OPTIMAL TAXATION OF FAMILIES: Mirrlees meets Becker*

Musab Kurnaz†
Tepper School of Business, Carnegie Mellon University

This Version: August 7, 2016

Abstract
This paper examines the optimal taxation of families in an environment in which (i) characteristics of a family, labor productivity and desire for children, are only observable by the family, and (ii) child-rearing requires both goods and parental time. Potential parents simultaneously decide labor income and number of children. The 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 parental time and the cost of goods involved in child-rearing have distinct impacts on the shape of optimal child tax credits. 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 are decreasing in the first three quarters 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. For median-income families, the credit for the second child equals 64% of the credit for the first child. I find that the optimal tax schedule generates a welfare gain equal to 1.3% of aggregate consumption.

JEL-Classification: H21, H53, D82, J13

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

*I am indebted to Christopher Sleet, Laurence Ales, and Isa Hafalir for invaluable advice, guidance, and encouragement throughout this project. I also thank Finn Kydland, Ali Shourideh, Kjetil Storesletten, Marla Ripoll, Randall Wright, and Hakki Yazici for very useful comments. I acknowledge seminar participants in Carnegie Mellon University, Koc University, Bilgi University, Bilkent University, Istanbul Technical University, Kadir Has University, TOBB-ETU, Ozyegin University, Midwest Macro Meetings Summer 2016, and Econometric Society Summer Meetings 2016.

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

The tax treatment of families has been in a debate throughout many decades in the United States. Policies regarding income tax schedules have frequently changed and tax benefits for parents have been improved. For example, the federal spending for Earned Income Tax Credits, which is mostly attained by parents, has been increased by 140% and the spending for Child Tax Credits has doubled up in the first decade of the 2000s. Along the same line, although there is a vast positive literature on the results of the policy changes, there is almost no normative work to guide governments on the design child tax credits. My paper fills this gap. In this paper, I ask a very important policy question: What should be the optimal child tax credits? To address this question, I study a model in which working families are heterogeneous and parents bear child-rearing costs. I find that child-rearing costs translate into U-shaped child tax credits in income. On the one hand, the reduction in parents’ consumption owing to the cost of goods relatively gets larger towards lower income parents. This enhances the government redistribution motives at the bottom of income distribution. Consequently, credits are increasing towards the lowest income parents. On the other, distortions on parents’ labor supply created by the time cost get greater towards the higher income parents. To reduce the efficiency loss owing to distortions, the government decreases the marginal taxes of high income parents, and hereby increases tax credits towards the highest income parents.

I make both theoretical and quantitative contributions. On the theoretical side, I embed a Becker model into an optimal tax framework. The former has been studied by labor economists to analyze the historical changes in fertility rates owing to changes in wages. This model allows me to study income taxation and child tax credits together. I show how the ingredients of this model affect the shape of optimal child tax credits. On the quantitative side, I estimate a parameter which is a key measure for marginal utility of a child given a structural form of utility from a discrete number of children. Such estimation is quite new to the public finance literature.

The public finance literature largely focuses on the incentive to supply labor by individual workers who are characterized by their privately observed labor productivity. In contrast, I study a family problem in which potential parents’ unobservable characteristics are twofold: their labor productivities and their tastes for children. Twofold

---

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

2In contrast, the main stream of normative tax literature solely focuses on income taxation.

3Tastes for children can be considered as a measure of child desire of households based on their exogenous characteristics, such as religion and race. See Becker (1960).
characteristics of potential parents form a two dimensional screening problem. Owing to the technical difficulties, the public finance literature dealing with multidimensional screening problems is sparse. My paper is one of exceptions. I cope with the difficulties by disentangling the effects of characteristics: The marginal benefit of a child is independent of parents’ labor productivities and the marginal cost of a child is independent of parents’ tastes. As a result, for a given labor productivity level, there exist critical tastes which equalize the marginal cost and the benefit of a child. Consequently, potential parents choose to have \( n \) children if their child tastes are between \( n^{th} \) and \( n + 1^{th} \) critical tastes. This optimality condition provides information on the underlying tastes of parents’ tastes. On the other hand, labor productivities of families are one of the key components of their income. Observable incomes of families provide information on families’ labor productivities. As a result, the government constructs an optimal tax schedule of families in a way that the families are not better off by neither changing their sizes nor generating another families’ income.

