrow \mathbb{R}_+$  a (measurable) productivity kernel giving the productivity of each talent in each task. As in Section 2.4, assignment is efficient in the competitive equilibria and planner's problems that we consider. Thus, as there, an indirect production function over effort allocations is recoverable from an assignment problem:

$$F(e) = \sup_{m, \Lambda} \left\{ H(m) : \begin{array}{l} \forall B \in \mathcal{V}, \int_B m(dv) = \int_B \int_{\Theta} \mu(\theta, v)e(\theta)\Lambda(d\theta, dv) \\ \forall B \in \mathcal{F}, \int_B P(d\theta) = \int_V \int_B \Lambda(d\theta, dv) \end{array} \right\}, \quad (\text{B.22})$$

where  $m$  is a distribution of effective labor across tasks and  $\Lambda$  is a distribution of workers across tasks and talents. The constraints in (B.22) ensure that these distributions are consistent with one another, the effort allocation and the underlying distribution of talent. It follows that, as in the main text, the planner's problem with assignment can be decomposed into an outer step (B.19) in which  $(c, e)$  are chosen and an inner step (B.22) in which  $(m, \Lambda)$  are chosen (to determine  $F$  at  $e$ ).

## B.7.2 Reformulation of the general framework

Enlarging the dimension of talent to allow for multiple attributes that interact differently across tasks greatly complicates the pattern of binding incentive constraints. [Rothschild and Scheuer \(2013\)](#) observe that if  $\Theta$  is uncountable, then (almost all) talents earning the same wage receive the same consumption and effort. Consequently, allocations can be re-expressed as functions of (one

---

<sup>11</sup>We continue to exclude externalities which is a focus of [Rothschild and Scheuer \(2014\)](#) and [Lockwood et al. \(2014\)](#).

dimensional) wages and in this form only local incentive constraints bind. This is an important simplifying insight that we utilize below. However, it is not costless as it requires the introduction of rather complicated constraints that relate the (endogenous) wage distribution to allocations (as functions of wages). In the remainder of this section, we make one simplification: in the inner assignment problem, we restrict attention to effective labor allocations described by (density) functions  $l : V \rightarrow \mathbb{R}_+$  rather than measures. Thus, we exclude atoms of effective labor in tasks. As before, let  $H$  denote the production function, but defined now on the domain of such densities. In addition, assume that  $H$  is (Fréchet) differentiable with derivative  $\partial H$ . The shadow price of task  $v$  output is  $\partial H(l; v)$  and the wage of a worker of talent  $\theta$  in task  $v$  is:  $\partial H(l, v)\mu(\theta, v)$ . Workers choose their tasks to maximize their wages:

$$v^*(\theta; l) \in \operatorname{argmax}_V \partial H(l; v)\mu(\theta, v). \quad (\text{B.23})$$

Let  $w^*(\theta; l) = \partial H(l; v^*(\theta; l))\mu(\theta, v^*(\theta, l))$  denote the (maximized) wage of a worker of talent  $\theta$  given the labor allocation  $l$ . By (B.23),  $l$  implies a distribution of workers over wages and tasks:

$$R(l)(w, v) := \int_{\Theta} 1(w^*(\theta; l) \leq w, v^*(\theta; l) \leq v) P(d\theta). \quad (\text{B.24})$$

In this setting, define an allocation to be a triple  $(l, \tilde{c}, \tilde{e})$  with  $\tilde{c} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and  $\tilde{e} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  mapping *wages* rather than talents to consumption and effort choices. Note that from (B.23), given  $l$ , any worker receiving wage  $w$  in task  $v$  has productivity  $\frac{w}{\partial H(l; v)}$  and an effective labor of  $\frac{w}{\partial H(l; v)} \tilde{e}(w)$ . Consistency of  $l$  with  $\tilde{e}$  thus requires:

$$\forall v : \quad l(v) = \int_0^\infty \frac{w \tilde{e}(w)}{\partial H(l; v)} R(l)(dw, v). \quad (\text{B.25})$$

In addition, the wage distribution  $Q$  must equal the wage marginal of  $R(l)$ :

$$\forall w : \quad Q(w) = \int_0^w \int_V R(l)(dw', dv'). \quad (\text{B.26})$$

Hence, the government's problem can be re-expressed as:

$$\sup_{\tilde{c}, \tilde{e}, l, Q} \int_0^\infty U(\tilde{c}(w), \tilde{e}(w)) Q(dw) \quad (\text{B.27})$$

subject to (B.25) and (B.26),  $\forall w, w'$ ,

$$U(\tilde{c}(w), \tilde{e}(w)) \geq U\left(\tilde{c}(w'), \frac{w' \tilde{e}(w')}{w}\right), \quad (\text{B.28})$$

and

$$\int_0^\infty \{w \tilde{e}(w) - \tilde{c}(w)\} Q(dw) \geq G. \quad (\text{B.29})$$

The main difficulties in (B.27) are the constraints (B.25) and (B.26) relating  $Q$  to  $l$  and  $\tilde{e}$ , absent which everything would reduce to a standard Mirrlees problem.

ROTHSCHILD AND SCHEUER'S SPECIALIZATION Rothschild and Scheuer (2013) make progress by assuming that: (i)  $V = \{1, 2\}$ , (ii)  $\Theta = \mathbb{R}^2$  with  $\theta = (\theta_1, \theta_2)$  and  $\mu(\theta, v) = \theta_v$  and (iii)  $P$  has a density  $p$ . These restrictions make (B.25) and (B.26) manageable.  $H$  is now a function of only two variables  $(l_1, l_2)$  and  $R$  is given by:

$$R(l)(w, 1) = \int_{\underline{\theta}}^{\frac{w}{\partial H(l;1)}} \int_{\underline{\theta}}^{\theta_1 \frac{\partial H(l;1)}{\partial H(l;2)}} p(\theta_1, \theta_2) d\theta_2 d\theta_1 \quad (\text{E.6'a})$$

$$R(l)(w, 2) = \int_{\underline{\theta}}^{\theta_2 \frac{\partial H(l;2)}{\partial H(l;1)}} \int_{\underline{\theta}}^{\frac{w}{\partial H(l;2)}} p(\theta_1, \theta_2) d\theta_2 d\theta_1. \quad (\text{E.6'b})$$

The low dimensionality of  $V$  (and, hence,  $l$ ) suggests an inner-outer approach to solving (B.27) quite distinct from that described in the main text. The inner component maximizes (B.27) over  $(\tilde{c}, \tilde{e})$  subject to (B.25), (B.28) and (B.29) with  $l$  fixed and  $Q$  set to  $Q(\cdot) = \sum_{v=1,2} R(l)(\cdot, v)$ . This problem is a standard Mirrlees problem augmented by (B.25). The outer component maximizes the resulting value function over  $l$ . Note that here assignment of effective labor across tasks  $l$  is solved for in the *outer step*, in contrast to the formulation in the main text where this is done in the inner step.

AN ALTERNATIVE SPECIALIZATION We now present a version of the general formulation given above in which workers are distributed across two talent attributes. The first interacts with tasks and affects comparative advantage, the second influences the ability to do all things and absolute advantage. Let  $V := [0, 1]$  and  $\Theta := \{1, \dots, K\} \times \mathbb{R}_+$ . Denote elements of  $\Theta$  by  $\theta := (k, \psi)$  and assume that  $\mu$  has the form  $\mu(\theta, v) := \psi a_k(v)$  where the function  $a$  is log super-modular in  $(k, v)$ . Workers of a given  $k$  type have the same profile of relative wages and the same preference ordering over tasks; variations in  $\psi$  cause workers to be more or less good at all things and underpin intra-task wage dispersion. Let  $\{\pi_k\}_{k=1}^K$  denote the distribution of workers across  $k$  and  $\{f_k\}_{k=1}^K$  the densities of workers over  $\psi$  conditional on  $k$ . The latter are assumed to satisfy  $E_k[\psi] = 1$ . Also let  $H(l) = [\int_0^1 b(v) l(v)^{\frac{\varepsilon-1}{\varepsilon}} dv]^{\frac{\varepsilon}{\varepsilon-1}}$ . As in the main text,  $k$ -types (i.e. workers with a common  $k$ , but potentially different values of  $\psi$ ) sort themselves across tasks with those of a given type  $k$  distributing themselves over a sub-interval  $[\tilde{v}_{k-1}, \tilde{v}_k]$  so as to ensure a common value for  $w_k := \partial H(l, v) a_k(v)$ . The wage received by a  $(k, \psi)$ -type is  $w = \psi w_k$  and, since  $E_k[\psi] = 1$ ,  $w_k$  is the average wage per unit of effort received by  $k$ -types. In this setting, it is useful to define the  $\psi$ -weighted average effort (“effective labor supply”) of the  $k$ -th type:

