OPTIMAL TAXATION OF FAMILIES: Mirrlees meets Becker*

Musab Kurnaz†

Tepper School of Business, Carnegie Mellon University

This Version: April 29, 2016

Abstract

This paper examines the optimal taxation of families in an environment in which (i) families’ earning abilities and tastes for children are families’ private information, and (ii) child-rearing requires both parental time and goods. Potential parents simultaneously decide labor income and number of children and a government uses information on family income and size to construct an optimal tax system: a combination of an income tax schedule with child tax credits. The optimal child tax credits are distinctly affected by the parental time and cost of goods involved in child-rearing. In the quantitative part, I calibrate my model to the US data and show that child-rearing costs translate into a pattern of optimal credits that is U-shaped in income. In particular, the credits to families are decreasing over the first half of the income distribution. In addition, the credits are decreasing by family size owing to economies of scale in the impact of child-rearing costs on family welfare. For median-income families, the credit for the second (third) child equals 28% (2%) of the credit for the first (second) child.

JEL-Classification: H21, H53, D82, J13

Keywords: Optimal Income Taxation, Child Tax Credits, Multidimensional Screening, Fertility

*I am indebted Christopher Sleet, Laurence Ales, and Isa Hafalir for invaluable advice, guidance, and encouragement throughout this project. I also thank Kjetil Storesletten, Marla Ripoll, and Randall Wright for very useful comments. I also acknowledge seminar participants in Carnegie Mellon University, Koc University, Bilgi University, Bilkent University, Istanbul Technical University, Kadir Has University, TOBB-ETU, and Ozyegin University.

†Musab Kurnaz: mkurnaz@andrew.cmu.edu
1 Introduction

The tax treatment of families is a debating point in the US throughout decades. Policy regarding with income tax schedule has frequently changed and the changes favored families with children. For example, the federal spending for Earned Income Credits, which is mostly attained by families with children, has increased by 140% and for Child Tax Credits has been doubled up in a decade. Consequently, the federal budget devoted to child tax credits is 13% of mean-tested transfers in 2011. Along the same line, although there is vast positive literature focuses on the effect of policy changes, there is almost no normative work to guide governments how to design child tax credits. My paper fills this gap. In this paper, I answer a very important policy question: What should be the optimal 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. On the one hand, the goods cost of child-rearing reduces the welfare of low-earning families more relative to the high. The reduction enhances the redistribution motives of the government towards low-earning families and consequently more credits are provided for lower-earning families. On the other, the opportunity cost of childcare gets heavier through high-earning families and makes them reluctant to work harder and creates an efficiency loss in total output. To alleviate the disincentive effect, the government has a motive for more provision to the wealthier. When I consider both costs of child-rearing, I show that the optimal child tax credits are U-shaped with respect to income.

This paper analyzes the optimal design of income taxes of families with different sizes and particularly focuses on child tax credits. I embed a Becker model into an optimal tax framework. In contrast to many optimal tax theory papers which focus on individual workers, I study a family problem in which potential parents do not only make labor choice but also fertility decision. Potential parents are characterized by their (earning) abilities and tastes for children which are only observable to themselves. A higher ability increases the labor productivity and a higher taste increases the desire to have larger family. Facing a problem of asymmetric information, a redistributive government max-

---

1Becker (1960) and Becker (1965) suggest child should appear as a good in family utility and should be considered as an output of goods and time.

2The main stream in the literature studies characterization of workers’ abilities. See Mirrlees (1971).

3Tastes for children can be considered as a measure of child desire of households based on their exoge-
imizes a social welfare by choosing labor income taxes and child tax credits. Optimal 
taxes of families are characterized by a formula which links marginal income tax rates 
to the sensitivity of family income to taxes, the exogenous ability distribution, and the 
redistributive motives of the government. An important contribution of this paper is that 
the tax formula has two dimensions in contrast to the literature. This novelty allows me 
to shed light on the design of child tax credits.

Characteristics of families are twofold which creates a two dimensional screening 
problem. The literature on optimal taxation dealing with multidimensional screening 
is sparse owing to the technical difficulties. My paper is one of exceptions. I cope with 
two dimensional screening problem by disentangling the characteristics in family utility 
and therefore their effects on the family choices. The separability assumption ensures 
that the marginal benefit of a child is independent of parents’ ability and the marginal 
cost of a child is independent of parents’ taste. This independence allows me to define 
critical tastes, the tastes which makes family indifferent between two consecutive num-
bors of children given an ability level. As a result, parents whose tastes are between \( n^{th} \) 
and \( n + 1^{th} \) critical tastes choose to have \( n \) children. This optimality condition allows to 
screen parents’ tastes by the critical taste definition. On the other hand, screening abili-
ties is conventionally handled by a first-order approach, i.e. \( \theta \) ability parents distinguish 
theremselves from other parents whose abilities are in the neighborhood of \( \theta \).

I calibrate my model to the US economy to analyze the optimal tax system quantita-
tively. First, I estimate the cost of child-rearing costs. The opportunity cost is estimated 
as the ratio of average forgone labor of parents owing to child care to the average labor 
of childless families. Besides, the goods cost is estimated using the information on the 
annual report of US Department of Agriculture. Second, using these estimates, I calculate 
families’ earning abilities from consumption-leisure margin of families. The margin is 
affected by taxes and the information on taxes is taken from the March release of the Cur-
rent Population Survey (CPS). CPS also provide weights of families which allows me to 
derive the earning ability distribution. Third, I assume a particular distribution for child 
tastes and use maximum likelihood estimation to capture the parameter determines the 
marginal benefit of children.

The quantitative analysis provides two important policy results which are sharply 
different from the US child tax credit schedule. First, optimal tax credits for each child 
is U-shaped according to income which is a result of two types of child-rearing costs.

4The conventional tax formula in the literature is one dimensional because the many papers focus solely 
on individual workers who intensively decide labor.

nous characteristics, such as religion and race. See Becker (1960).
On the one hand, the goods cost decreases the welfare of low-income families relatively more than the high. On the other, the burden of time cost is more intense for high-income families than the low. These costs push the child credits up for low and high-income families, respectively. As a result, the optimal credits are U-shaped in income. The credits are decreasing in the first half of the income distribution and increases in the rest. In contrast, the US child tax credits are declining according to family income. This has an implication of that the US government focuses on the impact of goods cost. However, the time cost significantly affects families’ labor choice and reduces the total output of the economy.\textsuperscript{5} Including the effect of time cost into construction of credits may improve social welfare.

Second, I show that the optimal credits are not same for each child in a family because of economies of scale in the impact of child-rearing on family welfare. Families have a convex preference over welfare and the impact of first (second) child on family welfare has relatively larger than the impact of second (third) child. In addition, there is an economies of scale in child-rearing costs. On the one hand, time cost of one child is 2.29 market labor hours per week while it is 4.16 hours for two children and 4.87 hours for three children. On the other, goods cost of two (three) children is 58% (83%) more than the goods cost of one child. The scale strengthens the impact of one child relative to the second on family welfare owing to the convexity in preference. The government values these impacts to design its redistribution motives. As a result, credits for each child differ for all family income levels. In particular, the credit for the second (third) child is 28% (2%) of the first (second) child credit for the median income families. In contrast, the child tax credits in US are almost similar for each child according to family income.

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 2. I introduce the model in Section 3 in which I also derive the two dimensional optimal tax schedule. Section 4 quantitatively analyzes the model. I check the robustness of results in Section 5 and conclude with Section 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 voluminous literature on fertility theories starting with Becker (1960) and Becker (1965). These two seminal works suggest that child should appear in family utility

\textsuperscript{5}The opportunity cost of childcare is 1.8\% of GDP in 1992 according to Haveman and Wolfe (1995).
as a production good of inputs of goods and time. More recently, Haveman and Wolfe (1995) estimate goods and the time costs of child-rearing. They find that child-rearing costs is 14.5% of 1992 GDP and two-thirds of the costs is spent by parents while the rest is financed by the government. This large share shows the importance of child-rearing costs on the US economy.

The goods cost is 82% of parental cost which includes expenditures on food, housing, health care, and clothes. The literature studies the impact of the goods cost is vast. Recent examples are Golosov, Jones, and Tertilt (2007) and Hosseini, Jones, and Shourideh (2013). The former paper studies the efficiency of the future allocations and the latter focuses on the consumption inequality in the long run. On the other hand, the time cost is the crucial ingredient of well known empirical evidence of negative correlation between fertility and family labor income (see Jones, Schoonbroodt, and Tertilt (2010)). This is mainly because the opportunity cost of childcare is higher for the high-wage workers. In addition, time cost increases the labor sensitivity of parents to the wage changes. 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 Frisch elasticity, in particular, is important for my work because it is one of the major components of the optimal tax system (see Saez (2001)).