The taxation of families consists of a combination of income tax schedule with child tax credits. The former redistributes towards low earning families and the latter reduces the income tax liabilities of parents who are made monetarily worse off by child-rearing costs. The reduction at the bottom and at the top is high owing to goods and time cost, respectively. Goods cost decreases the consumption of low income parents relatively more than the high. This implies the decrease in the marginal utility of consumption, which is the key measure of government’s redistribution motives, is much higher at the bottom. As a result, the government increases credits towards lowest income parents. On the other hand, the time cost makes high income parents’ labor more sensitive to the tax changes as well as makes them reluctant to work harder comparing to non-parents. To alleviate the disincentive effect, the government has a motive for more provision to the wealthier parents and decreases their marginal taxes. The mechanical effect of decrement implies that tax credits are increased towards the highest income parents. Consequently, high income parents are incentivized to work harder not only by reducing their marginal taxes but also by increasing their consumption. The combination of the distinct effects of goods and time cost creates an optimal child tax credits that is U-shaped with respect to income.

I calibrate my model to the US economy and analyze the implications of child-rearing costs for optimal policy. First, I find estimates for the child-rearing costs. The goods cost is estimated using the information on the annual report of the US Department of Agriculture. Besides, the opportunity cost is considered as the ratio of forgone labor of parents owing to childcare to the labor of non-parents. Second, I use Current Population Survey
(CPS) to calculate families’ labor productivities from their consumption-labor margin by assuming as if the data was generated by an equilibrium of my model. CPS includes detailed information on taxes and weights of observations which allows me to derive the probability distribution of productivities. Third, I assume a particular distribution for child tastes and use a maximum likelihood estimation to capture a parameter which is a key determinant of the marginal benefit of a child.

The quantitative analysis provides two important policy results which are sharply different from the current US child tax credits. First, optimal tax credits for each child is U-shaped according to income. In particular, the credits are decreasing in the first three quarters of the income distribution and are increasing in the rest. In contrast, the US child tax credits are decreasing in family income. This has an implication of that the US government focuses on the impact of goods cost. However, the time cost significantly affects parents’ labor supply and reduces the total output of the economy.\(^4\) Therefore, including the effect of time cost into design 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 child-rearing effects. Since families have convex preferences over their welfare, the impact of first child has relatively larger than the impact of the second on family welfare. In addition, there is an economies of scale in child-rearing costs. On the one hand, time cost of one child is 96 market labor hours per year while it is 136 hours for two children. On the other, goods cost of two children is only 58% more than the goods cost of one child. The scale strengthens the impact of convex preference. The government values these impacts on the design of its redistribution motives and provides more credits for the first child. In particular, the credit for the second child is 64% of the first for the median income families. In contrast, the child tax credits in US are almost similar for each child according to family income.

The sharp differences between the optimal and the actual tax schedule bring the question of what the potential welfare gain from implementing the optimum is. I find that the gain is 1.3% in terms of equivalent increase in consumption for all families. This gain suggests that U-shaped tax credit schedule improves social welfare significantly. In addition, the optimal tax schedule makes 50% of population better off which implies the optimal policy can be quite attractive for the society.

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. Next, 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

\(^4\)The opportunity cost of childcare is 1.8% of GDP in 1992 according to Haveman and Wolfe (1995).
Section 6 and conclude with Section 7.

**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 as a consumption good, which is produced by inputs of goods and time. I refer these inputs as child-rearing costs. Haveman and Wolfe (1995) measure the child-rearing costs using Consumer Expenditure Survey (CEX) data. They find that total child-rearing costs is 14.5% of 1992 GDP. This large share shows the importance of child-rearing costs in the US economy. In addition, two-thirds of the costs is financed by parents which shows the significance of costs on family welfare.

Haveman and Wolfe (1995) find that the goods cost is 82% of total parental cost which includes expenditures on food, housing, health care, and clothes. The impact of the goods cost on the economy is studied by many works such as Golosov, Jones, and Tertilt (2007) and Hosseini, Jones, and Shourideh (2013). The former paper studies the efficiency of the allocations for future generations 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 that the fertility rate is negatively correlated by family labor income (see Jones, Schoonbroodt, and Tertilt (2010)). This is mainly because the opportunity cost of childcare is higher for the higher wage workers. In addition, time cost increases the sensitivity of parental labor to the wage changes. Blundell, Meghir, and Neves (1993) find that parents have higher Frisch elasticity than non-parents. 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 require more time and reduce available time for labor. The Frisch elasticity, in particular, is important because it is one of the major components of an optimal tax system (see Saez (2001)).

In contrast to many works in fertility literature that focuses on either goods cost or time cost, I study with both and show that their impact on family welfare is very important for the optimal policy design.