$$\ell_k = \int_0^\infty \psi e(k, \psi) f_k(\psi) d\psi = \int_0^\infty \frac{w}{w_k} \tilde{e}(w) f_k\left(\frac{w}{w_k}\right) \frac{dw}{w_k}, \quad (\text{B.31})$$

where the second term expresses effort as a function of  $(k, \psi)$  and the third re-expresses it as a function of the wage and uses  $w = \psi w_k$  to change variables from  $\psi$  to  $w$ . The vector of effective labor supplies  $\ell = \{\ell_k\}$  (rather than the full allocation of effective labor over tasks  $l$ ) is sufficient to determine the average wages  $\{w_k\}$  and, hence, the complete wage distribution. Similar to the main text it can be shown that given  $\ell = \{\ell_k\}$ , final output is:

$$Y(\ell) := \max_{\{\tilde{v}_k\}} \left[ \sum_{k=1}^K B_k(\tilde{v}_{k-1}, \tilde{v}_k) \ell_k^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}, \quad (\text{B.32})$$

with  $B_k(\tilde{v}_{k-1}, \tilde{v}_k) = (\int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv)^{\frac{1}{\varepsilon}}$ . In addition, the wage terms  $\{w_k\}$  are given by:

$$w_k(\ell) = \left( \frac{Y(\ell)}{\ell_k} \right)^{\frac{1}{\varepsilon}} \tilde{B}_k(\ell), \quad (\text{B.33})$$



where  $\tilde{B}_k(\ell)$  is the value of  $B_k(\tilde{v}_{k-1}, \tilde{v}_k)$  at the optimized task thresholds from (B.32) and the notation makes the dependence of  $w_k$  on  $\ell$  explicit. Combining (B.31) and (B.33) gives an analogue of (B.25):

$$\ell_k = \int_0^\infty \frac{w}{w_k(\ell)} \tilde{e}(w) f_k \left( \frac{w}{w_k(\ell)} \right) \frac{dw}{w_k(\ell)}. \quad (\text{E.7''})$$

Variations in absolute advantage  $\psi$  create variations in the wages paid to members of a given  $k$ -type group. Thus, they create wage variation within each partition of the occupational space  $[\tilde{v}_k, \tilde{v}_{k+1}]$ . The overall wage distribution  $Q$  satisfies an analogue of (B.26):

$$Q(w) = \int_0^w \sum_{k=1}^K f_k \left( \frac{w'}{w_k(\ell)} \right) \frac{\pi_k}{w_k(\ell)} dw'. \quad (\text{E.8''})$$

The government's problem can then be expressed as:

$$\sup_{\tilde{c}, \tilde{e}, \ell, Q} \int U(\tilde{c}(w), \tilde{e}(w)) Q(dw) \quad (\text{B.34})$$

subject to (E.7''), (E.8'') and (B.28), and the resource constraint:

$$\int \{w\tilde{e}(w) - \tilde{c}(w)\} Q(dw) \geq G.$$

Assuming that the first order approach is valid (i.e. that the workers' envelope condition is sufficient for incentive compatibility), the implied optimal marginal tax paid by a worker earning  $w$  is:

$$\tau(w) = \frac{\frac{1-Q(w)}{wq(w)} \Psi(w) \mathcal{H}(w)}{1 + \frac{1-Q(w)}{wq(w)} \Psi(w) \mathcal{H}(w)} + \frac{\mathcal{N}(w)}{1 + \frac{1-Q(w)}{wq(w)} \Psi(w) \mathcal{H}(w)}, \quad (\text{B.35})$$

where  $q(w)$  is the wage density,  $\Psi(w)$  is the normalized multiplier on the incentive constraint,  $\mathcal{H}(w) = \frac{1+\mathcal{E}^u}{\mathcal{E}^c}$ , with  $\mathcal{E}^u$  the uncompensated and  $\mathcal{E}^c$  the compensated labor supply elasticities. The first right hand side component of (B.35) is the conventional Mirrlees tax term; the second is the wage compression term. A tax induced effort perturbation at  $w$  impacts the effective labor of each  $k$ -type population (since each includes some workers receiving wage  $w$ ). This, in turn, after task migration of  $k$  types, affects the  $w_k$  terms and, hence, the entire wage distribution. These effects

are seen by decomposing the wage compression term numerator  $\mathcal{N}$  as:

$$\mathcal{N}(w) = N \cdot L(w),$$

where  $L(w)$  is a column vector giving the impact of a perturbation in  $\tilde{e}(w)$  on the vector of effective labor supplies  $\ell = \{\ell_k\}$  and  $N$  is a row vector giving the shadow value of a perturbation in  $\ell$  on the distribution of wages. Specifically, the  $k$ -th element of  $N$  is:

$$N_k := \frac{1}{\ell_k} \int_0^\infty \lambda(w') \sum_{\kappa=1}^K \left( \frac{w_\kappa(\ell)}{q_\kappa(w'; w_\kappa(\ell))} \frac{\partial q_\kappa(w'; w_\kappa(\ell))}{\partial w_\kappa} \right) \left( \frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k} \right) q_\kappa(w'; w_\kappa(\ell)) dw',$$

where  $\frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k}$  is the elasticity of the  $\kappa$ -th comparative advantage type's average wage with respect to the  $k$ -th type's effective labor (and an analogue of the relative wage-effort elasticities from the main text),  $\frac{w_\kappa(\ell)}{q_\kappa(w'; w_\kappa(\ell))} \frac{\partial q_\kappa(w'; w_\kappa(\ell))}{\partial w_\kappa}$  is the elasticity of the  $\kappa$ -th type's conditional wage density at  $w'$  with respect to  $w_\kappa$  and  $\lambda(w') = V(w') + \chi\{w'\tilde{e}(w') - \tilde{c}(w')\}$  is the societal value of the allocation given to workers earning  $w'$  (i.e. the utility plus the shadow value of the resource surplus these workers generate). As in the main text, technical change impacts the wage functions  $w_\kappa$  and the wage elasticities  $\frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k}$  directly and via the reallocation of  $k$ -types across tasks. The affect of these changes on the wage distribution and tax policy's ability to shape this distribution is now diffused through the densities  $q_k$ . They modify the hazard term  $\frac{1-Q(w)}{q(w)w}$  (the analogue of  $\frac{\Delta w_{k+1}^*}{w_{k+1}^*} \frac{1-\Pi_k}{\pi_k}$  in the main text) and, hence, the Mirrlees term in (B.35). They also change the numerator of the wage compression term which incorporates the elasticities  $\frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k}$ .

**BRINGING THE FORMULATION TO DATA** Via the introduction of the additional talent attribute  $\psi$ , the above formulation permits intra-task wage variation and an unbounded wage distribution. In these dimensions it advances the baseline model in the main text. However, numerically solving (B.34) becomes challenging as  $K$  becomes large. Thus, in contrast to the simpler baseline model in which the set of  $k$  values is uncountable,  $K$  is restricted to equal four in the calculations below.<sup>12</sup> The above formulation inherits many of its parameters from the baseline model. In our calculations below, we retain earlier values for these parameters. In particular, utility parameters are reused, the final goods production function is assumed to be Cobb-Douglas and our earlier estimates of  $b$  and

<sup>12</sup>These calculations use the non-linear optimizer SNOPT. The code is available on request.

$a$  retained. Values for the conditional densities  $\{f_k\}$ , the new parameters in (B.34), are required. To derive estimates of these, and following the approach in the main text, we use a worker's occupation to infer his or her  $k$ -type. Specifically, we order occupations by average wage, use this ordering and a worker's occupation to rank workers and then recombine workers into  $K = 4$  ranked and equally sized  $k$  groups. Thus, a worker is of type  $k = 1$  if she belongs to the first  $\frac{1}{K}$  of workers by (ordered) occupation, she is of type  $k = 2$  if she belongs to the next  $\frac{1}{K}$  workers by (ordered) occupation and so on. To capture  $\psi$  dispersion, we fit Burr Type XII distributions to each demeaned  $k$ -group. The density of a Burr Type XII distribution converges asymptotically to a Pareto density in the right tail, but admits a non-Paretian form that better accounts for wage data over the remainder of its domain.<sup>13</sup> The Burr XII is estimated by maximum likelihood; the estimation procedure accounts for top coded observations. For the  $k = 4$  group, our estimated Pareto tail parameter is 3.07. The fitted Burr XII distributions account for about 75% of the overall wage dispersion in our sample.

**RESULTS** Table B.4 gives the resulting optimal marginal tax rates. Relative to the benchmark case considered in the main text, this exercise generates similar values for marginal tax rates on middle incomes, but, as expected, quite different values for marginal rates on incomes in the tails of the distribution. The latter are shaped by the tails of the Burr XII distributions.