In contrast to many works in fertility literature that focuses on one type of child-rearing costs, I study with both type of costs and show that their interaction with family income is important for optimal policies.

My paper also contributes to the public finance literature, which is founded on the trade-off between efficiency and equity beginning with Mirrlees (1971). The trade-off arises because there is a friction in the information. Workers’ abilities are only observable to workers’ themselves in many papers of this literature. In my paper, however, not only abilities but also tastes for children are families’ private information, and hence the friction is two dimensional. Because of the technical difficulties of multi screening problems, there are few works study such an environment. 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 earners’ earning abilities and secondary earners’ work costs 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

6Baron and Myerson (1982) and Rochet and Chone (1998) provide some additional requirements to solve a multi screening problem in the industrial organization literature.
studies an 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 households. In this paper, I derive optimal taxes for an arbitrary number of family sizes.

2 Children in the US Income Tax Code

I first document the changes in federal spending of the most important US government policies to motivate my analysis. Over 100 programs, 28% of the federal budget for welfare programs is spent for children and 33% of this expense is related to tax credits and exemptions such as Child Tax Credit (CTC), Child and Dependent Care Tax Credit (CDCTC), Dependent Exemptions, and Earned Income Tax Credit (EITC).\(^7\) I refer Appendix A.1 for detailed information of these programs. Here, I particularly focus on the drastic changes in federal budget for CTC and EITC. The federal spending for CTC and EITC has been increased by 210% and 140%, respectively (see Figure 1). The increase in CTC budget is because of the increase in the credit amounts and relaxation of the eligibility conditions. The increase in EITC budget mostly depends on the expansions of the program during 2000s.

The changes in the federal budgets for tax credits (especially changes in EITC spending) is the focus of positive literature (see Hotz and Scholz (2003) and references there). However, the normative analysis is very sparse. My paper is one of the first papers to study the design of the tax credits. In the following section, I introduce a static model, in which heterogeneous potential parents decide labor choice and number of children by observing their own labor productivities and desire for children. Raising children requires both monetary expenses (goods) and parental time. The effects of these costs of family welfare are the determinants of the design of the optimal child tax credits.

3 Model

A continuum of potential parents (hereafter families), where the size is normalized to 1, have identical preferences over consumption \(c \in \mathbb{R}_+\), labor income \(z \in [0, \bar{z}]\), and number of children \(n \in \mathcal{N} = \{0, 1, \ldots, N\}\) described by a utility function \(U : \mathbb{R}_+ \times [0, \bar{z}] \times \mathcal{N} \rightarrow \mathbb{R}_+\).\(^7\) Refer to http://www.urban.org/sites/default/files/alfresco/publication-pdfs/412599-Data-Appendix-to-Kids-Share-.PDF
\( U(\cdot, \cdot, n) \) is assumed to be concave, twice differentiable on the interior its domain, with for each \( z \in [0, Z] \) and for all \( n \in \mathcal{N}, U(\cdot, z, n) \) increasing and for each \( c \in \mathbb{R}_+ \) and for all \( n \in \mathcal{N}, U(c, \cdot, n) \) decreasing and strictly concave. First and second partial derivatives of \( U \) are denoted \( U_x(\cdot, \cdot, n) \) and \( U_{xy}(\cdot, \cdot, n) \) with \( x, y \in \{c, z\} \). \( U \) satisfies the Inada conditions: for all \( c \in \mathbb{R}^+ \) and for all \( n \in \mathcal{N}, \lim_{z \downarrow 0} U_z(c, \cdot, n) = 0 \) and \( \lim_{z \uparrow Z} U_z(c, \cdot, n) = -\infty \). In addition, \( U \) satisfies the Spence-Mirrlees single crossing property: \( \frac{\partial^2 U_z(c, \cdot, n)}{\partial z \partial c} \) is decreasing in \( \theta \).

The characteristics of families are twofold. Each family has an (earning) ability \( \theta \in \Theta = (\bar{\theta}, \bar{\theta}) \) and a (child) taste: \( \beta \in B = (\bar{\beta}, \bar{\beta}) \) in the population. The family characteristics \( \gamma = (\beta, \theta) \) are distributed according to a continuous density distribution over \( \Gamma = B \times \Theta \). Let \( \Pi(\gamma) \) 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 \( \theta \). Family characteristics, \( \gamma \), are families’ private information and two dimensional unlike the most of public finance papers starting with Mirrlees (1971) studies one dimensional private information.

Following Becker (1965), \( n \) children are output of exogenous inputs of \( e_n \) amount of goods (expenditures) and \( b_n \) amount of parental working time. This assumption justifies an equality criterion for children across families in the perspective of the government. In addition, the costs can be interpreted as the minimum input to produce \( n \) children. Families are allowed to spend more goods and time, but the extra of the minimum input

---

8This assumption is constructed for moral reasons that the government equally cares every children.
is a part of family consumption and leisure, respectively.

The government collects taxes according to observable family choices: family income $z$ and number of children $n$. The tax system is nonlinear in income: $T(z, n)$. For notational sake, I denote $T(z, n)$ as $T_n(z)$. This system can be transformed into the US income tax format if every family pay income taxes according to their income $T_0(z)$ and get credits depending on the number of children $\sum_{j=1}^n k_j(z)$ where $k_j(z) = T_{j-1}(z) - T_j(z)$ is the child tax credit for the $j^{th}$ children. Hence, $T_n(z) = T_0(z) - \sum_{j=1}^n k_j(z)$.

Observing their characteristics, child-rearing costs, and taxes, a $\gamma$ family solves:

$$\max_{c,z,n} \mathcal{U}(c, z, n) \quad \text{s.t.} \quad c \leq z - T_n(z) - e_n$$

where its utility function is:

$$\mathcal{U}(c, z, n) = \Psi \left( u(c) - \theta h \left( \frac{z}{\theta} + b_n \right) + m(n, \beta) \right)$$

where $h$ is an increasing and convex function of class $C^2$ and normalized so that $h(0) = 0$ and $h'(1) = 1$ (see Kleven, Kreiner, and Saez (2009)) and $m(n, \beta)$ represents the utility of having $n$ children and concave in $n$ which is modeled according to Becker (1960)'s suggestion that children should be considered as consumption goods. The utility is determined by the number of children and families' taste over children. The taste parameter may depend on family's religion, race, etc.

Note that child choice is discrete, and hence first-order conditions are not immediately applicable. Therefore, I solve the family problem in two steps. Initially, families determine how many children to have. Second, consumption and income are chosen given the number of children.

### 3.1 Family Problem

I use backward induction: Given $n$, the optimal consumption, $c_n$, and optimal income, $z_n$, should satisfy the first-order condition and the budget set:

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

$$c_n = z_n - T_n(z_n) - e_n.$$
These equations imply $c_n$ and $z_n$ are independent from $\beta$ but $b_n$ and $e_n$. Next, I define the indirect utility of $n-$child families:

$$V_n(\theta) := u(c_n) - \theta h\left(\frac{z_n}{\theta} + b_n\right).$$

(5)

Note that child rearing costs are captured by $V_n(\theta)$. Therefore the marginal cost of $n$ children equals $V_{n-1}(\theta) - V_n(\theta)$. The marginal benefit of $n$ children is $M(n, \beta) := m(n, \beta) - m(n - 1, \beta)$. The $\gamma$ family 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, $\gamma$ families decide to have $n$ children if and only if $\beta \in (\beta_n(\theta), \beta_{n+1}(\theta))$, where

$$\beta_n(\theta) := M^{-1}(V_{n-1}(\theta) - V_n(\theta)), \quad (6)$$

for $n = 1, 2, \ldots, N$.

In Figure 2, I illustrate the child choice for $n = 0, 1$. For a particular ability level, the families with $\beta \in (\underline{\beta}, \beta_1(\theta))$ choose to have no children because the marginal benefit of one child is less than its costs. $\beta_1(\theta)$ taste level equalizes the benefit and cost having one child. The families with $\beta \in (\beta_1(\theta), \beta_2(\theta))$ 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 2: Critical Child Taste Levels](image)

Next, I define tax equilibrium.