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. The main stream of the literature assumes workers’ productivities are only observable to workers themselves. In my pa-
per, working families privately observe their tastes for children as well as their labor productivities. This implies that the friction in the information is two dimensional. The literature on multidimensional screening problem is sparse owing to the technical difficulties.\footnote{Baron and Myerson (1982) and Rochet and Chone (1998) provide some additional requirements to solve multidimensional screening problems in the industrial organization literature.} Kleven, Kreiner, and Saez (2009) and Jacquet, Lehmann, and der Linden (2013) are notable exceptions. The former focuses on the jointness of couples’ income taxation in which primary earners’ labor productivities and secondary earners’ work costs are privately observed by families. They show that marginal income tax rates of the primary earner should be smaller if his or her spouse works. The latter studies an environment in which workers privately know their labor productivities and their taste of work and make an extensive labor decision. They provide a rationale for non-negative marginal rates. Unlike the studies above, I focus on not only the marginal rates but also the 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).\footnote{http://www.urban.org/sites/default/files/alfresco/publication-pdf/412699-Data-Appendix-to-Kids-Share--.PDF} 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 spending for former program is because per child credit is increased from $400 to $1,000 gradually and the eligibility conditions to claim CTC is relaxed. The increase in EITC spending mostly depends on the expansions of the program during 2000s.
The change in the federal budgets for tax credits (especially changes in the EITC program) is in 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 goods and parental time. The effects of these costs on family welfare are the key determinants of the tax credit design.

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 $z \in [0, Z]$, and number of children $n \in \mathcal{N} = \{0, 1, \ldots, N\}$ described by a utility function $U : \mathbb{R}_+ \times [0, Z] \times \mathcal{N} \to \mathbb{R}$. $U(\cdot, \cdot, n)$ is assumed to be concave, twice differentiable on its interior 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: $\frac{-U_x(c, z/\theta, \cdot)}{U_{zx}(c, z/\theta, \cdot)}$ is decreasing in $\theta$. 

Figure 1: Government Spending for EITC and CTC
The characteristics of families are twofold. Each family have a labor productivity \( \theta \in \Theta = (\theta, \bar{\theta}) \) and a taste for children: \( \beta \in B = (\beta, \bar{\beta}) \) in the population.\(^7\) 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 \( \theta \): \( \Pi(\beta, \theta) = \int_\Theta F(\beta|\theta) f(\theta) d\theta \) where \( f(\theta) \) is the unconditional distribution of \( \theta \). Family characteristics, \( \gamma \), are not observable to the rest of society.

Following Becker (1965), \( n \) children are output of exogenous inputs of \( e_n \) amount of goods and \( b_n \) amount of parental working time. This assumption justifies that the government equally care child-rearing across different income groups of families. 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 is considered 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 \( T_n(z) \). Hence, families with \( n \) children pay income taxes according to \( T_0(z) \) and get credit \( \sum_{j=1}^n k_j(z) \) where \( k_j(z) = T_{j-1}(z) - T_j(z) \) is the child tax credit for the \( j \)th child. Hence, tax liability of a \( n \) child family is \( T_n(z) = T_0(z) - \sum_{j=1}^n k_j(z) \).

Observing child-rearing costs, and the tax system, and their own characteristics, a \( \gamma \) family solves:

\[
\begin{align*}
\max_{c,z,n} & \quad \mathcal{U}(c, z, n) \quad \text{s.t.} \quad c \leq z - T_n(z) - e_n \\
\end{align*}
\]

where its utility function is:

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

\( u \) is a weakly concave function and \( h \) 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)). \( m(n, \beta) \) is concave in \( n \) and measures the interaction of the number of children and families’ taste over children. The utility function, \( \mathcal{U} \), is in line with Becker (1960) who suggests children should appear as a consumption good in family utilities.

Note that child choice is discrete, and hence first-order conditions are not immediately

---

\(^7\)As the main stream in public finance literature, I do not specifically model the production side of the economy and consider the labor productivities are exogenously given to families. There are few exceptions which specifically study how the production side affect productivities and its consequences on taxes. See Ales, Kurnaz, and Sleet (2015) and references there.
applicable. Therefore, I solve the family problem in two steps. Initially, families determine how many children to have. Next, 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 of families with \( n \) children:

\[
\begin{align*}
    u'(c_n) \left( 1 - \frac{\partial T_n(z_n)}{\partial z} \right) &= \frac{\partial h \left( \frac{z_n}{\theta} + b_n \right)}{\partial z}, \tag{2} \\
    c_n &= z_n - T_n(z_n) - e_n. \tag{3}
\end{align*}
\]

These equations imply \( c_n \) and \( z_n \) are independent from \( \beta \). By defining the indirect utility of \( n \)-child families:

\[
V_n(\theta) := u(c_n) - \theta h \left( \frac{z_n}{\theta} + b_n \right), \tag{4}
\]