Table B.4: Optimal Tax Rates on Real Labor Income.

	Decade	Percentiles of Income			
		10 <sup>th</sup>	50 <sup>th</sup>	90 <sup>th</sup>	99 <sup>th</sup>
<b>Marginals</b>	70s	41.8	48.2	42.9	44.5
	00s	43.2	45.7	43.7	43.4

The response of optimal marginal tax rates to technological change, while dampened, is qualitatively similar to before: marginal rates rise over the lowest two income deciles, fall over low-mid income deciles (between the second and seventh decile), rise over higher income deciles (seventh

<sup>13</sup>The Burr Type XII distribution has density  $g(x) = \frac{cdx^{c-1}}{(1+x^c)^{d+1}}$  and hazard  $\frac{1-G(x)}{g(x)x} = \frac{1+x^c}{cdx^c}$ , where  $c$  and  $d$  are parameters. It was proposed as flexible description of income distributions by Singh and Maddala (1976). McDonald (1984) concludes that the Burr XII outperforms many other heavy-tailed distributions in describing the distribution of income.

to ninth) and fall at the very top, see Figure B.11 which gives changes in rates. The dampening of the response, marginal tax rates change by at most 2.5 points at a given income compared to a maximum of 8.5 points in the baseline case, stems from the introduction of variation in absolute advantage that diffuses the impact of both technical change and taxes on wages. However, by assuming only  $K = 4$  comparative advantage groups and by attributing all residual measured wage variation to absolute advantage, we make this dampening force very large. We interpret these numerical results as lower bounds on responsiveness of optimal policy to technical change.

After decomposition, the Mirrlees and Wage Compression terms display similar movements to those in the main text: changes in the Mirrleesian term prevail over most of the income quantile domain, but are dampened by offsetting movements in the wage compression term. For example, at the fourth income decile, the Mirrlees term falls by more than 4 points, while the wage compression term rises by about 2 points. The logic behind changes to the Mirrlees and compression terms is essentially that in the main text. Expressed in terms of changes to the wage distribution rather than the structure of binding incentive constraints, technical change has compressed wage differentials at the bottom, but increased them at the top. This has thinned out the wage distribution in the tails, but fattened it in the middle. Since, other things equal, it is desirable to impose higher marginal taxes in areas of the distribution where the wage density is low and few will be distorted, the Mirrlees component falls on the low-middle incomes, but rises on high incomes (and on very low incomes in the extreme lower tail). The wage compression effect is enhanced by technical change as reduced substitutability of workers across tasks gives the government, through tax policy, more leverage over task shadow prices and wages. This creates a wage compression motive for raising taxes at the bottom and lowering them at the top. Overall change in the tax code depends on the balance of these forces, with the wage compression force only predominating in the extreme tails.

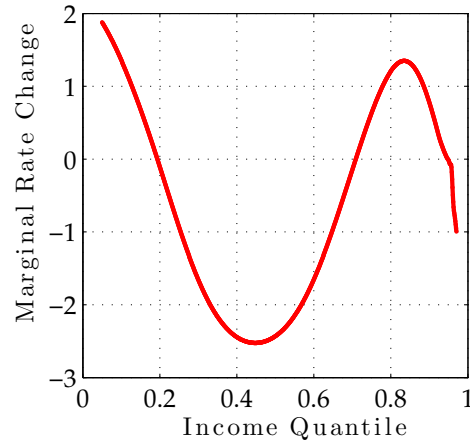


Figure B.11: Change in Marginal Taxes.

Overall, while quantitative responses are more muted than in the benchmark case, the broad policy prescription of modest marginal tax reductions over a broad band of lower-middle income quantiles combined with an increase over higher quantiles emerges as a robust finding.

## B.8 Intra-occupational task variation

The O\*NET database collects information on the content of occupations from two sources - occupational analysts and a direct survey of US workers and establishments. The latter permits some unbundling of occupations since those surveyed are asked to assess the knowledge (educational and training) requirements of the occupations with which they are associated. Variation in responses gives an indication of the variety of tasks that might be performed by different workers within the same occupation.<sup>14</sup> The O\*NET reports statistics summarizing the distribution of these responses. From the statistics contained within the latest release of the O\*NET we construct Leik's ordinal variation indices (see, [Weisberg \(1992\)](#));<sup>15</sup> we then average the indices associated with every occupation to create a single index for each. We interpret this index as a measure of disagreement

<sup>14</sup>Variations in assessments of the skill and ability content of occupations would be preferable, but these assessments are obtained from analysts and are not requested in the establishment survey. At issue is whether the higher levels of training and education thought necessary by some respondents are indicative of higher productivities within a task or whether they are indicative of the performance of more complex tasks within a set of tasks defining the occupation. Our stance is that they are at least partly the latter.

<sup>15</sup>This index provides a measure of variation for ranked, categorical variables. The index is equal to zero for a degenerate distribution and assumes the maximum value of one for a polarized distribution with equal weight on the two extremal categories.

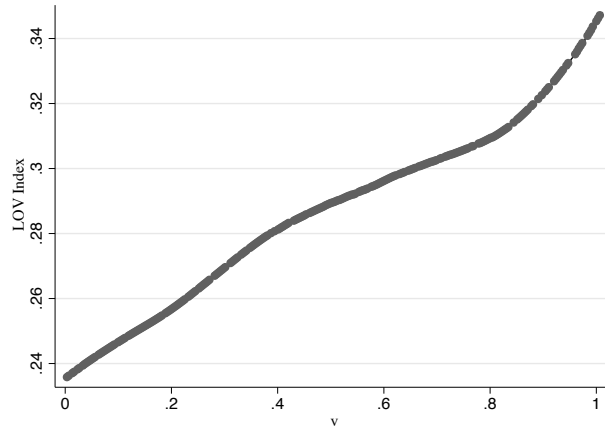


Figure B.12: Average Leik's ordinal variation (LOV) index concerning the knowledge requirements of occupations.

amongst workers and establishments as to the knowledge content of an occupation (with a value of zero indicating complete agreement and a value of one maximal disagreement) and as a proxy for the variety of (complexity-ranked) tasks within an occupation. In Figure B.12, we plot this (Lowess-smoothed) 'disagreement' index across all ranked occupations  $v$ . As the figure indicates, this index is greater than zero across all  $v$ 's. This suggests a degree of dispersion and disagreement in the survey answers. In addition, the value is increasing in  $v$  - the regression coefficient on non-smoothed data is statistically significant with value is .092 (.013). Overall, these results are consistent with intra-occupational, complexity-ranked task variety that is increasing in the occupation's average wage. This, in turn, suggests that variation in average occupational wages is a lower bound for wage variation across (complexity-ranked) tasks, especially at the upper end of the wage distribution.

## Appendix C

# Appendix for Chapter 3

### C.1 Appendix

*Proof of Proposition 4.* The proof follows from an argument similar to that in Example 1. Suppose every bidder other than bidder 1 uses no resale equilibrium strategy. Consider bidder 1 with value 1 and the alternative strategy  $(\beta_1^N(1), \beta_1^N(1), 0, \dots, 0)$  (similar strategies can be found for other values in  $(0, 1)$ ). With this strategy, this bidder will receive two items. She can sell the second item by using a second-price auction with a reserve price  $\frac{1}{2}$ , which gives her an expected revenue that is strictly higher than  $\mathbb{E} [Y_k^{(n-1)}]$ . This is because  $\mathbb{E} [Y_k^{(n-1)}]$  is the expected revenue of the second price auction *with no reserve price*, and the expected revenue of a second price auction with optimal reserve price is strictly higher than that. Since this bidder has paid  $\mathbb{E} [Y_k^{(n-1)}]$  for the second item and gets strictly more than  $\mathbb{E} [Y_k^{(n-1)}]$  in the resale stage, this deviation strictly increases her utility.  $\beta^N(x)$  is not an equilibrium of discriminatory auctions with resale.  $\square$

*Proof of Lemma 4.* The first equation follows from noting that  $\psi_x(z) = \left(z - \frac{F(x)-F(z)}{f(z)}\right)$  and  $F_1^{(n-1)}(z | x) = \left(\frac{F(z)}{F(x)}\right)^{n-1}$ . Denote  $H(x, z) := \frac{n-1}{F(x)^{n-1}} \left(z - \frac{F(x)-F(z)}{f(z)}\right) F(z)^{n-2} f(z)$ . Then, we have

$$\gamma(x) = \int_{\psi_x^{-1}(0)}^x H(x, z) dz$$

and

$$\gamma'(x) = \left( \int_{\psi_x^{-1}(0)}^x \frac{\partial}{\partial x} H(x, z) dz + H(x, x) - H(x, \psi_x^{-1}(0)) \right).$$