**Definition 1.** Let $G \in \mathbb{R}_+$ be a fixed public spending amount. Given $G$, a tax equilib-
rium is a tax system $T : \mathbb{R}_+ \times \mathcal{N} \to \mathbb{R}$, and allocation $\{c(\gamma), z(\gamma), n(\gamma)\}_{\gamma \in \Gamma}$ such that $(c(\gamma), z(\gamma), n(\gamma))$ solves (1) and $G \leq \int_{\Gamma} [z(\gamma) - c(\gamma) - e_n(\gamma)]d\Pi(\gamma)$. Let $\mathcal{T}$ be the set of tax equilibria (given $G$), which I take to be nonempty.

### 3.2 The Government’s Problem

A government attaches Pareto weight $\xi(\gamma)$ to families with type $\gamma$ with weights normalized to satisfy $\int_{\Gamma} \xi(\gamma)d\gamma = 1$. The government selects a tax equilibrium to solve:

$$\max_{\mathcal{T}} \int_{\Gamma} \xi(\gamma) \mathcal{U}((c(\gamma), z(\gamma), n(\gamma)) \ d\Pi(\gamma).$$

Let $T^*$ and $\{c^*(\gamma), z^*(\gamma), n^*(\gamma)\}_{\gamma \in \Gamma}$ denote an optimal tax equilibrium. Optimal marginal tax rate of families with $n$ children is defined as:

$$\frac{\partial T^*_n(z_n)}{\partial z} = 1 + \frac{U_c(c^*(\gamma), z^*(\gamma), n^*(\gamma))}{U_c(c^*(\gamma), z^*(\gamma), n^*(\gamma))}$$

for all $n \in \mathcal{N}$.

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

$$\max_{\{c(\gamma), z(\gamma), n(\gamma)\} \in \mathbb{R}_+ \times [0, z], \times \mathcal{N}} \int_{\Gamma} \xi(\gamma) \mathcal{U}((c(\gamma), z(\gamma), n(\gamma)) \ d\Pi(\gamma)$$

subject to the incentive constraints

$$\max_{\gamma' \in \Gamma} \Psi \left( u(c(\gamma')) - \theta h \left( \frac{z(\gamma')}{\theta} + b_n \right) + m(n(\gamma'), \beta) \right) \leq \mathcal{U}((c(\gamma), z(\gamma), n(\gamma))$$

for all $\gamma \in \Gamma$

(7)

and the resource constraint

$$G \leq \int_{\Gamma} [z(\gamma) - c(\gamma) - e_n(\gamma)]d\Pi(\gamma).$$

In (MDP), the government chooses a report-contingent allocation of consumption, income, and number of children $\{c(\gamma), z(\gamma), n(\gamma)\}$ for all $\gamma \in \Gamma$ that induces every family truthfully report its type $\gamma$ and produce $z(\gamma)$ income and have $n(\gamma)$ children. Incentive constraints (Equation (7)) ensure the optimality of truthful reporting. If type $\gamma$ pretends
to be type $\gamma'$, she must produce $z(\gamma')$ income and have $n(\gamma')$ number of children.

The characteristics of families are their private information and therefore misreporting can be two dimensional. To handle two dimensional misreporting strategies, I use the definitions of indirect utility and critical tastes for children. First, given $n$, I use first-order approach to prevent misreporting $\theta$:

$$
V_n(\theta) := \frac{\partial V(\theta)}{\partial \theta} = -h\left(\frac{z_n}{\theta} + b_n\right) + h'\left(\frac{z_n}{\theta} + b_n\right) \frac{z_n}{\theta} \geq 0 \quad \forall n \in \mathcal{N}.
$$

(8)

Second, using the indirect utilities, I handle misreporting $\beta$ by Equation (6) which implies families with different sizes are not better off by misreporting their $\beta$. In addition, for a particular family size, misreporting $\beta$ does not change the family utility.

I adjust (MDP) with these definitions. 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”.

### 3.2.1 Pseudo-Mechanism Design Problem

The government problem solves the following problem:

$$
\max \{c_n(\theta), z_n(\theta)\} \in \mathbb{R} \times [0, \pi] \int\sum_{n \in N} \int_{\beta_{n+1}(\theta)}^{\beta_n(\theta)} \zeta_n(\theta) \Psi \left(V_n(\theta) + m(n, \beta)\right) p(\beta|\theta) f(\theta) d\beta d\theta \quad \text{(PMDP)}
$$

subject to Equation (5), (6), and (8) and the resource constraint:

$$
G \leq \int\sum_{n \in N} \int_{\beta_{n+1}(\theta)}^{\beta_n(\theta)} \left(z_n(\theta) - c_n(\theta) - e_n\right) p(\beta|\theta) f(\theta) d\beta d\theta
$$

(9)

where $\beta_0(\theta) = \beta$ and $\beta_{N+1}(\theta) = \bar{\beta}$. Adjusting the problem does not alter the solution:

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

**Proof.** See Appendix A.2.

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 for all $n \in \mathcal{N}$:

$$
\frac{T_n'(\theta)}{1 - T_n'(\theta)} = \frac{1}{\epsilon_n(\theta)} \times \frac{1}{\theta H_n(\theta)} \times R_n(\theta)
$$

(10)
where

\[
\varepsilon_n(\theta) = \frac{\partial \log z_n(\theta)}{\partial \log(1 - T_n'(z_n(\theta)))} = \frac{h'(z_n(\theta)/\theta + b_n)}{(z_n(\theta)/\theta)h''(z_n(\theta)/\theta + b_n)},
\]

\[
H_n(\theta) = f(\theta)(P(\beta_{n+1}(\theta)|\theta) - P(\beta_n(\theta)|\theta)),
\]

\[
R_n(\theta) = \theta \hat{\theta} \left[ (1-g_n(\theta)) u'(c_n(\theta)) (\frac{P(\beta_{n+1}(\theta')|\theta') - P(\beta_n(\theta')|\theta'))}{u'(c_n(\theta'))} \right] u'(c_n(\theta)) f(\theta') d\theta'
\]

where \(z_n, T_n\) is continuous in \(\theta\), and \(\varepsilon_n(\theta)\) is the elasticity of income of \(n\)-child families with respect to marginal taxes, and \(g_n(\theta)\) is the weight assigned by the government to \(\theta\)-ability families with \(n\) children, and \(\Delta T_n(\theta) := T_n(\theta) - T_{n+1}(\theta)\) where \(T_n(\theta) = z_n(\theta) - \varepsilon_n(\theta)\).

Proof. See Section 6.

I let \(u(c) = c\) and provide an heuristic proof of Proposition 1 for one-child families. Suppose that the government increases the taxes of one-child families with \(\theta' \geq \theta\) earning abilities by \(dT\) (see Figure 3a).

![Figure 3: Heuristic Proof of Proposition 1](image)

This change creates three effects. First, for a \(\theta'\) one-child family, there will be a welfare loss for the society by \(g_1(\theta')dT\) due to the decrease in consumption of one-child families by \(dT\) where \(g_1(\theta')\) measures the average cost of taking an extra dollar more from \(\theta'\)

---

---

\(^9\)I let \(\Delta T_{-1}(\theta) = 0\) when \(n = 0\) and \(\Delta T_N(\theta) = 0\) when \(n = N\).
families with one-child in terms of the public good.  

\[ g_1(\theta) := \mathbb{E}_\beta \left[ \frac{\tilde{g}_1(\theta) \Psi'(V_1(\theta) + m(1, \beta))}{\lambda} | \beta_1(\theta) < \beta < \beta_2(\theta) \right]. \]

On the other hand, the government revenue increases by \(dT\). Therefore the net effect for a \(\theta'\) ability family with one child is: \(dT(1 - g_1(\theta'))\). The aggregate effect for all \(\theta' \geq \theta\) is:

\[ dM = dT \int_{\theta}^{\tilde{\theta}} (1 - g_1(\theta')) \left[ P(\beta_1(\theta')|\theta') - P(\beta_2(\theta')|\theta') \right] f(\theta') d\theta'. \]