I can measure the marginal cost of \( n^{th} \) child by \( V_{n-1}(\theta) - V_n(\theta) \) which is independent from families’ child taste. On the other hand, the marginal benefit of \( n^{th} \) child depends on the child taste: \( M(n, \beta) := m(n, \beta) - m(n - 1, \beta) \). The \( \gamma \) family have \( n \) children if and only if the marginal benefit of \( n^{th} \) child is larger than its marginal cost while the marginal benefit of \( n + 1^{th} \) child is less than its marginal cost. Formally, \( \gamma \) families decide to have \( n \) children if and only \( \beta \in (\beta_n(\theta), \beta_{n+1}(\theta)) \), where

\[
\beta_n(\theta) := M^{-1}(V_{n-1}(\theta) - V_n(\theta)), \tag{5}
\]

for \( n = 1, 2, \ldots, N \). Note that \( \beta_n(\theta) \) equalizes the marginal benefit and the cost of \( n^{th} \) child. I call \( \beta_n(\theta) \) as the critical taste level for \( n^{th} \) child for \( \theta \) productive families. Note that critical tastes depend on productivities. Given a productivity level, \( \{\beta_n(\theta)\}_{n \in \mathcal{N}} \) provides information on the underlying tastes of families. Using the family problem solution, I state the definition of a tax equilibrium:

**Definition 1.** Let \( G \in \mathbb{R}_+ \) be a fixed public spending amount. Given \( G \), a **tax equilibrium** 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 (FP) 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 of type $\gamma$ with weights normalized to satisfy $\int_\Gamma \xi(\gamma) d\gamma = 1$. The government selects a tax equilibrium to solve:

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

(3.1)

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:

$$\frac{\partial T^*_n(z_n)}{\partial z} = 1 + \frac{U_z(c^*(\gamma), z^*(\gamma), n^*(\gamma))}{U_c(c^*(\gamma), z^*(\gamma), n^*(\gamma))} \quad \forall n \in \mathbb{N}.$$  

(3.2)

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 part of tax equilibrium. The mechanism design problem associated with (GP) can be formulated as:

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

(3.3)

subject to the incentive constraints

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

(4.1)

and the resource constraint

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

(4.2)

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 (6) 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.

Families’ unobservable characteristics are two dimensional and therefore families can misrepresent any of dimensions of characteristics. Studying such environment is technically very hard. I handle technical difficulties by preventing misrepresentation sequentially as in the family problem solution. Given families do not find misrepresenting their
\( \beta \) beneficial, I use a first-order approach, which is the conventional procedure to prevent misrepresentation of \( \theta \):

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

(7)

Second, given the first-order conditions, I use critical tastes for children to prevent misrepresentation of \( \beta \).

I adjust (MDP) with (4), (5), and (7). The 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:

\[
\max \left\{ \{c_n(\theta), z_n(\theta)\} \in \mathbb{R}^+_\times [0, \bar{z}] \right\}_{n \in \mathcal{N}} \int_{\Theta} \sum_{n \in \mathcal{N}} \int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \xi_n(\theta) \Psi(V_n(\theta) + m(n, \beta)) p(\beta|\theta) f(\theta) d\beta d\theta \quad \text{(PMDP)}
\]

subject to 4, (5), and (7) and the resource constraint:

\[
G \leq \int_{\Theta} \sum_{n \in \mathcal{N}} \int_{\beta_n(\theta)}^{\beta_{n+1}(\theta)} \left( z_n(\theta) - c_n(\theta) - e_n \right) p(\beta|\theta) f(\theta) d\beta d\theta \quad \text{(8)}
\]

where \( \beta_0(\theta) = \beta \) and \( \beta_{N+1}(\theta) = \bar{\beta} \). The adjustment in the government’s 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) gives the optimal taxation of families which 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 \( n \in \mathcal{N} \):

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

(9)
where

\[
\varepsilon_n(\theta) = \frac{\partial \log z_n(\theta)}{\partial \log (1 - T'_n(z_n(\theta))))} = \frac{h'(z_n(\theta)/\theta + b_n)}{(z_n(\theta)/\theta)h''(z_n(\theta)/\theta + b_n)},
\]

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

\[
R_n(\theta) = \tilde{\theta}\left[ (1 - g_n(\theta')) \left( P(\beta_{n+1}(\theta')|\theta') - P(\beta_n(\theta')|\theta') \right) \right]
\]

\[+ \sum_{j=n}^{n+1} \Delta T_{j-1}(\theta') p(\beta_j(\theta')|\theta') \frac{\partial \beta_j(\theta')}{\partial V_n(\theta')} \right] u'(c_n(\theta)) f(\theta') d\theta'
\]

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 \(\theta\)-productive families with \(n\) children, and \(\Delta T_n(\theta) := T_n(\theta) - T_{n+1}(\theta)\) where \(T_n(\theta) = z_n(\theta) - c_n(\theta)\).\(^8\)

Proof. See Appendix.