Note that  $H(x, x) = (n-1) \frac{xf(x)}{F(x)}$  and  $H(x, \psi_x^{-1}(0)) = 0$ . Hence,

$$\gamma'(x) = \left( \int_{\psi_x^{-1}(0)}^x \frac{\partial}{\partial x} H(x, z) dz + (n-1) \frac{xf(x)}{F(x)} \right).$$

Moreover,

$$\begin{aligned} \frac{\partial}{\partial x} H(x, z) &= (n-1) F(z)^{n-2} f(z) \frac{-(n-1)}{F(x)^n} f(x) \left( z - \frac{F(x) - F(z)}{f(z)} \right) \\ &\quad + (n-1) F(z)^{n-2} f(z) \frac{1}{F(x)^{n-1}} \left( -\frac{f(x)}{f(z)} \right) \\ &= -H(x, z) \frac{(n-1) f(x)}{F(x)} + (n-1) F(z)^{n-2} f(z) \frac{1}{F(x)^{n-1}} \left( -\frac{f(x)}{f(z)} \right) \end{aligned}$$

Thus,

$$\begin{aligned} \int_{\psi_x^{-1}(0)}^x \frac{\partial}{\partial x} H(x, z) dz &= \int_{\psi_x^{-1}(0)}^x \left( -H(x, z) \frac{(n-1) f(x)}{F(x)} + (n-1) F(z)^{n-2} f(z) \frac{1}{F(x)^{n-1}} \left( -\frac{f(x)}{f(z)} \right) \right) dz \\ &= -\frac{(n-1) f(x)}{F(x)} \int_{\psi_x^{-1}(0)}^x H(x, z) dz - \frac{(n-1) f(x)}{F(x)^{n-1}} \int_{\psi_x^{-1}(0)}^x F(z)^{n-2} dz \\ &= -\frac{(n-1) f(x)}{F(x)} \gamma(x) - \frac{(n-1) f(x)}{F(x)^{n-1}} \int_{\psi_x^{-1}(0)}^x F(z)^{n-2} dz \end{aligned}$$

Hence, we have

$$\begin{aligned} \gamma'(x) &= \left( \int_{\psi_x^{-1}(0)}^x \frac{\partial}{\partial x} H(x, z) dz + (n-1) \frac{xf(x)}{F(x)} \right) \\ &= (n-1) \left( \frac{f(x)}{F(x)} (x - \gamma(x)) - \frac{f(x)}{F(x)^{n-1}} \int_{\psi_x^{-1}(0)}^x F(y)^{n-2} dy \right) \end{aligned}$$

□

*Proof of Lemma 5.* Consider optimally selling two items to three bidders whose valuations are uniformly distributed in  $[0, y]$ . The optimal mechanism is a uniform price auction with reserve price  $\frac{y}{2}$  and revenue given by

$$\frac{3}{8} \times \frac{y}{2} + \frac{3}{8} \times 2 \times \frac{y}{2} + \frac{1}{8} \times 2 \times \frac{5y}{8} = \frac{23}{32} y$$



If a bidder with value  $x$  wins three items, his prior belief is that three remaining bidders all have values smaller than  $\beta^{-1}(\theta(x))$ . Hence the conclusion follows.  $\square$

*Proof of Lemma 6.* Consider optimally selling one item to two bidders whose valuations are uniformly distributed in  $[0, y]$ . The optimal mechanism is a second price auction with reserve price  $\frac{y}{2}$  and revenue given by

$$2 \times \frac{1}{2} \times \frac{1}{2} \times \frac{y}{2} + \frac{1}{2} \times \frac{1}{2} \times \frac{2y}{3} = \frac{5}{12}y$$

When a bidder with value  $x$  bids  $\delta(x) = \theta(x)$  he wins two items. This means that highest valued bidder among the competitor has a value greater than  $\beta^{-1}(\delta(x))$  and each of the two lowest value bidders has values smaller than  $\beta^{-1}(\theta(x)) = \beta^{-1}(\delta(x))$ . Since the two lowest valued bidders are potential buyers in the resale market, the conclusion follows.  $\square$

*Proof of Lemma 7.* First, by method of contradiction, suppose that  $\beta(x) > \frac{23}{64}x$  for some  $x \in [0, c]$ . Consider a bidder with value  $\theta^{-1}(\beta(x)) \equiv y$ . When this bidder receives three items from the auctioneer, his total payment for second and third object is  $\delta(y) + \theta(y) \geq 2\theta(y) = 2 \times \beta(x) > \frac{23}{32}x$ , whereas his expected revenue from resale for this case is only  $\frac{23}{32}x$ . Therefore he makes a loss when he receives 3 items. So, he is better off by deviating to  $(\beta(y), \delta(y), 0)$  from  $(\beta(y), \delta(y), \theta(y))$ .  $\square$

*Proof of Lemma 8.* For (i), by method of contradiction, suppose that there exist  $t \in [0, \theta(1)]$  such that  $(\beta^{-1}(t))' > \frac{32}{23}$  and  $\delta^{-1}(t) < \theta^{-1}(t)$ . Then we can argue that type  $\theta^{-1}(t) \equiv y$  strictly benefits by deviating to  $(\beta(y), \delta(y), t + \varepsilon)$  for small enough  $\varepsilon$ . This is because, by deviating to  $t + \varepsilon$  from  $t$  for his third bid (and this is feasible since  $\delta^{-1}(t) < \theta^{-1}(t)$ ), this bidder (i) increases the probability of getting three items, and (ii) increases his net utility when he sells two items to unassigned bidders (his payment is increases by  $\varepsilon$ , and his expected revenue increases by strictly more than  $\frac{23}{32} \times \frac{32}{23} \times \varepsilon = \varepsilon$ .)

Next, since  $\beta(0) = 0$  and  $\beta(x) \leq \frac{23}{64}x$  for all  $x \in [0, c]$ , we have  $\beta'(0) \leq \frac{23}{64}$  or  $(\beta^{-1}(0))' \geq \frac{64}{23} > \frac{32}{23}$ . Since  $\beta$  is continuously differentiable, there exists  $d \leq c$  such that  $\beta(x)' > \frac{23}{32}$  for all  $x \in [0, d]$  and part (i) implies that  $\delta(x) = \theta(x)$  for all  $x \in [0, d]$ .  $\square$

*Proof of Lemma 9.* By Lemma 8, we know that for all  $x \in [0, d]$ , we have  $\beta(x) \in [0, \frac{23}{64}x]$  and  $\delta(x) = \theta(x)$ . Let us first calculate the net utility of a buyer with value  $x \in [0, d]$  when he is a seller in the resale stage. Appealing to Lemmas 5 and 6, this is given by

$$r(x) \equiv \beta^{-1}(\theta(x))^3 \left( \frac{23}{32} \beta^{-1}(\theta(x)) - 2\theta(x) \right) + 3\beta^{-1}(\theta(x))^2 \left( x - \beta^{-1}(\theta(x)) \right) \left( \frac{5}{12} \beta^{-1}(\theta(x)) - \theta(x) \right).$$

First of all, if  $r'(x) > 0$  then we have  $\theta(x) = \beta(x)$ . This is because whenever  $r'(x)$  is positive, a bidder with value  $x$  becomes strictly better off by increasing his second and third bids by  $\varepsilon$ , and doing this would be feasible if  $\beta(x) > \theta(x)$ .

Next, by method of contradiction, suppose that there exists no  $e \in (0, 1]$  such that for all  $x \in [0, e]$ ,  $\beta(x) = \delta(x) = \theta(x)$ . This means there exists  $f > 0$  such that we have  $\beta(x) > \theta(x)$  for all  $x \in (0, f]$ . Now, we argue that for all  $x \in (0, f]$ , we have  $r(x) > 0$ . This is because for all  $x \in (0, f]$ , (i)  $x > \beta^{-1}(\theta(x))$ , (ii)  $\beta^{-1}(\theta(x)) \geq \frac{64}{23}\theta(x)$ , hence

$$\begin{aligned} r(x) &> \beta^{-1}(\theta(x))^3 \left( \frac{23}{32} \frac{64}{23} \theta(x) - 2\theta(x) \right) + 3\beta^{-1}(\theta(x))^2 \left( x - \beta^{-1}(\theta(x)) \right) \left( \frac{5}{12} \beta^{-1}(\theta(x)) - \theta(x) \right) \\ &\geq 3\beta^{-1}(\theta(x))^2 \left( x - \beta^{-1}(\theta(x)) \right) \left( \frac{5}{12} \frac{64}{23} \theta(x) - \theta(x) \right) \\ &> 0 \end{aligned}$$

since  $\frac{5}{12} \frac{64}{23} \cong 1.159 > 1$ . Since  $r(0) = 0$  and  $r(x) > 0$  for all  $x \in (0, f]$ , there exists  $y \in (0, f]$  such that  $r'(y) > 0$ , which implies  $\theta(y) = \beta(y)$ , a contradiction.