Note that \(dM\) is a mechanical effect and 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 \(\tau = \tilde{\tau} \tilde{\theta}^{\varepsilon_1} \), where \(\tilde{\tau}\) represents the change in the marginal tax rates on income (see Figure 3a). This increment creates a behavioral response, i.e. the families in the small band decrease their income by \(dz = \frac{\varepsilon_1(\theta) \tilde{\theta}}{\tilde{T}_1(\theta)}\), where \(\varepsilon_1(\theta) := \frac{\partial \log z_1(\theta)}{\partial \log (1 - \tilde{T}_1(\theta))}\) is the elasticity of income of one child families with respect to marginal tax rates. Combining the terms gives the first behavioral effect is:

\[ dB_1 = -T'_1(\theta) dz f(\theta) d\theta = -dT \frac{T_1(\theta)}{1 - T_1(\theta)} \varepsilon_1(\theta) \theta \left[ P(\beta_1(\theta)|\theta) - P(\beta_2(\theta)|\theta) \right] f(\theta). \]

Third, \(dT\) affects family sizes. Families with tastes for children are in the neighborhood of \(\beta_1(\theta)\) and \(\beta_2(\theta)\) alter their children choice which is shown in Figure 3b. 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: \(\Delta T_0(\theta') \frac{\partial \beta_1(\theta')}{\partial V_1(\theta')} p(\beta_1(\theta')|\theta') f(\theta')\) where \(\frac{\partial \beta_1(\theta')}{\partial V_1(\theta')}\) is the mechanical effect of \(V_1(\theta')\) on \(\beta_1(\theta')\) and \(p(\beta_1(\theta')|\theta') f(\theta')\) is the density of these families. Similarly, one-child families in the neighborhood of \(\beta_2(\theta)\) prefers to have two children. For this case, the effective change is: \(\Delta T_1(\theta') \frac{\partial \beta_2(\theta')}{\partial V_1(\theta')} p(\beta_2(\theta')|\theta') f(\theta').\)

---

10 See Equation (13) for a general definition.
11 To change the marginal rates over abilities by \(\tau\), the marginal rates on income should increase by \(\tilde{\tau}\).
12 The new critical tastes are represented by \(\tilde{\beta}_1(\theta)\) and \(\tilde{\beta}_2(\theta)\). See Figure 3b.
The aggregate effect of the change of the family size is represented by:

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

Together with the second behavioral effect, the original mechanism is optimal if \( dM + dB_1 + dB_2 = 0 \). This equality gives the equation in Proposition 1 when \( u(c) = c \) for \( n = 1 \). Note that, this method can be processed for any \( n \in \mathcal{N} \) 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 labor force. 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 \( h(x) = \frac{x^{1+\frac{1}{r}}}{1+\frac{1}{r}} \). Then: \( \varepsilon_n(\theta) = \varepsilon(1 + \frac{b_n}{z_n/\theta}) \).

**Proof.** Note that \( \varepsilon_n := \frac{\log^r \partial z_n}{\log^r (1-T_n')} = \frac{1-T_n'}{z_n} \frac{\partial z_n}{\partial (1-T_n')} \). Equation (3) implies: \( (1-T_n') = h' \left( \frac{z_n}{\theta} + b_n \right) \).
Take derivative with respect to \( (1-T_n') \) and rewrite: \( \varepsilon_n = \frac{h'(\frac{z_n}{\theta} + b_n)}{h'' \left( \frac{z_n}{\theta} + b_n \right) \frac{z_n}{\theta}} = \varepsilon \left( 1 + \frac{b_n}{z_n/\theta} \right) \).

It is straightforward to see that \( \varepsilon_n(\theta) \) depends on \( z_n \), and hence the elasticity of income of parents is endogenous. Moreover, the elasticity for families with children is larger than the elasticity of childless families: \( \varepsilon_n(\theta) > \varepsilon_0(\theta) \). This result is in line with Blundell, Meghir, and Neves (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: \( f(\theta) \left( P(\beta_{n+1} | \theta) - P(\beta_{n} | \theta) \right) \). 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 abilities but also by insuring families with more children. Families with more children have to spend more goods and time to raise children. As a result their wel-
fare decreases. Hence, the government redistributes these families to reduce the welfare reduction due to child-rearing costs.

Next, I provide an interpretation of the terms of Proposition 1.

**Interpretation of the terms:** The interaction of the terms in Equation (10) is complex. Here, I go over term by term and provide a basic interpretation of each term. First, the elasticity, $\varepsilon_n(\theta)$, 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 dead-weight loss for the economy and reduce efficiency. Hence, the government reduces the marginal taxes of those with higher income elasticity.

Second, the density of family sizes, $H_n(\theta)$, decreases the marginals. Intuitively, if the density is large, the impact of the distortions created by the marginal taxes will be large and reduce efficiency. Therefore, the government decreases marginal rates.

Third, $G_n(\theta)$ measures the redistribution tastes of the government. If the government wants to redistribute a particular family type, the government decreases their tax liabilities (or increases the transfers to those families) and subsequently increases their marginals.

### 3.3 Understanding the Shape of Credits

In this subsection, I will give an example to show the main forces behind the shape of the credits. For simplicity, I let $u(c) = c$ and $N = \{0, 1\}$. Families who earn $z$ pay income taxes of $T(z)$ and those who have children get credit of $k(z)$.

#### 3.3.1 Only Goods Cost

Suppose that $b_1 = 0$. Suppose that there are two abilities $\theta_L, \theta_H$ such that $\theta_L < \theta_H$. Without government intervention, families produce $z_n(\theta_j) = \theta_j$ for $n \in N$ and $j = L, H$. The percentage change in potential consumption of families due to the goods cost of child-rearing is much higher for low abilities (incomes). Therefore, the change in marginal utility of consumption is much higher for low income families. Consequently, the government redistribution motives are stronger at the bottom and hence more credits are given to the low income families.
3.3.2 Only Time Cost

Suppose that \( e_1 = 0 \). Assume that there are three abilities: \( \theta \in \{ \theta_L, \theta_M, \theta_H \} \) with \( \theta_L < \theta_M < \theta_H \). The downward (binding) incentive constraints of families are:

\[
z_0(\theta_M) - h\left(\frac{z_0(\theta_M)}{\theta_M}\right)\theta_M - T(z_0(\theta_M)) \geq z_0(\theta_L) - h\left(\frac{z_0(\theta_L)}{\theta_M}\right)\theta_M - T(z_0(\theta_L)) \quad \text{(IC-0M)}
\]

\[
z_1(\theta_H) - h\left(\frac{z_1(\theta_H)}{\theta_H}\right)\theta_H - T(z_1(\theta_H)) + k(z_1(\theta_H)) \geq z_1(\theta_M) - h\left(\frac{z_1(\theta_M)}{\theta_H}\right)\theta_H - T(z_1(\theta_M)) + k(z_1(\theta_M)) \quad \text{(IC-1H)}
\]

Also, let exogenous abilities are chosen such that \( z_1(\theta_H) = z_0(\theta_M) \) and \( z_1(\theta_M) = z_0(\theta_L) \) under the optimal government interventions. I check how optimal credits should be constructed under optimal tax system. Define

\[
K(z) := \left[ \theta_M h\left(\frac{z}{\theta_M}\right) - \theta_H h\left(\frac{z}{\theta_H} + b\right) \right]
\]

IC-0M and IC-1H imply:

\[
k(z_1(\theta_H)) - k(z_1(\theta_M)) = K(z_1(\theta_M)) - K(z_1(\theta_H)).
\]

Note that \( K'(z) < 0 \) if and only if \( \frac{z}{\theta_M} < \frac{z}{\theta_H} + b \). This inequality holds for \( z_1(\theta_H) \), and consequently for all \( z < z_0(\theta_M) \). Therefore

\[
k(z_0(\theta_M)) - k(z_0(\theta_L)) > 0. \quad (11)
\]

The intuition behind Equation (11) is the following. Owing to the opportunity cost of child care, one child families has higher abilities than zero child families who have same earnings. Consequently, the incentive constraints for one child families are much tighter comparing to those of the zero child families, because it is much easier for one child families to produce a lower income. This tightness gets stronger towards high income families. Hence, the government increases credits for high income families.

The effect of cost types are distinct. To see the overall effect and quantitatively analyze optimal taxes, I bring my model to the US data.

---

13Most of optimal tax papers assume and show downward incentive binds and upwards do not.

14\( z_1(\theta_H) = \theta_H (1 - b) = z_0(\theta_M) = \theta_M (1 - T'(z_0(\theta_M))) \), where equalities come from first order conditions for \( z_0(\theta_M) \) and \( z_1(\theta_H) \), respectively.
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.