To understand the economic intuition behind equations of Proposition 1, I provide an heuristic derivation of (9) for one-child families based on the effects of a small tax reform around the optimal tax system. For simplicity, let \(u(c) = c\) and suppose that the government increases the taxes of one-child families with \(\theta' \geq \theta\) productivities by \(dT\) (see Figure 2a).

![Figure 2: Heuristic Proof of Proposition 1](attachment:Fig2.png)

This change creates three effects. First, there will be a welfare loss for the society because consumption of one-child families is decreased by \(dT\). Note the average welfare

---

\(^8\) I let \(\Delta T_{-1}(\theta) = 0\) when \(n = 0\) and \(\Delta T_N(\theta) = 0\) when \(n = N\).
Similarly, one-child families in the neighborhood of mechanical effect of ply and their incomes decrease by $\frac{\partial}{\partial z}$. For this case, the effective change is:

$$\Delta \tau(z) = \frac{\partial}{\partial z} f(z')$$.

On the other hand, the government’s collects $dT$ amount of taxes from $\theta'$ families. Therefore the net effect is $dT(1 - g(\theta'))$ for a $\theta'$ family with one-child. Integration of net effects give the aggregate effect:

$$dM = dT \int_\theta (1 - g(\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 include any behavioral responses. Next, I focus on the behavioral responses of families to $dT$.

Second, families whose productivities are in $[\theta, \theta + d\theta]$ generate less income. To increase taxes by $dT$, the government should increase the marginal taxes of families whose productivities are in $[\theta, \theta + d\theta]$ by $\tau = \frac{d\tau}{d\theta}$, where $\tau$ represents the change in the marginal tax rates on income (see Figure 2a). Consequently, these families reduce their labor supply and their incomes decrease by $dz = \frac{z_1}{1 - T_1(z_1)}$, where $\epsilon_1(\theta) := \frac{\partial \log z_1(\theta)}{\partial \log(1 - T_1(z_1))}$ is the elasticity of income of one-child families with respect to marginal tax rates. Combining the terms gives the first behavioral effect is:

$$dB_1 = -T_1'(\theta) dz f(\theta) d\theta = -dT \left[ \frac{T_1'(\theta)}{1 - T_1'(\theta)} \epsilon_1(\theta) \theta [P(\beta_1(\theta)|\theta) - P(\beta_2(\theta)|\theta)] f(\theta') \right].$$

Third, families with tastes for children are in the neighborhood of $\beta_1(\theta)$ and $\beta_2(\theta)$ alter their sizes (see Figure 2b). 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: $\Delta T_1(\theta') \frac{\partial \beta_2(\theta')}{\partial V_1(\theta')} p(\beta_2(\theta')|\theta') f(\theta')$.

---

9To change the marginal rates over productivities by $\tau$, the marginal rates on income should increase by $\tau$.

10The new critical tastes are represented by $\tilde{\beta}_1(\theta)$ and $\tilde{\beta}_2(\theta)$. See Figure 2b.
The aggregate effect of the change of the family size is represented by:

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

To sum up, total effect of any small changes in the tax system should be zero, if the tax system is optimal: \( dM + dB_1 + dB_2 = 0. \) (9) directly emerges by the previous equality for \( u(c) = c \) and \( n = 1. \) Note that, this method can be processed for any \( n \in \mathbb{N} \) to get the full schedule.

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

**Interpretation of the terms:** The interaction of the terms in Equation (9) is complex. Here, I provide interpretation on the effect of each term on the marginal taxes. First, the elasticity, \( \varepsilon_n(\theta), \) is reciprocally correlated with the marginal taxes. The marginal taxes create distortions on labor supply and hence family income and the distortions are greater 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 optimal marginal rates are negatively correlated with family sizes.

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) and mechanically increases their marginals.