□

*Proof of Lemma 10.* Consider a bidder with value  $x \in (0, e)$  who bids as if his value is  $z$  (which is very close to  $x$ .) His expected utility is given by

$$u(x, z) = z^3 \left( x - 3\beta(z) + \frac{23}{32}z \right) + R(x, z)$$

where  $R(x, z)$  is his expected utility from resale stage when he is a buyer and is given by

$$\begin{aligned} R(x, z) = & 6 \int_z^{\max\{1, 2x\}} \int_{\frac{k}{2}}^x \int_{\frac{k}{2}}^l (x - m) dm dl dk + 6 \int_z^{\max\{1, 2x\}} \int_x^k \int_{\frac{k}{2}}^x (x - m) dm dl dk \\ & + 6 \int_z^{\max\{1, 2x\}} \int_{\frac{k}{2}}^k \int_0^{\frac{k}{2}} \left(x - \frac{k}{2}\right) dm dl dk + 6 \int_z^{\max\{1, 2x\}} \int_0^{\frac{k}{2}} \int_0^l \left(x - \frac{k}{2}\right) dm dl dk \end{aligned}$$

(where  $k, l, m$  denote the realizations for highest, the second highest, and the third highest values among the competitors, the first two terms in the summation represent the cases in which the bidder with value  $x$  pays the third highest value, and the last two terms in the summation represent the cases in which the bidder with value  $x$  pays the reserve price).

A necessary condition for this to be an equilibrium is  $\frac{\partial u(x, z)}{\partial z} \Big|_{z=x} = 0$ . Note that  $\frac{\partial R(x, z)}{\partial z} = -6 \left( \int_{\frac{x}{2}}^x \int_{\frac{x}{2}}^l (x - m) dm dl + \int_x^z \int_{\frac{x}{2}}^x (x - m) dm dl + \int_{\frac{x}{2}}^z \int_0^{\frac{x}{2}} (x - \frac{z}{2}) dm dl + \int_0^{\frac{x}{2}} \int_0^l (x - \frac{z}{2}) dm dl \right)$  which equals to  $x^3 + \frac{5}{8}z^3 - 3x^2z$ .

Hence, optimality requires:

$$\frac{\partial}{\partial z} \left( z^3 \left( x - 3\beta(z) + \frac{23}{32}z \right) \right) - (x^3 + \frac{5}{8}z^3 - 3x^2z) \Big|_{z=x} = 0$$

This differential equation will have a unique solution which is  $\beta(x) = \frac{3}{8}x$ . □

*Proof of Proposition 5.* We show that the no-resale equilibrium, which is bidding only one positive bid with

$$\beta^N(x) = \mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < x \right],$$

remains an equilibrium with resale.

Suppose that all bidders but bidder 1 are bidding according to above strategy.

If bidder 1 wins one additional unit for a bid  $b$ , he expects to sell it for

$$\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_{k-1}^{(n-1)} < \beta^{-1}(b) \right].$$

This is because, when bidder 1 wins two units, he knows that the highest losing value is  $k - 1$  out of  $n - 1$  opponents, and in a second price auction, he could sell to this person at the  $k - th$  highest of  $n - 1$ .

Now we claim that this is less than  $b$ :

$$\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_{k-1}^{(n-1)} < \beta^{-1}(b) \right] < \mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < \beta^{-1}(b) \right] = b$$

First, consider the deviation of the form  $(b, 0, \dots, 0)$ . For this deviation the expected selling price is

$$\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < \beta^{-1}(b) \right] = b.$$

Hence, resale cannot be profitable. Now consider the deviation  $(b_1, b_2, \dots, b_l, 0, \dots, 0)$ . Expected selling price per unit is going to be

$$\mathbb{E} \left[ Y_k^{(n-1)} \mid Y_{k-l}^{(n-1)} < \beta^{-1}(b_l) \right] < \mathbb{E} \left[ Y_k^{(n-1)} \mid Y_k^{(n-1)} < \beta^{-1}(b_l) \right] = b_l.$$

Hence, again, resale is not profitable (assuming all  $l$  is won, otherwise, change to  $l' < l$ ).

So for all deviation cases, deviations are not profitable.  $\square$

*Proof of Proposition 6.* We first look at the deviation where a bidder deviates for the first bid. Suppose the bidder with value  $x$  deviates to  $(\beta^R(z), \beta^R(x), \dots, \beta^R(x))$  where  $z > x$ . Note that if  $z > x > Y_1^{(n-1)}$  then since he just increases the expected payment to the auctioneer such a deviation is not profitable. Thus that deviation may increase the payoff only if  $z > Y_1^{(n-1)} > x > Y_k^{(n-1)}$ . In this case, if he did not deviate, his expected utility would be

$$\Pi(x, x) = \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)}) (x - \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}]). \quad (\text{C.1})$$

After deviation his expected utility becomes

$$\Pi(x, z) = \Pr(Y_1^{(n-1)} < z) (x - \beta^R(z)).$$

Since  $\Pr(Y_1^{(n-1)} < z) = F_1^{(n-1)}(z)$  the deviation expected utility becomes

$$\Pi(x, z) = F_1^{(n-1)}(z) \left( x - \frac{1}{F_1^{(n-1)}(z)} \int_0^z t f_k^{(n-1)}(t) dt \right) = F_1^{(n-1)}(z) x - F_k^{(n-1)}(z) z + \int_0^z F_k^{(n-1)}(t) dt \quad (\text{C.2})$$

For equation (C.1) we know that

$$\begin{aligned} F_k^{(n-1)}(t \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}) &= \Pr(Y_k^{(n-1)} < t \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}) \\ &= \frac{\Pr(Y_k^{(n-1)} < t)}{\Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)})} = \frac{F_k^{(n-1)}(t)}{\Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)})} \end{aligned}$$

and the expectation term is

$$\mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}] = \int_0^x t f_k^{(n-1)}(t \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}) dt.$$

After some tedious algebra (C.1) can be stated as:

$$\Pi(x, x) = x(\Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)}) - F_k^{(n-1)}(x)) + \int_0^x F_k^{(n-1)}(t) dt.$$

Then the profit of deviation becomes

$$\begin{aligned} \Pi(x, z) - \Pi(x, x) &= F_1^{(n-1)}(z)x - F_1^{(n-1)}(z)z + \int_0^z F_k^{(n-1)}(t) dt - \int_0^x F_k^{(n-1)}(t) dt \\ &= F_1^{(n-1)}(z)(x - z) + \int_x^z F_k^{(n-1)}(t) dt \\ &= F_k^{(n-1)}(z)(x - z) + \int_x^z F_k^{(n-1)}(t) dt < 0. \end{aligned}$$

Therefore such a deviation is not profitable.

In the previous deviation we increased the first bid while keeping the other bids the same. Now we do reverse:  $(\beta^R(z_1), \beta^R(z_2), \dots, \beta^R(z_k))$  where  $x > z_1 > z_2 > \dots > z_k$ . The first row of the following expected utility is for the case where  $z_k > Y_1^{(n-1)}$ . He wins  $k$  items in the bidding stage and sells  $k - 1$  items with a price  $\mathbb{E}[Y_k^{(n-1)} \mid Y_1^{(n-1)} < z_k] = \beta^R(z_k)$ . It is similar for the second row. The utility is the case for  $z_k < Y_1^{(n-1)} < z_{k-1}$ .  $k - 1$  items are won in the bidding stage and  $k - 2$  items will be sold in the resale stage with a price of  $\mathbb{E}[Y_k^{(n-1)} \mid Y_1^{(n-1)} < z_{k-1}] = \beta^R(z_{k-1})$ . Note that

$$-F_1^{(n-1)}(z_k) \left[ x - \sum_{j=1}^{k-1} \beta^R(z_j) + (k - 2)\beta^R(z_{k-1}) \right] + F_1^{(n-1)}(z_k) \left[ x - \sum_{j=1}^k \beta^R(z_j) + (k - 1)\beta^R(z_k) \right] \leq 0.$$