4.1 Utility Function Form

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 \( h(\frac{z}{\beta}) = \frac{(\frac{z}{\beta} + b_n)^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \). This form implies childless families have a constant elasticity of income with respect to marginal rates: where \( \epsilon := \frac{\partial \log z}{\log (1-T')} \). 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.3. In the benchmark, I use \( \epsilon = 0.56 \). 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.

The utility from number of children is new to the literature. I let a weakly super-modular function: \( m(n, \beta) = -(N-n)\rho(\beta - \bar{\beta}) \) where \( \rho \geq 1 \) and I estimate \( \rho \).

Finally, I let \( \Psi(U) = \frac{U^{1-\sigma}}{1-\sigma} \). Chetty (2006) suggests that \( \sigma := \frac{-U_{cc}}{U_{cc}} \) is an upper bound of the curvature of utility over wealth and should be less than two to have positive labor supply response to increase in wages. Kleven, Kreiner, and Saez (2009) studies with the UK data and set \( \sigma = 1 \). I set \( \sigma = 0.8 \) as in Jacquet, Lehmann, and der Linden (2013) who use US Current Population Survey (CPS).

4.2 Data and Sample Selection

I use the March release of the CPS administered by the US Census Bureau and the US Bureau of Labor Statistics.\(^{15}\) I use the sample of 2005-2014 years in which families report both their child tax credits and their marginal tax rates.

The sample selection has four main criteria: Marital status, employment status, age of spouses, and source of income.

First, I fix number of spouses in families and in the benchmark. This restriction eliminates potential time difference between one-spouse and two-spouse families.

Second, I focus on only employed spouses which rules out the extensive margin decision and helps to capture a fine estimate for ε. In the benchmark, I focus on two-spouse families who jointly file their income taxes. The main reason is that, child benefits are available for married households filling jointly under the current US income tax code. In the robustness analysis, I focus on one-spouse families.

Third, 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. 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, Schoonbroodt, and Tertilt (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.

Fourth, 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)). Non-labor income is positively correlated with fertility (see Jones, Schoonbroodt, and Tertilt (2010)) and ruling out the non-labor income shows the interaction between labor and fertility. In addition, I rule out families who earn less than $20,000 which is around 130% of federal poverty level for two-people families in 2011 and families below threshold benefit from some other mean-tested programs. Since I focus on only income tax benefits for children, this is a valid assumption. I also rule out families who earn more than $200,000 because they are in phase-in region of the current tax credit programs. Finally, I follow Heathcote, Perri, and Violante (2010) and eliminate families who earn less than $250. The final sample has 31,066 families.

With this sample restriction, I plot the family labor income and the number of children in the family in Figure 4. The figure implies the well-known empirical evidence that the fertility rate is negatively correlated with family labor income.

\[^{16}\text{See http://www.bls.gov/news.release/wkyeng.t03.htm}\]
Next, I focus on the child-rearing costs and find estimates for $b_n$ and $e_n$.

### 4.3 Parental Time

The assumption on $h$ normalizes the cost of working (see Kleven, Kreiner, and Saez (2009)). In laissez faire, Equation (3) implies $z_n(\theta) = \theta(1 - b_n)$, $\theta b_n$ is forgone earnings due to child care. Accordingly, I interpret $b_n$ as the opportunity cost of child care. Knowles (1999) and de la Croix and Doepke (2003) follow Haveman and Wolfe (1995) and set opportunity cost of 15% of motherhood time per child. Since this number is only based on females, I estimate $b_n$ using the 2003 wave of American Time Use Survey (ATUS) sample which is also used by Aguiar and Hurst (2007). With the sample selection in Section 4.2, I show the market labor of families in Table 1.
Table 1: Opportunity Cost of Child Care

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. Opportunity cost of $n$ children is calculated as $b_n := \frac{l_0 - l_n}{l_0}$.

Table 1 shows that there is an economics of scale in the time cost of child-rearing. The estimates $b_n$ are much lower comparing to 15%. The main reason is my sample includes also males. If I particularly focus on market hours of females, the estimates are very close to the Haveman and Wolfe (1995) (see Table 2).

<table>
<thead>
<tr>
<th></th>
<th>0 child family</th>
<th>1 child families</th>
<th>2 children families</th>
<th>3+ children families</th>
</tr>
</thead>
<tbody>
<tr>
<td>labor ($l_n$)</td>
<td>42.34</td>
<td>40.05</td>
<td>38.18</td>
<td>37.47</td>
</tr>
<tr>
<td>$b_n \simeq$</td>
<td>0</td>
<td>0.054</td>
<td>0.098</td>
<td>0.115</td>
</tr>
<tr>
<td>sample size</td>
<td>320</td>
<td>550</td>
<td>884</td>
<td>415</td>
</tr>
</tbody>
</table>

Table 2: Opportunity Cost of Child Care for Females

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 females. Opportunity cost of $n$ children is calculated as $b_n := \frac{l_0 - l_n}{l_0}$.

<table>
<thead>
<tr>
<th></th>
<th>0 child family</th>
<th>1 child families</th>
<th>2 children families</th>
<th>3+ children families</th>
</tr>
</thead>
<tbody>
<tr>
<td>labor ($l_n$)</td>
<td>37.8</td>
<td>32.58</td>
<td>31.41</td>
<td>29.58</td>
</tr>
<tr>
<td>$b_n \simeq$</td>
<td>0</td>
<td>0.138</td>
<td>0.168</td>
<td>0.217</td>
</tr>
<tr>
<td>sample size</td>
<td>175</td>
<td>302</td>
<td>484</td>
<td>204</td>
</tr>
</tbody>
</table>

4.4 Cost of Goods

Haveman and Wolfe (1995) use Consumer Expenditure Survey (CEX) data and suggest that the goods cost is $12,151 per child (in terms of 2011$). 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 (2012), analyzes the goods cost of child-rearing for families with different wealth. 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 3.

Families with Average Income

<table>
<thead>
<tr>
<th></th>
<th>1 child</th>
<th>2 children</th>
<th>3 children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Income</td>
<td>38,000</td>
<td>11,313-12,463</td>
<td>18,100-19,940</td>
</tr>
<tr>
<td>Middle Income</td>
<td>79,940</td>
<td>15,463-17,900</td>
<td>24,740-26,690</td>
</tr>
<tr>
<td>High Income</td>
<td>180,040</td>
<td>25,575-30,638</td>
<td>40,920-49,020</td>
</tr>
</tbody>
</table>

Table 3: Expenditures on Child-rearing

Costs and the information in Table-1 of Lino (2012) 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 in 2011$.

In addition, Figure 2 of Lino (2012) provides the shares of expenditures:

<table>
<thead>
<tr>
<th>Food</th>
<th>Clothing</th>
<th>Health care</th>
<th>Housing</th>
<th>Transportation</th>
<th>Care &amp; Education</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>16%</td>
<td>6%</td>
<td>8%</td>
<td>30%</td>
<td>14%</td>
<td>18%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Table 4: Share of Expenditures on Child-rearing

Note that the ranges of expenditures in Table 3 are narrow. In benchmark case, I set expenditures of low income families is minimum level. Moreover, I consider the minimum requirement for child expenditure should contain food, clothing, and health care expenditure. Hence, I use $e_1 = 3,600$, $e_2 = 5,700$, and $e_3 = 6,600$. I interpret the other expenditures for all families and extra costs for middle and high income families as a part of their family consumption. Note that $e_1$ is quite close to the personal exemption amount of 2011 tax year (which $3,700).