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

**Novelty of the tax formula:** One the most important contributions of this paper is that the tax formula is two dimensional. Two dimensions allow me to study income taxes and child tax credits together. Next, I focus on the effect of second dimension (number of children) on 9. The tax formula varies from the conventional tax formulas in the literature in three ways owing to the number of children. First, the elasticity component, \( \varepsilon_n, \) is *endogenous*. The endogeneity arises because time is perfectly substitutable between childcare and labor force. Time devoted to childcare 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 \( h(x) = \frac{x^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}. \) Then: \( \varepsilon_n(\theta) = \varepsilon(1 + \frac{b_n}{z_n/\theta}). \)
Proof. Note that \( \varepsilon_n := \frac{\log \partial z_n}{\log \partial (1 - T_n')} = \frac{1 - T_n'}{\partial z_n} \cdot \frac{\partial z_n}{\partial (1 - T_n')} \). Equation (2) implies: \( (1 - T_n') = h'(z_n + b_n) \).

Take derivative with respect to \( (1 - T_n') \) and rewrite: \( \varepsilon_n = \frac{h'(z_n + b_n)}{h'(z_n + b_n) \frac{\partial z_n}{\partial (1 - T_n')}} = \varepsilon \left( 1 + \frac{b_n}{z_n/\theta} \right) \). □

It is straightforward to see \( \varepsilon_n(\theta) \) 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: \( \varepsilon_n(\theta) > \varepsilon_0(\theta) = \varepsilon \) which is in line with Blundell, Meghir, and Neves (1993).

Second, a novel term, the density of family sizes appear in the formula: \( f(\theta)(P(\beta_n+1(\theta)|\theta) - P(\beta_n(\theta)|\theta)) \). This term is also 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 redistribution motives are shaped by the marginal utility consumption of families. The government not only redistributes from high to low income families but also from small size families to large size families. The former is characterized the difference between the marginal utility of consumption of high and low income families. The latter is shaped by child-rearing costs.

### 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 \( N = \{0, 1\} \). Moreover, there are two different child tastes: \( \beta = 0 \) and \( \bar{\beta} = \infty \) with positive probability to ensure some families will have 0 children and some will have 1 child. Families who earn \( z \) pay income taxes of \( T(z) \) and those who have a child get credit of \( k(z) \). Next, I consider environments in which there exists only one type of child-rearing costs.

#### 3.3.1 Only Goods Cost

Suppose that \( b_1 = 0 \) and there are two productivities \( \theta_L, \theta_H \) 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 parents owing to the goods cost is much higher for low productivities (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 parents.
3.3.2 Only Time Cost

In this subsection, I show the effect of time cost on the incentive constraints. Let \( e_1 = 0 \) and suppose there are three productivities: \( \theta \in \{ \theta_L, \theta_M, \theta_H \} \) with \( \theta_L < \theta_M < \theta_H \). Three types of productivities ensure comparison of incentive constraints across family sizes. The downward (binding) incentive constraints are:\(^{11}\)

\[
\begin{align*}
z_0(\theta_M) - h\left( \frac{z_0(\theta_M)}{\theta_M} \right) \theta_M - T(z_0(\theta_M)) &\geq z_0(\theta_L) - h\left( \frac{z_0(\theta_L)}{\theta_M} \right) \theta_M - T(z_0(\theta_L)) \quad \text{(IC-0M)} \\
z_1(\theta_H) - h\left( \frac{z_1(\theta_H)}{\theta_H} + b \right) \theta_H - T(z_1(\theta_H)) + k(z_1(\theta_H)) &\geq z_1(\theta_M) - h\left( \frac{z_1(\theta_M)}{\theta_H} + b \right) \theta_H - T(z_1(\theta_M)) + k(z_1(\theta_M)) \quad \text{(IC-1H)}
\end{align*}
\]

For tractability of the comparison, I let all exogenous variables be chosen such that \( z_1(\theta_H) = z_0(\theta_M) \) and \( z_1(\theta_M) = z_0(\theta_L) \) given an optimal tax system. In addition, suppose that income taxes are constructed such that zero-child families do not misrepresent their \( \theta \). Now, I check how optimal credits look like under this particular optimal tax system. First, I define

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

Hence, 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 \( z_1(\theta_H) \),\(^{12}\) and consequently for all \( z < z_1(\theta_H) = z_0(\theta_M) \). Therefore

\[
k(z_1(\theta_H)) - k(z_1(\theta_M)) > 0. \tag{10}
\]

The intuition behind (10) is the following. Given a particular income level of families, parents have higher productivities than non-parents owing to the opportunity cost of childcare. Consequently, the incentive constraints of one-child families are much tighter comparing to those of zero-child families, because producing low income is much more tempting owing to the convexity in \( h \). The tightness gets stronger towards higher produc-

\(^{11}\)I follow the main stream of literature that shows downward incentive constraints bind.