Let  $\mathbf{x} = [x, x, \dots, x]$  and  $\mathbf{z} = [z_1, z_2, \dots, z_k]$ .<sup>1</sup> The expected utility of deviation is

$$\begin{aligned}
\Pi(\mathbf{z}, x) &= F_1^{(n-1)}(z_k) \left[ x - \sum_{j=1}^k \beta^R(z_j) + (k-1)\beta^R(z_k) \right] \\
&\quad + \left( F_1^{(n-1)}(z_{k-1}) - F_1^{(n-1)}(z_k) \right) \left[ x - \sum_{j=1}^{k-1} \beta^R(z_j) + (k-2)\beta^R(z_{k-1}) \right] \\
&\quad + \dots \\
&\quad + \left( F_1^{(n-1)}(z_1) - F_1^{(n-1)}(z_2) \right) \left[ x - \beta^R(z_1) \right] \\
&\quad + \left( \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)}) (x - \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}]) \right) \\
&\leq F_1^{(n-1)}(z_1) \left[ x - \beta^R(z_1) \right] + \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)}) (x - \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}]) \\
&\leq F_1^{(n-1)}(x) \left[ x - \beta^R(x) \right] + \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)}) (x - \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}]) \\
&= \Pi(\mathbf{x}, x)
\end{aligned}$$

So such a deviation is not profitable. Therefore the  $k$ -tuple  $(\beta^R(x), \beta^R(x), \dots, \beta^R(x))$  is an equilibrium.  $\square$

*Proof of Proposition 7.* We will compare the expected payments of a bidder with value  $x$  in each equilibrium. If they are the same, the revenues will also be the same. The expected payment for a bidder for the first equilibrium is

$$\begin{aligned}
m^N(x) &= \Pr(Y_k^{(n-1)} < x) \mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x] = F_k^{(n-1)}(x) \frac{1}{F_k^{(n-1)}(x)} \int_0^x t f_k^{(n-1)}(t) dt \\
&= \int_0^x t f_k^{(n-1)}(t) dt = F_k^{(n-1)}(x) x - \int_0^x F_k^{(n-1)}(t) dt \\
&= x \left( \sum_{j=0}^{k-1} \binom{n-1}{j} F(x)^{n-1-j} (1-F(x))^j \right) - \int_0^x \left( \sum_{j=0}^{k-1} \binom{n-1}{j} F(z)^{n-1-j} (1-F(z))^j \right) dz
\end{aligned}$$

The next equation is the expected payment of the bidder if bidders follow the second equilibrium. The first summand is when the bidder wins the bidding stage and the second summand is

---

<sup>1</sup>Each  $j$ -th component represents the deviation value.

when the bidder loses but gets one item in the resale stage. So it becomes

$$\begin{aligned}
m^R(x) &= \Pr(Y_1^{(n-1)} < x)(k\beta^R(x) - (k-1)\mathbb{E}[Y_k^{(n-1)} \mid Y_1^{(n-1)} < x]) \\
&\quad + \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)})\mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}] \\
&= \Pr(Y_1^{(n-1)} < x)\beta^R(x) + \Pr(Y_k^{(n-1)} < x < Y_1^{(n-1)})\mathbb{E}[Y_k^{(n-1)} \mid Y_k^{(n-1)} < x < Y_1^{(n-1)}]. \quad (\text{C.3})
\end{aligned}$$

The cumulative distribution of  $Y_k^{(n-1)}$  under the condition  $Y_1^{(n-1)} < x$  is:

$$\begin{aligned}
F_k^{(n-1)}(z \mid Y_1^{(n-1)} < x) &= \sum_{j=0}^{k-1} \binom{n-1}{j} \left(\frac{F(z)}{F(x)}\right)^{n-1-j} \left(1 - \frac{F(z)}{F(x)}\right)^j \\
&= \frac{1}{F(x)^{n-1}} \sum_{j=0}^{k-1} \binom{n-1}{j} F(z)^{n-1-j} (F(x) - F(z))^j.
\end{aligned}$$

Note  $\beta^R(x) = \mathbb{E}[Y_k^{(n-1)} \mid Y_1^{(n-1)} < x] = \int_0^x z f_k^{(n-1)}(z \mid Y_1^{(n-1)} < x) dz = x - \frac{1}{F(x)^{n-1}} \int_0^x \left( \sum_{j=0}^{k-1} \binom{n-1}{j} F(z)^{n-1-j} (F(x) - F(z))^j \right) dz.$

Now the first summand of (C.3) becomes

$$\Pr(Y_1^{(n-1)} < x)\beta^R(x) = F(x)^{n-1}x - \int_0^x \left( \sum_{j=0}^{k-1} \binom{n-1}{j} F(z)^{n-1-j} (F(x) - F(z))^j \right) dz.$$

The second summand of (C.3) is:

$$\left( \sum_{j=1}^{k-1} \binom{n-1}{j} F(x)^{n-1-j} (1 - F(x))^j \right) x - \int_0^x \left( \sum_{j=1}^{k-1} \binom{n-1}{j} F(z)^{n-1-j} \left( (1 - F(z))^j - (F(x) - F(z))^j \right) \right) dz.$$

As a result, it is easy to see that  $m^R(x) = m^N(x)$ . Hence, the equilibria are revenue equivalent.  $\square$

*Proof of Proposition 8.* Consider a deviation  $(\beta(z), 0, \dots, 0)$ . For a bidder with value  $x \geq r_1^*$  the expected utility is

$$\Pi(x, z) = (x - \beta^{RR}(z)) \Pr(z > Y_k^{(n-1)}) = (x - \beta^{RR}(z)) F_k^{(n-1)}(z).$$

This should attain the maximum at  $z = x$  so

$$\frac{\partial \Pi(x, z)}{\partial z} \Big|_{z=x} = x f_k^{(n-1)}(x) - \left( \beta^{RR}(x) F_k^{(n-1)}(x) \right)' = 0.$$

Then

$$x f_k^{(n-1)}(x) = \left( \beta^{RR}(x) F_k^{(n-1)}(x) \right)'.$$

If we take integral:

$$\int_{r^*}^x y f_k^{(n-1)}(y) dy = \int_{r^*}^x \left( \beta^{RR}(y) F_k^{(n-1)}(y) \right)' dy$$

then we can see that

$$\beta^{RR}(x) = \mathbb{E}[\max\{Y_k^{(n-1)}, r_1^*\} \mid Y_k^{(n-1)} < x].$$

This is the unique candidate for an equilibrium under the condition that bidders bid only for the first item. Note that it is the standard equilibrium of a discriminatory auction without resale opportunity where the auctioneer is allowed to set a reserve price. Now consider a deviation  $(\beta^{RR}(x), b, 0, \dots, 0)$ . In this deviation the bidder gets an additional unit. He will sell the item to the  $k - 1$  highest value bidder with the price  $\max\{r_2^*, Y_k^{(n-1)}\}$ . So the expected return from resale is  $\mathbb{E}[\max\{r_2^*, Y_k^{(n-1)}\} \mid Y_{k-1}^{(n-1)} < (\beta^{RR})^{-1}(b)]$ . But the expected payment is  $\mathbb{E}[\max\{r_1^*, Y_k^{(n-1)}\} \mid Y_k^{(n-1)} < (\beta^{RR})^{-1}(b)] = b$ . And since  $r_2^* \leq r_1^{*2}$  and the event  $Y_{k-1}^{(n-1)} < (\beta^{RR})^{-1}(b)$  is less likely than to  $Y_k^{(n-1)} < (\beta^{RR})^{-1}(b)$  so

$$\mathbb{E}[\max\{r_2^*, Y_k^{(n-1)}\} \mid Y_{k-1}^{(n-1)} < (\beta^{RR})^{-1}(b)] \leq \mathbb{E}[\max\{r_1^*, Y_k^{(n-1)}\} \mid Y_k^{(n-1)} < (\beta^{RR})^{-1}(b)] = b.$$

Consider a general deviation  $(b_1, b_2, \dots, b_l, 0, \dots, 0)$ . He wins  $l$  items sells  $l - 1$  items. The expected

---

<sup>2</sup>In the bidding stage, the auctioneer chooses the optimal reserve price from the interval  $[0, 1]$  for  $n$  bidders. In the resale stage, the auctioneer (or the winner of the bidding stage whose value is  $x$ ) chooses the optimal reserve price from the interval  $[0, x]$  for  $n - k + 1$  bidders. Since  $1 \geq x$ , it is easy to see  $r_1^* \geq r_2^*$ .



return from resale is :

$$\mathcal{R} = \left( \begin{array}{c} r_2^* \Pr((\beta^{RR})^{-1}(b_l) > Y_{k-l+1}^{(n-1)} > r_2^* > Y_{k-l+2}^{(n-1)}) \\ + 2r_2^* \Pr((\beta^{RR})^{-1}(b_l) > Y_{k-l+1}^{(n-1)} > Y_{k-l+2}^{(n-1)} > r_2^* > Y_{k-l+3}^{(n-1)}) \\ + \dots + (l-1)r_2^* \Pr((\beta^{RR})^{-1}(b_l) > Y_{k-l+1}^{(n-1)} > \dots > Y_{k-1}^{(n-1)} > r_2^* > Y_k^{(n-1)}) \\ + \left( (l-1)\mathbb{E}[Y_k^{(n-1)} \mid (\beta^{RR})^{-1}(b_l) > Y_k^{(n-1)} > r_2^*] \right) \end{array} \right).$$