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

### 4.5 Estimation of the Distribution of Earning Abilities

The quasi-linear preference structure allows me to find earning abilities of families (see Equation (3)):

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

(12)

Note that CPS has information about family structure and detailed sources of family income and taxes.\textsuperscript{18} Given the complexity of state rates, I focus only on federal tax rates. I

\textsuperscript{18}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.
add earned income tax rates to the reported federal marginal tax rates. Using \( b_n \) values from Table 1 and the weights assigned to the family types (given by the data), I derive the distribution of earning abilities and show it in Figure 5a.

\[
\begin{align*}
\theta & \text{ in 000s of 2011} \\
f(\theta) & \times 10^{-3} \\
0 & 2 4 6 8
\end{align*}
\]

(a) Ability Distribution: \( f(\theta) \)

\[
\begin{align*}
\beta & \\
p(\beta) & 0.01 0.008 0.006 0.004 0.002
\end{align*}
\]

(b) Child Taste Distribution: \( p(\beta) \)

Figure 5: Probability Distribution for Family Characteristics

### 4.6 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 \( \beta \sim \text{i.i.d. } [1, \infty) \) is distributed according to an exponential function: 
\[
P(\beta) = 1 - \beta^{-\lambda}
\]
where \( \lambda := -\frac{\partial \log(1-P(\beta))}{\partial \log \beta} \) is the negative of the elasticity of having \( N \) children with respect to the marginal cost of \( N^{th} \) children.\(^{20}\) Unfortunately, the literature on elasticity of fertility with respect to its cost is very sparse. In an interesting work, Cohen, Dehejia, and Romanov (2013) use an Israeli data and particularly estimate a price elasticity for the third children of \(-0.42\). The overall price elasticity is estimated as \(-0.54\) and benefit elasticity of \(0.19\). Milligan (2005) using Canadian data find benefit elasticity of \(0.105\) and Laroque and Salanie (2008) using French data estimates benefit elasticity of \(0.2\). Whittington, Alm, and Peters (1990) estimates suggests the elasticity of the US benefits is in between \(0.12\) and \(0.23\). Empirical evidences suggests benefit elasticities are similar across countries. Therefore, I set \( \lambda = 0.5 \) and fix \( N = 3 \). I plot the distribution of child

\[^{19}\text{The families report how much earned income credit they received. Yet, the data does not provide if credits are in the phase-in or out region. I use the information on EITC for years 2005-2014 to figure out the marginal effect of the credit.}\]

\[^{20}\text{Given a } \theta, \text{ the fraction of } N-\text{child families is } 1 - P(\beta_N(\theta)) \text{ and marginal cost of } N^{th} \text{ children is equal to } \beta_N(\theta).\]
taste in Figure 5b.

I use Bernoulli maximum likelihood estimation to find the estimate of $\rho$. First, I discretize $\Theta$ to its percentiles and calculate $V_n(\theta_j)$ for each $j^{th}$ percentile. For all $\theta_j$, Equation (6) implies that the theoretical probability of having $n$—children is:

$$P_n(\theta_j) := P(\beta_{n+1}(\theta_j)) - P(\beta_n(\theta_j)) \quad \forall n \in \mathcal{N}$$

where $\beta_0(\theta_j) = \underline{\beta}$ and $\beta_{N+1}(\theta_j) = \overline{\beta}$. Second, I can calculate the fraction of $n$—child families from data: $\pi_n(\theta_j)$. Finally, I derive the Bernoulli maximum likelihood function:

$$\max_\rho L = \prod_{n=0}^{3} \prod_{j=0}^{100} P_n(\theta_j)^{\pi_n(\theta_j)}.$$ (ML)

The solution of (ML) is $\hat{\rho} = 3.96$ where $\hat{\sigma_\rho} = 1.85$.

### 4.7 Deriving the Optimal Tax System

In benchmark, I set $G = $9,561, since the government collects $9,561 per capita taxes from the sample selection. I numerically solve the government’s problem (i.e. PMDP), which is an optimal control problem and its Hamiltonian is stated in Section 6, at my selected and estimated parameters using the numerical solver DIDO version 7.3.7.

Figure 6 compares optimal and statutory income taxes of the 2011 tax year. Optimal taxes are much lower comparing to the statutory taxes.

Figure 7 compares optimal and statutory tax credits of the 2011 tax year. There are two important results. First, optimal credits are U-shaped according to income. In contrast, the US credits are mainly decreasing over income. The shape of the optimal credits are based on the effect of child-rearing costs. The impact of goods costs increases credits for low income families, on the other hand, the government increases credits for high income families (with children) to incentivize them to work harder. The shape of statutory credits implies that the US government only focuses the goods costs and does not deal with the time cost.

Second, the US credits are (almost) same for each child for a large range of income. In contrast, the optimal credits are decreasing by the number of children in family. The

---

21See the estimation strategy in Appendix A.4.
22Per capita income of sample is $92,869.
23For details on the solution algorithm, refer to Ross and Fahroo (2003).
24There is small jumps in the statutory credits which are due to the effect of personal exemption when marginals change.
economies of scale on the impact of child-rearing costs on families is the source of decreasing credit over family size which is disregarded by the US government. On other hand, the UK government has recently proposed to cut the third child benefits which is in line with this result.
5 Robustness

In this section, I analyze robustness of U-shaped tax credits. First, I focus on the effect of $\gamma$. Second, I study the effect of $\lambda$ and $\varepsilon$, elasticity of probability of having three children with respect to the marginal cost of third children and elasticity of family income with respect to one minus marginal taxes.

5.1 Effect of $\gamma$

Figure 7 plots the tax credits under different $\gamma$. When $\gamma$ increases, the credits are getting larger for each children. The intuition is straightforward. The more risk averse families are, the stronger redistribution motives emerge.

![Figure 7: Effect of $\gamma$ on Optimal Tax Credits](image)

(a) Optimal Tax Credits when $\gamma = 0.5$

(b) Optimal Tax Credit when $\gamma = 1.5$

Figure 8: Effect of $\gamma$ on Optimal Tax Credits

5.2 Effect of $\lambda$ and $\varepsilon$

In this subsection, I study the effect of elasticities of having children and income. First, I set $\lambda = 1.5$ and reestimate $\rho$: $\hat{\rho} = 4.64$ with $\hat{\sigma}_\rho = 0.59$. With new estimates, I numerically solve the government problem again. Figure 9a shows the optimal credits. The credits are U-shaped according to income.

Next, I analyze effect of $\varepsilon$ by studying the optimal taxation of single females. I adjust $b_n$ according to Table 2 and goods cost as $e_1 = \$3,420$, $e_2 = \$5,400$, and $e_3 = \$6,420$. These costs are the 30% of the average costs in Table 8 of Lino (2012)).
set the elasticity of income as $\varepsilon = 0.8$ (see Blundell, Pistaferri, and Saporta-Eksten (2012)). In addition to these adjustments, I also set a boundary on earning abilities at $\bar{\theta} = 150$ because there are very few single females whose abilities are bigger than this boundary. Figure 9b plots the optimal tax credits for single females. First of all, credits are bigger. This is mainly due that the opportunity cost is much larger for single females. Second, the U-shape credits still optimal. Credits for first child is U-shaped according to income. However, the effect child-rearing costs are counterbalancing each other. As a result, credits for second children is almost constant and credits for the third is increasing. The main reason is that goods costs are lower and opportunity costs are much bigger. As expected, increasing credits will be more demanding and credits for third children is increasing.

![Figure 9: Robustness Analysis on $\lambda$ and $\varepsilon$](image)

### 6 Conclusion

This paper studies optimal income taxation and child tax credits in a static Mirrlees model in which potential parents are heterogeneous and have private information on their child tastes and earning abilities. Households decide how much income to generate and how many children to have by considering child-rearing costs. A Utilitarian government maximizes social welfare and determines the equity-efficiency trade-off owing to the redistribution motives and informational friction, respectively. An optimal tax mechanism is founded on this trade-off and combines income taxes of 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 redistribution 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 credits are decreasing by family size because of economies of scale in the impact of child-rearing cost on family welfare.

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. 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.
References


Appendix

Proof of Proposition 1. The Hamiltonian of the problem is:

\[ \mathcal{H} = \sum_{n=0}^{n=N} \int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \left( \xi_n(\theta) \Psi(V_n(\theta) + m(n, \beta)) + \lambda [z_n(\theta) - c_n(\theta)] \right) p(\beta | \theta) f(\theta) d\beta \]

\[ + \sum_{n=0}^{n=N} \mu_n(\theta) \left( -h \left( \frac{z_n}{\theta} + b_n \right) + h' \left( \frac{z_n}{\theta} + b_n \right) \frac{z_n}{\theta} \right) \]  

(Hamiltonian)

\[ \text{where } c_n(\theta) := u^{-1} \left( V_n(\theta) + h \left( \frac{z_n}{\theta} + b_n \right) \theta \right) \text{ and } \mu_n(\theta) = \mu_n(\bar{\theta}) = 0. \]

The first-order conditions are:

\[ z_n(\theta) : \lambda \left( 1 - \frac{\partial c_n(\theta)}{\partial z_n(\theta)} \right) (P(\beta_{n+1} | \theta) - P(\beta_n | \theta)) f(\theta) = -\frac{\mu_n(\theta)}{\theta} h' \left( \frac{z_n(\theta)}{\theta} + b_n \right) \frac{z_n(\theta)}{\theta} \quad \forall n \in \mathcal{N}. \]