\(^{12}\)\( z_1(\theta_H) = \theta_H(1-b) = z_0(\theta_M) = \theta_M(1-T'(z_0(\theta_M))). \) where equalities come from first order conditions for \( z_1(\theta_H) \) and \( z_0(\theta_M) \), respectively.
tive as a result for higher income families. Hence, the government reduces the marginal
taxes of high-income parents and increases their tax credits.

The effects of goods cost and time cost on tax credits are quite distinct. To see the
overall effect for optimal policies, I bring my model to the US data.

4 Quantitative Analysis

In this section, I estimate the child-rearing costs and family preference over children as
well as the productivity distribution to examine the government’s problem (PMDP) quan-
titatively. First, I construct the forms of functions stated in the family utility (1) and create
criteria for sample selection. Second, I find estimates for child-rearing costs. Third, I de-
rive the productivity distribution and estimate the parameter that determines marginal
benefit of a child given a structural child taste distribution.

4.1 Family Preferences and Sample Selection

This subsection introduces family preferences and sample selection criteria in the bench-
mark. Section 6 provides robustness analysis on the preferences and the criteria.

4.1.1 Family Preference

I specify functions stated in (1). First, I assume \( u(c) = c \) based on the fact that most
of empirical studies find small income effect comparing to the substitution effect (see
Blundell and Macurdy (1999)). In addition, it is natural to eliminate the non-labor income
effect on labor to understand the relationship between labor income and fertility.

Second, I set \( h \left( \frac{\bar{z}}{\bar{b}} + b_n \right) = \left( \frac{\bar{z} + b_n}{1 + \bar{z}} \right)^{1 + \frac{1}{2}} \) which implies childless families \((b_n = 0)\) have a
constant elasticity of income with respect to marginal rates: \( \varepsilon := \frac{\partial \log z}{\partial \log (1 - T)} \). I set \( \varepsilon = 0.56 \)
by using individual level of elasticities in the literature (see Appendix A.3). Note that this
number is quite close to the elasticity estimates in Chetty (2012).

Third, I let a weakly super-modular function for the utility from children: \( m(n, \beta) = \)
\(- (N - n) \rho (\beta - \bar{\beta}) \). I estimate \( \rho \) in Section 4.3.2.

Finally, I choose a constant relative risk averse function over the family welfare: \( \Psi(U) = \)
\( \frac{U^{1 - \sigma}}{1 - \sigma} \). Chetty (2006) suggests that \( \sigma := - \frac{U_c}{U}\) is an upper bound of the curvature of utility
over the family wealth and should be less than two to have positive labor supply response

\[^{13}\text{Chetty (2012) construts a common confidence interval for the elasticities of many different studies.}\]
to positive wage changes. Jacquet, Lehmann, and der Linden (2013) set $\sigma = 0.8$ for their numerical analysis using the CPS. I let $\sigma = 0.8$.\(^{14}\)

### 4.1.2 Sample Selection

Data source is the March release of the CPS administered by the US Census Bureau and the US Bureau of Labor Statistics.\(^{15}\) I use 2005-2014 years’ sample because families report both their federal taxes, child tax credits, and their marginal tax rates in this period. In addition, I restrict this sample based on four main criteria: number of spouses, employment status, age of spouses, and family income.

First, the sample has families with same number of spouses. This restriction eliminates potential time difference between one-spouse and two-spouse families. In the benchmark, the sample consists of two-spouse families. I study with the sample of one-spouse families in Section 6.

Second, spouses in families are employed which rules out the extensive margin decision and helps to capture a fine estimate for $e_n, b_n$, and $e_n$.

Third, the age of each spouse in families are between 35-45. This restriction is used by many positive works that study the relationship between fertility and family income (see Docquier (2004), Jones and Tertilt (2008), and Jones, Schoonbroodt, and Tertilt (2010)). The main reason of this restriction is to eliminate the age effect on income stream of spouses as well as parenthood decision. According to Bureau of Labor Statistics, (median) earnings of households increase in the early ages (16-35) and become stabilized after the age of 35.\(^{16}\) Moreover, early age households may postpone parenthood decision owing to socioeconomic factors and the age restriction rules out the effect of postponement. The other reason of the restriction is to minimize the possibility that some children have grown up and left the family. The restriction implies parents of the sample bear child-rearing costs.

Fourth, families’ total labor income is between 80% and 120% of their total income (refer to Ales, Kurnaz, and Sleet (2015)). The former restriction guarantees that labor is the main source of family income. The latter sets a boundary on a negative business or other income. In addition, I make same restriction on individual labor income of each spouse. In total, income restriction minimizes the non-labor income effect on fertility decision.\(^{17}\) Finally, families who can get other mean-tested transfers and who do not benefit

\(^{14}\)Kleven, Kreiner, and Saez (2009) focus on the UK data and set $\sigma = 1$ for their quantitative exercise.