Note that  $\mathcal{R} \leq (l-1)\mathbb{E}[\max\{r_2^*, Y_k^{(n-1)}\} \mid r_2^* < Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)]$ . On the other hand, the expected payment for  $l-1$  items is:  $(l-1)\mathbb{E}[\max\{r_1^*, Y_k^{(n-1)}\} \mid Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)]$ . Since  $r_2^* \leq r_1^*$  and the event  $(\beta^{RR})^{-1}(b_l) > Y_k^{(n-1)} > r_2^*$  is less likely comparing to  $Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)$  we have

$$\mathbb{E}[\max\{r_2^*, Y_k^{(n-1)}\} \mid r_2^* < Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)] \leq \mathbb{E}[\max\{r_1^*, Y_k^{(n-1)}\} \mid Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)].$$

Hence  $\mathcal{R} \leq (l-1)\mathbb{E}[\max\{r_1^*, Y_k^{(n-1)}\} \mid Y_k^{(n-1)} < (\beta^{RR})^{-1}(b_l)]$ . So such deviation is not profitable.

Therefore  $(\beta^{RR}(x), 0, \dots, 0)$  is an equilibrium.  $\square$



# Bibliography

Acemoğlu, Daron, 2002, Directed technical change, *The Review of Economic Studies* 69, 781–809.

Acemoğlu, Daron, and David Autor, 2011, Skills, tasks and technologies: Implications for employment and earnings, in David Card, and Orley Ashenfelter, eds., *Handbook of Labor Economics*, volume 4B, 1043–1171 (Elsevier Press).

Aguiar, Mark, and Erik Hurst, 2007, Measuring Trends in Leisure: The Allocation of Time Over Five Decades, *The Quarterly Journal of Economics* 122, 969 – 1006.

Ales, Laurence, Musab Kurnaz, and Christopher Sleet, 2015, Technical Change, Wage Inequality, and Taxes, *American Economic Review* 105, 3061 – 3101.

Athey, Susan, 2001, Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information, *Econometrica* 69, 861–889.

Ausubel, Lawrence M., Peter Cramton, Marek Pycia, Marzena Rostek, and Marek Weretka, 2014, Demand Reduction and Inefficiency in Multi-Unit Auctions, *The Review of Economic Studies* Forthcoming.

Autor, D., and D. Dorn, 2013, The growth of low-skill service jobs and the polarization of the US labor market, *American Economic Review* 103, 1553–1597.

Autor, D., L. Katz, and M. Kearney, 2006, The polarization of the US labor market, *American Economic Review* 96, 189–194.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger, 1998, Computing inequality: Have computers changed the labor market?, *Quarterly Journal of Economics* 113, 1169–1213.

- Autor, David H, Frank Levy, and Richard J Murnane, 2003, The skill content of recent technological change: An empirical exploration, *Quarterly Journal of Economics* 118, 1279–1333.
- Baron, David P., and Roger B. Myerson, 1982, Regulating a Monopolist with Unknown Costs, *Econometrica* 50, 911 – 930.
- Blau, David, 2003, Child Care Subsidy Programs, in *Means-Tested Transfer Programs in the United States*, 443 – 516 (University of Chicago Press).
- Blundell, Richard, and Thomas Macurdy, 1999, Labor supply: A Review of Alternative Approaches, in O. Ashenfelter, and D. Card, eds., *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 27, 1559 – 1695 (Elsevier).
- Blundell, Richard, Costas Meghir, and Pedro Neves, 1993, Labour Supply and Intertemporal Substitution, *Journal of Econometrics* 59, 137 – 160.
- Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten, 2012, Consumption Inequality and Family Labor Supply, Working Paper 18445, National Bureau of Economic Research.
- Boca, Daniela Del, Christopher Flinn, and Matthew Wiswall, 2013, Household Choice and Child Development, *The Review of Economic Studies* 81, 137 – 185.
- Bonnans, J., and A. Shapiro, 2000, *Perturbation Analysis of Optimization Problems* (Springer-Verlag: Springer Series in Operations Research).
- Bound, John, Charles Brown, Greg J Duncan, and Willard L Rodgers, 1994, Evidence on the validity of cross-sectional and longitudinal labor market data, *Journal of Labor Economics* 12, 345–368.
- Bound, John, and Alan B Krueger, 1991, The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?, *Journal of Labor Economics* 9, 1–24.
- Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt, 2002, Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence, *The Quarterly Journal of Economics* 117, 339–376.
- Buera, F., and J. Kaboski, 2012, The rise of the service economy, *American Economic Review* 102, 2540–2569.

- Bukhchandani, Sushil, and Chi-fu Huang, 1989, Auctions with Resale Markets: An Exploratory Model of Treasury Bill Markets, *Review of Financial Studies* 2, 311–339.
- Cheng, Harrison, 2011, Auctions with Resale and Bargaining Power, *Journal of Mathematical Economics* 47, 300 – 308.
- Cheng, HarrisonH., and Guofu Tan, 2010, Asymmetric Common-value Auctions with Applications to Private-value Auctions with Resale, *Economic Theory* 45, 253–290.
- Chetty, Raj, 2012, Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, *Econometrica* 80, 969 – 1018.
- Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber, 2011, Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins, *The American Economic Review* 101, 471–475.
- Costinot, A., and J. Vogel, 2010, Matching and inequality in the world economy, *Journal of Political Economy* 118, 747–786.
- Docquier, Frederic, 2004, Income Distribution, Non-convexities and the Fertility - Income Relationship, *Economica* 71, 261 – 273.
- Engelbrecht-Wiggans, Richard, and Charles M. Kahn, 1998a, Multi-unit Auctions with Uniform Prices, *Economic Theory* 12, 227–258.
- Engelbrecht-Wiggans, Richard, and Charles M. Kahn, 1998b, Multi-Unit Pay-Your-Bid Auctions with Variable Awards, *Games and Economic Behavior* 23, 25 – 42.
- Filiz-Ozbay, Emel, Kristian Lopez-Vargas, and Erkut Y. Ozbay, 2015, Multi-object auctions with resale: Theory and experiment, *Games and Economic Behavior* 89, 1 – 16.
- French, Eric, 2005, The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour, *The Review of Economic Studies* 72, 395 – 427.
- Garratt, Rod, and Thomas Troger, 2006, Speculation in Standard Auctions with Resale, *Econometrica* 74, pp. 753–769.

- Garratt, Rodney J., Thomas Troger, and Charles Z. Zheng, 2009, Collusion via Resale, *Econometrica* 77, 1095–1136.
- Godbey, Geoffrey, and John P. Robinson, 1999, *Time for Life: The Surprising Ways Americans Use their Time* (The Pennsylvania State University Press).
- Goldin, Claudia, and Lawrence F Katz, 1998, The origins of technology-skill complementarity, *The Quarterly Journal of Economics* 113, 693–732.
- Golosov, Mikhail, Larry E. Jones, and Michele Tertilt, 2007, Efficiency with Endogenous Population Growth, *Econometrica* 75, 1039 – 1071.
- Goos, Maarten, and Alan Manning, 2007, Lousy and lovely jobs: The rising polarization of work in Britain, *Review of Economics and Statistics* 89, 118–133.
- Grossman, E., and E. Rossi-Hansberg, 2008, Trading tasks: A simple theory of offshoring, *American Economic Review* 98, 1978–1997.
- Gupta, Madhurima, and Bernard Lebrun, 1999, First Price Auctions with Resale, *Economics Letters* 64, 181 – 185.
- Guryan, Jonathan, Erik Hurst, and Melissa Kearney, 2008, Parental Education and Parental Time with Children, *Journal of Economic Perspectives* 22, 23–46.
- Hafalir, Isa, and Vijay Krishna, 2008, Asymmetric Auctions with Resale, *American Economic Review* 98, 87–112.
- Hafalir, Isa, and Vijay Krishna, 2009, Revenue and Efficiency Effects of Resale in First-price Auctions, *Journal of Mathematical Economics* 45, 589 – 602.
- Haile, Philip A, 2000, Partial Pooling at the Reserve Price in Auctions with Resale Opportunities, *Games and Economic Behavior* 33, 231 – 248.
- Haile, Philip A., 2001, Auctions with Resale Markets: An Application to U.S. Forest Service Timber Sales, *American Economic Review* 91, 399–427.
- Haile, Philip A., 2003, Auctions with Private Uncertainty and Resale Opportunities, *Journal of Economic Theory* 108, 72 – 110.