Also, the co-states are:

\[ -\frac{\dot{\mu}_n(\theta)}{\lambda f(\theta)} = \int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \left( \xi_n(\theta) \Psi'(V_n(\theta) + m(n, \beta)) - \frac{\partial c_n(\theta)}{\partial V_n(\theta)} \right) p(\beta | \theta) d\beta + \sum_{j=n}^{n+1} \Delta T_{j-1}(\theta) p(\beta_j(\theta) | \theta) \frac{\partial \beta_j(\theta)}{\partial V_n(\theta)} \quad \forall n \in \mathcal{N}. \]

where \( T_n(\theta) = z_n(\theta) - c_n(\theta) \) and \( \Delta T_n(\theta) = T_n(\theta) - T_{n+1}(\theta). \)

Boundary conditions imply:

\[ -\frac{\mu_n(\theta)}{\lambda} = \int_{\theta}^{\bar{\theta}} \left[ \left( 1 - g_n(\theta') \right) (P(\beta_{n+1} | \theta') - P(\beta_n | \theta')) + \sum_{j=n}^{n+1} \Delta T_{j-1}(\theta') p(\beta_j(\theta') | \theta') \frac{\partial \beta_j(\theta')}{\partial V_n(\theta')} \right] f(\theta') d\theta' \]

where

\[ g_n(\theta) = \mathbb{E}_\beta \left[ \xi_n(\theta) \Psi'(V_n(\theta) + m(n, \beta)) u'(c_n(\theta)) | \beta_n(\theta) < \beta < \beta_{n+1}(\theta) \right] \]

\[ = \frac{\int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \xi_n(\theta) \Psi'(V_n(\theta) + m(n, \beta)) u'(c_n(\theta)) p(\beta | \theta) f(\theta) d\beta}{\lambda (P(\beta_{n+1}(\theta) | \theta) - P(\beta_n(\theta) | \theta)) f(\theta)} \quad (13) \]

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.

\[ ^{26} \text{Let } \beta_0 = \bar{\beta}, \text{ and } \beta_{N+1} = \bar{\beta}. \]

\[ ^{27} T_n(\theta) \text{ is the taxes paid by } \theta \text{-ability families with } n \text{ children. Let } \Delta T_{-1}(\theta) = 0 \text{ and } \Delta T_N(\theta) = 0. \]
Combining the of previous terms shows that the optimal tax function should satisfy:

\[
\frac{T_n'(\theta)}{1 - T_n'(\theta)} = \frac{1}{\varepsilon_n(\theta)} \times \frac{1}{\theta f(\theta)(P(\beta_{n+1}(\theta)|\theta) - P(\beta_n(\theta)|\theta))} \times \\
\int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - s_n(\theta'))}{u'(c_n(\theta'))} \left( P(\beta_{n+1}(\theta')|\theta') - P(\beta_n(\theta)|\theta') \right) \right] \\
+ \sum_{j=n}^{n+1} \Delta T_{j-1}(\theta') p(\beta_j(\theta')|\theta') \frac{\partial \beta_j(\theta')}{\partial V_n(\theta')} \right] u'(c_n(\theta)) f(\theta') d\theta' \quad \forall n \in N
\]

where \(\varepsilon_n(\theta)\) is the compensated elasticity of income of \(n\)-child families with respect to marginal tax rates.\(^{28}\)
A ONLINE APPENDICES

A.1 US CHILD TAX PROGRAMS

The Child Tax Credit (CTC) was enacted as a temporary provision in the Taxpayer Relief Act of 1997. The credit has gradually increased from $400 to $1,000 from 2001 to 2010 and become refundable for all families by the Economic Growth and Tax Relief Reconciliation Act of 2001. The credit is reduced by $50 for each $1,000 when aggregate gross income (AGI) is above $110,000 for married tax payers filing jointly. Finally, the credit has become permanent by the American Taxpayer Relief Act of 2012.

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 (Refer to Blau (2003) for more details).

Dependent Exemptions program decreases the AGI by an amount per child. The amount gradually increased from $2,800 to $3,700 from 2000 to 2011. This program is also mean-tested transfer and the exemption decreases beginning with phase-out income.

The Earned Income Tax Credit (EITC) is another mean-tested transfer program for working families. The maximum credit and phase in and out rates drastically change with the number of children in families. Table 5 shows the EITC rates for 2011. Families with more children are given more credits (Refer to Hotz and Scholz (2003) for more details).

<table>
<thead>
<tr>
<th># of children</th>
<th>earnings ≤</th>
<th>credit rate</th>
<th>max credit</th>
<th>phase-out begins</th>
<th>phase-out rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$6,070</td>
<td>0.08</td>
<td>$464</td>
<td>$12,670</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>$9,100</td>
<td>0.34</td>
<td>$3,094</td>
<td>$21,770</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>$12,780</td>
<td>0.40</td>
<td>$5,112</td>
<td>$21,770</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>$12,780</td>
<td>0.45</td>
<td>$5,751</td>
<td>$21,770</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 5: EITC for married tax payers filling jointly for 2011


29If 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.

30According to Falk and Crandall-Hollick (2016), %97 of EITC budget is spent for families with children.
A.2 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, Lehmann, and der Linden (2013) and Kleven, Kreiner, and Saez (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 characteristics. However, in this work, any existing child requires parental time, which is perfectly substitutable with market labor.

Let $\gamma = (\beta, \theta) \in B \times \Theta = \Gamma$ be the private information of a family. If the family report $\gamma$ as their type, the government chooses optimal allocation $(c(\gamma), z(\gamma), n(\gamma))$ and associated utility is:

$$U((c(\gamma), z(\gamma), n(\gamma))) = \Psi \left( u(c(\gamma)) - \theta h \left( \frac{z(\gamma)}{\theta} + b_n \right) + m(n(\gamma), \beta) \right).$$

This mechanism should satisfy the revelation principle, by which any government mechanism can be decentralized by a truthful mechanism $(c(\gamma), z(\gamma), n(\gamma))_{\gamma \in \Gamma}$ such that

$$U(c(\gamma), z(\gamma), n(\gamma)) \geq \Psi \left( u(c(\gamma')) - \theta h \left( \frac{z(\gamma')}{\theta} + b_n \right) + m(n(\gamma'), \beta) \right) \quad \forall (\gamma \times \gamma') \in \Gamma^2.$$

Let $U(((c(\gamma'), z(\gamma'), n(\gamma'))), \gamma) := \Psi \left( u(c(\gamma')) - \theta h \left( \frac{z(\gamma')}{\theta} + b_n \right) + m(n(\gamma'), \beta) \right)$ is the utility of the $\gamma$ family who report $\gamma'$ and gets $(c(\gamma'), z(\gamma'), n(\gamma'))$.

Family characteristics are two-fold, and hence a possible mimicking strategy has two dimensions. The possibility of double deviation in the mimicking strategy is handled by the indirect utility of $n$ child families (5) and critical tastes for children (6) 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 3.** Any truthful mechanism $(c(\gamma), z(\gamma), n(\gamma))_{\gamma \in \Gamma}$ can be replaced by a new truthful mechanism $\{c_n(\theta), z_n(\theta)\}_{\theta \in \Theta}$, such that $\forall \theta \in \Theta$ and $\forall n \in N$, there is a $\beta_n(\theta)$ such that if $\beta \in (\beta_n(\theta), \beta_{n+1}(\theta))$, then $U(c_n(\theta), z_n(\theta), n, \gamma) \geq \max_{\gamma'} U(((c(\gamma'), z(\gamma'), n(\gamma'))), \gamma)$. The new mechanism generates same utility as the original mechanism and the government collects as much taxes as the original mechanism.

---

$^{31}$I let $\beta_0 = \beta$ and $\beta_{N+1} = \tilde{\beta}$. 

34
Proof. For each \( \theta \), partition the set \( B \) into \( N + 1 \) sets such that if \( \beta \in B_j \) then \( n(\beta, \theta) = j \) for all \( j \in \mathcal{N} \). 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 \( \beta, \beta' \in B_j \), the truthfulness of the original mechanism implies:

\[
\begin{align*}
u(c(\beta, \theta)) - h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta + m(\beta, \theta) &\geq \nu(c(\beta', \theta)) - h\left(\frac{z(\beta', \theta)}{\theta} + b_j\right) \theta + m(\beta, \beta) \\
u(c(\beta', \theta)) - h\left(\frac{z(\beta', \theta)}{\theta} + b_j\right) \theta + m(\beta', \beta') &\geq \nu(c(\beta, \theta)) - h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta + m(\beta, \beta').
\end{align*}
\]

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 \( \nu(c(\beta, \theta)) - h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta \) is constant for all \( \beta \in B_j \), and let \( V_j(\theta) \) be its value.