- Haveman, Robert, and Barbara Wolfe, 1995, The Determinants of Children's Attainments: A Review of Methods and Findings, *Journal of Economic Literature* 33, 1829 – 1878.
- Heathcote, J., K. Storesletten, and G. Violante, 2014, Optimal tax progressivity: An analytical framework, Staff Report 496. FRB Minneapolis.
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante, 2010a, Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006, *Review of Economic Dynamics* 13, 15–51.
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante, 2010b, Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006, *Review of Economic Dynamics* 13, 15 – 51.
- Hellwig, Martin, 2007, The undesirability of randomized income taxation under decreasing risk aversion, *Journal of Public Economics* 91, 791–806.
- Hosseini, Roozbeh, Larry E. Jones, and Ali Shourideh, 2013, Optimal Contracting with Dynastic Altruism: Consumption and Family Size, *Journal of Economic Theory* 148, 1806 – 1840.
- Hotz, V. Joseph, and John Karl Scholz, 2003, The Earned Income Tax Credit, in *Means-Tested Transfer Programs in the United States*, 141 – 198 (University of Chicago Press).
- Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden, 2013, The Optimal Marginal Tax Rates with both Extensive and Intensive Responses, *Journal of Economic Theory* 148, 1770 – 1805.
- Jones, Larry E., Alice Schoonbroodt, and Michele Tertilt, 2010, Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?, in John B. Shoven, ed., *Demography and the Economy*, 43–100 (University of Chicago Press).
- Jones, Larry E., and Michele Tertilt, 2008, An Economic History of Fertility in the U.S.: 1826 - 1960, in Peter Rupert, ed., *Frontiers of Family Economics*, 165 – 230 (Emerald Press, Bingley, UK).
- Kaplan, Steven N, and Joshua Rauh, 2013, It's the market: The broad-based rise in the return to top talent, *The Journal of Economic Perspectives* 35–55.

- Keane, Michael P., 2011, Labor Supply and Taxes: A Survey, *Journal of Economic Literature* 49, 961 – 1075.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, 2009, The Optimal Income Taxation of Couples, *Econometrica* 77, 537 – 560.
- Krishna, Vijay, 2002, *Auction Theory* (San Diego: Elsevier Science, Academic Press).
- Krusell, P., L.E. Ohanian, J.V. Ríos-Rull, and G.L. Violante, 2000, Capital-skill complementarity and inequality: A macroeconomic analysis, *Econometrica* 68, 1029–1053.
- Lane, Julia I, Laurie A Salmon, and James R Spletzer, 2007, Establishment wage differentials, *Monthly Labor Review* 130, 3–17.
- Lebrun, Bernard, 2010, First-price Auctions with Resale and with Outcomes Robust to Bid Disclosure, *The RAND Journal of Economics* 41, 165–178.
- Lino, Mark, 2014, Expenditures on Children by Families, 2013. U.S. Department of Agriculture, Center for Nutrition Policy and Promotion. Miscellaneous Publication No. 1528-2013.
- Lockwood, Benjamin B, Charles G Nathanson, and E Glen Weyl, 2014, Taxation and the allocation of talent, Technical report, Mimeo, Harvard University and University of Chicago.
- Mankiw, N. Gregory, and Matthew Weinzierl, 2010, The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution, *American Economic Journal: Economic Policy* 2, 155 – 176.
- McAdams, David, 2003, Isotone Equilibrium in Games of Incomplete Information, *Econometrica* 71, 1191–1214.
- McDonald, James B, 1984, Some generalized functions for the size distribution of income, *Econometrica* 52, 647–663.
- McFadden, Daniel, Peter Diamond, and Miguel Rodriguez, 1978, Measurement of the elasticity of factor substitution and the bias of technical change, in Melvyn Fuss, and Daniel McFadden, eds., *Production Economics: A Dual Approach to Theory and Applications* (Amsterdam: Elsevier).
- Milgrom, P., and C. Shannon, 1994, Monotone comparative statics, *Econometrica* 62, 157–180.



- Mirrlees, James, 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* 38, 175–208.
- Mouw, Ted, and Arne. Kalleberg, 2010, Occupations and the structure of wage inequality in the united states, 1980s to 2000s, *American Sociological Review* 75, 402–431.
- Noussair, Charles, 1995, Equilibria in a Multi-Object Uniform Price Sealed Bid Auction with Multi-Unit Demands, *Economic Theory* 5, pp. 337–351.
- Pagnozzi, Marco, 2007, Bidding to Lose? Auctions with Resale, *The RAND Journal of Economics* 38, pp. 1090–1112.
- Pagnozzi, Marco, 2010, Are Speculators Unwelcome in Multi-object Auctions?, *American Economic Journal: Microeconomics* 2, 97–131.
- Pagnozzi, Marco, and Krista Jabs Saral, 2013, Multi-object Auctions with Resale: an Experimental Analysis, Working paper.
- Patterson, Michael A, and Anil V Rao, 2013, GPOPS-II: A MATLAB software for solving multiple-phase optimal control problems using hp-adaptive Gaussian quadrature collocation methods and sparse nonlinear programming, Working Paper.
- Pencavel, John, 1986, Labor supply of Men: A Survey, in O. Ashenfelter, and R. Layard, eds., *Handbook of Labor Economics*, volume 1 of *Handbook of Labor Economics*, chapter 1, 3 – 102 (Elsevier).
- Reny, Philip J., 1999, On the Existence of Pure and Mixed Strategy Nash Equilibria in Discontinuous Games, *Econometrica* 67, pp. 1029–1056.
- Reny, Philip J., 2011, On the Existence of Monotone Pure-Strategy Equilibria in Bayesian Games, *Econometrica* 79, 499–553.
- Rochet, Jean-Charles, and Philippe Chone, 1998, Ironing, Sweeping, and Multidimensional Screening, *Econometrica* 66, 783 – 826.
- Rothschild, C., and L. Chen, 2014, Screening with endogenous preferences, Unpublished Paper. Wellesley College.

- Rothschild, C., and F. Scheuer, 2014, A theory of income taxation under multi-dimensional skill heterogeneity, NBER WP 19822.
- Rothschild, Casey, and Florian Scheuer, 2013, Redistributive taxation in the Roy model, *Quarterly Journal of Economics* 128, 623–662.
- Roy, A., 1950, The distribution of earnings and of individual output, *Economic Journal* 60, 489–501.
- Sadka, Efraim, 1976, On Income Distribution, Incentive Effects and Optimal Income Taxation, *The Review of Economic Studies* 42, 261 – 268.
- Saez, E., 2001, Using elasticities to derive optimal income tax rates, *Review of Economic Studies* 68, 205–229.
- Saez, Emmanuel, 2002, Optimal income transfer programs: Intensive versus extensive labor supply responses, *Quarterly Journal of Economics* 117, 1039–1073.
- Salanié, B., 2011, *Economics of Taxation* (MIT Press, USA).
- Sattinger, M., 1975, Comparative advantage and the distributions of earnings and abilities, *Econometrica* 43, 455–468.
- Schultz, T. Paul, 1986, The Value and Allocation of Time in High-Income Countries: Implications for Fertility, *Population and Development Review* 12, 87 – 108.
- Seade, Jesus K., 1977, On the Shape of Optimal Tax Schedules, *Journal of Public Economics* 7, 203 – 236.
- Singh, S., and G. Maddala, 1976, A function for size distribution of incomes, *Econometrica* 44, 976–983.
- Slavík, C., and H. Yazıcı, 2014, Machines, buildings and optimal dynamic taxes, *Journal of Monetary Economics* 66, 47–61.
- Stiglitz, Joseph, 1982, Self selection and Pareto efficient taxation, *Journal of Public Economics* 17, 213–240.

- Teulings, C., 1995, The wage distribution in a model of the assignment of skills to jobs, *Journal of Political Economy* 103, 280–315.
- Virag, Gabor, 2013, First-price Auctions with Resale: The Case of Many Bidders, *Economic Theory* 52, 129–163.
- Weisberg, H., 1992, *Central tendency and variability* ((Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–083.) Newbury Park, CA: Sage).
- Xu, Xiaoshu, Dan Levin, and Lixin Ye, 2013, Auctions with Entry and Resale, *Games and Economic Behavior* 79, 92 – 105.
- Zheng, Charles Z., 2014, Existence of Monotone Equilibria in First-price Auctions with Resale, Working paper.
- Ziliak, J. P., and T. J. Kniesner, 2005, The Effect of Income Taxation on Consumption and Labor Supply, *Journal of Labor Economics* 23, 769 – 796.