Note that at least as much taxes should be collected with the new mechanism. Let \( Z_j(\theta) = \{z(\beta, \theta)|\beta \in B_j(\theta)\} \). Define \( t_j = \sup_{z \in Z_j(\theta)} z - u^{-1}(V_j(\theta) + h(\frac{z}{\theta} + b_j)) \). 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 \( z \to \infty \). So there is a \( z_j(\theta) \in \overline{Z_j(\theta)} \) such that \( t_j = z_j(\theta) - u^{-1}(V_j(\theta) + h(\frac{z(\theta)}{\theta} + b_j)) \). \(^{32}\)

Define \( c_j(\theta) := u^{-1}(V_j(\theta) - h(\frac{z(\theta)}{\theta} + b_j)) \).

Note that \( (c_j(\theta), z_j(\theta)) \) maximizes the taxes over the closure of the set \( \{c(\beta, \theta), z(\beta, \theta)\}_{\beta \in B_j(\theta)} \).

These procedures can be followed for all \( j \in \mathcal{N} \).

Finally, I define \( \overline{\beta}_n(\theta) := M_n^{-1}(V_n(\theta) - V_{n+1}(\theta)) \) where \( M_n(\beta) := m(n + 1, \beta) - m(n, \beta) \) for all \( n \in \mathcal{N}. \)\(^{33}\) \( \overline{\beta}_n(\theta) \) are the critical tastes for children for each \( \theta \) and for each \( n \). Note that truthfulness of original mechanism implies: for all \( \beta \in B_j(\theta) \) the family chooses \( n = j \) and \( (z_j(\theta), c_j(\theta)) \), i.e. \( V_j(\theta) + m(j, \beta) \geq V_j(\theta) - m(j', \beta) \) for all \( j' = 0, 1, \ldots, N \). Pick \( j' = j - 1 \) and \( j' = j + 1 \). Then it is easy to see that \( M_n^{-1}(V_j(\theta) - V_{j+1}(\theta)) \geq \beta \geq M_n^{-1}(V_{j-1}(\theta) - V_j(\theta)) \). \(^{34}\)

Therefore \( B_j(\theta) = (\overline{\beta}_{j-1}(\theta), \overline{\beta}_j(\theta)). \)

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

\(^{32}\) \( Z_j(\theta) \) is the closure of the \( Z_j(\theta) \)

\(^{33}\) Let \( \overline{\beta}_0 = \beta \) and \( \overline{\beta}_{N+1} = \overline{\beta} \).

\(^{34}\) Note that I let \( m \) to be concave in its first dimension and therefore \( \overline{\beta}_{j-1}(\theta) < \overline{\beta}_j(\theta). \)
$B_j(\theta')$:

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

where the inequality is from the truthfulness of the initial mechanism.\textsuperscript{35} 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 $\theta, \theta', \beta \in B_j(\theta), \beta' \in B_j(\theta')$:

$$U(c_j(\theta), z_j(\theta), (\beta, \theta)) = \Psi(V_j(\theta) + m(j, \beta)) \geq \Psi(V_{j'}(\theta) + m(j', \beta))$$

$$\geq \Psi \left( u(c(\beta', \theta')) - h \left( \frac{z(\beta', \theta')}{\theta} \right) + m(j', \beta) \right)$$

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 $j \in N$. \hfill \Box

This lemma allows me to move from \{$(c(\beta, \theta), z(\beta, \theta), n(\beta, \theta))$\}$_{(\beta, \theta) \in B \times \Theta}$ schedule to \{$(c_n(\theta), z_n(\theta))$\}$_{\theta \in \Theta}$ schedule. I directly use the one-dimensional implementation requirement as long as the single-crossing condition is satisfied.

**Definition 2.** \{$z_n(\theta)$\}$_{\theta \in \Theta}$ is implementable if and only if there exist transfer functions \{$c_n(\theta)$\}$_{\theta \in \Theta}$ such that \{$(c_n(\theta), z_n(\theta))$\}$_{\theta \in \Theta}$ is a truthful mechanism.

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

**Lemma 4.** The income profile \{$z_n(\theta)$\}$_{\theta \in \Theta}$ is implementable if and only if $z_n(\theta) := \frac{\partial a(\theta)}{\partial \theta} \geq 0$.

**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: $z \geq 0$ if and only if there is $c(\theta)$ such that $u(c(\theta)) - h \left( \frac{z(\theta)}{\theta} + b_n \right) \theta \geq u(c(\theta')) - h \left( \frac{z(\theta')}{\theta} + b_n \right) \theta$ for all $\theta, \theta'$.

\textsuperscript{35} Note that $(c_j(\theta'), z_j(\theta'))$ is in the closure of the set $(c(\beta, \theta), z(\beta, \theta))_{\beta \in B_j(\theta)}$. 

36
For the "if" side of the lemma, I directly apply the one-dimensional implementability condition: for all \( n \in \mathcal{N} \), let \( \{ z_n(\theta) \}_{\theta \in \Theta} \) implementable. Then for a particular \( n \), truthfulness implies
\[
u(c_n(\theta)) - h\left(\frac{z_n(\theta) + b_n}{\theta}\right) \geq \nu(c_n(\theta')) - h\left(\frac{z_n(\theta') + b_n}{\theta}\right)
\]
for all \( \theta, \theta' \). As a result, the one-dimensional result suggests that for each \( n \in \mathcal{N} \) 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
\[
u(c_n(\theta)) - h\left(\frac{z_n(\theta) + b_n}{\theta}\right) \geq \nu(c_n(\theta')) - h\left(\frac{z_n(\theta') + b_n}{\theta}\right)
\]
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 same in the proof of previous lemma where I show that cross-sectional deviation is not profitable.

\[\Box\]

**A.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 (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.\(^{36}\)

\(^{36}\)Blundell, Pistaferri, and Saporta-Eksten (2012) finds that the Frisch elasticity of married men is 0.4. For
The research on Frisch elasticity of females is not as large as on male elasticities. Blundell, Pistaferri, and Saporta-Eksten (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 40% of the family income (see Figure 10). Hence, \( \varepsilon = 0.6 \times 0.4 + 0.4 \times 0.8 = 0.56 \).

![Figure 10: Female Share of Family Income](image)

Average ratio is very close to 0.4. Note that this graph suggests that the gender gap for this sample is 0.67, which is quite close to the actual gender gap in the US (0.7).

### A.4 Estimation Strategy

There are 8 data variables in ML that I should have to estimate \( \rho: \{V_n(\theta), \pi_n(\theta)\}_{n=3}^{n=0} \). I use data of CPS and follow these steps:

1. Input Earned Income Tax rates and child rearing costs: \( b_n \) and \( e_n \)
2. Use Equation (3) and calculate \( \theta \) for each family
3. Calculate \( V_n(\theta) \) for each family (Federal taxes \( T_n(z_n) \) and \( T_n'(z_n) \)) are reported)
4. Discretize \( \Theta \) set to its percentiles \( \Theta = \{\theta_j\}_{j=0}^{100} \)

...different models the value goes up to 0.6.
5. Calculate \( \{ V_n(\theta_j), z_n(\theta_j), T_n(z_n(\theta_j)), T'_n(z_n(\theta_j)), \pi_n(\theta) \} \) as a weighted average of original variables for all \( n \) and for all \( j \)

6. For each \( z_n(\theta_j) \), I use Equation (3) and find \( \tau'(z_n) \) and compare with \( T'_n(z_n(\theta_j)) \) (see Figure 11)

7. To prevent null variables, I interpolate \( z_n(\theta_j) \) over discretized ability set \( \Theta \)

8. To check whether \( V_n(\theta_j) \) are fine, I put interpolated \( z_n(\theta_j) \)'s as an input to TAXSIM 9.2 version to calculate statutory taxes: \( T^{st}(z_n) \) (see Figure 12)

9. I get \( V^{st}_n(\theta) \) from statutory taxes compare with \( V_n(\theta) \) (see Figure 13)

10. I use \( \{ V_n(\theta), \pi_n(\theta) \} \) to estimate \( \rho \)

---

Figure 11: Comparison of \( \tau \)
Figure 12: Comparison of $T$

Figure 13: Comparison of $V$