\(^{16}\)See [http://www.bls.gov/news.release/wkyeng.t03.htm](http://www.bls.gov/news.release/wkyeng.t03.htm)

\(^{17}\)Non-labor income is positively correlated with fertility (see Jones, Schoonbroodt, and Tertilt (2010)).
from tax credit programs are extracted from sample. Therefore, families’ labor incomes are between $20,000 and $200,000. The sample consists of 28,303 families within these restrictions.

Figure 3 shows how the average number of children change with the labor income distribution. The figure implies the well-known empirical evidence that the fertility rate is negatively correlated with family labor income.

![Figure 3: Income-Fertility Relation](image)

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 greater than $20,000 and less than $200,000 and converted to 2011$ using CPI deflator. Spouses are aged from 35 to 45. The sample size is 28,303.

### 4.2 Child-rearing Costs

I find estimates of child-rearing costs in this subsection. First, I analyze the goods cost stated in the literature. Next, I focus on the opportunity cost childcare and find estimates using CPS. I set $N = 2$ because the marginal cost of child-rearing for the third child is relatively low (See Table 1 and See Table 2) and the families with three or more children received EITC of two-child families until 2009. A robustness analysis on $N$ is studied in Appendix A.4.

---

18 The former level is around 130% of federal poverty level for a two-people family in 2011. Families below threshold may take benefits from other mean-tested programs which are beyond this paper. The latter level is in the beginning of the phase-in region of current tax credit programs.
4.2.1 Goods Cost: $e_i$

Haveman and Wolfe (1995) find that the annual goods cost is $12,151 per child (in terms of 2011$) based on Consumer Expenditure Survey (CEX). Examples of such costs include expenditures on food, housing, transportation, clothing, and health care. More recently, Lino (2012) analyzes the goods cost of child-rearing for families with different wealth and different size.\(^{19}\) This work particularly provides information on expenditures for children with different age. Using this information, I create a range of expenditures for two-spouse families in Table 1. For one child, $12,151 is the range of low income families’ expenses. However, Table 1 shows there is economies of scale in expenses. In addition, high income parents spend more. The minimum requirement of goods cost is considered as the spending of low income families on food, clothing, health care, and childcare and education expenditures. Figure 2 of Lino (2012) implies the share of these expenses is around 48% of total expenses. Hence, I set $e_1 = $5,500 and $e_2 = $9,500. The remaining expenditures are considered as a part of family consumption.

<table>
<thead>
<tr>
<th>Families with</th>
<th>Average Income</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>
<td>21,177-23,330</td>
</tr>
<tr>
<td>Middle Income</td>
<td>79,940</td>
<td>15,463-17,900</td>
<td>24,740-26,690</td>
<td>28,946-33,509</td>
</tr>
<tr>
<td>High Income</td>
<td>180,040</td>
<td>25,575-30,638</td>
<td>40,920-49,020</td>
<td>47,876-57,353</td>
</tr>
</tbody>
</table>

Table 1: Expenditures on Child-rearing

The ranges are constructed by the Table-1 of Lino (2012). The first column categorizes families according to their income. 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$.

4.2.2 Parental Time: $b_n$

The assumption on $h$ normalizes the cost of working (see Kleven, Kreiner, and Saez (2009)). In laissez faire, (2) implies $z_n(\theta) = \theta(1 - b_n)$, hence $\theta b_n$ is forgone family earnings owing to childcare. Accordingly, I interpret $b_n$ as the opportunity cost of childcare. Knowles (1999) and de la Croix and Doepke (2003) use the estimate of Haveman and Wolfe (1995) and set opportunity cost as 5-10% and 15% respectively. Haveman and Wolfe (1995) calculate opportunity cost as forgone income of full time working females owing to having children.\(^{20}\) I follow a similar approach and calculate $b_n$ as the fraction of forgone family earnings.


\(^{20}\)Knowles (1999) considers the opportunity cost calculated by Haveman and Wolfe (1995) an upper bound because some of childcare can be considered as leisure (see Godbey and Robinson (1999)).
gone labor work (hence income) of a family owing to children. CPS has information on weekly labor hours and number of weeks worked in a year. Therefore, I calculate the average of annual labor hours of families with different sizes (see Table 2). The opportunity costs are lower comparison to Haveman and Wolfe (1995) because the sample consists of males also. However, when I particularly compare the opportunity cost of females, the measures are in line with Knowles (1999).

<table>
<thead>
<tr>
<th>Families with</th>
<th>0 children</th>
<th>1 child</th>
<th>2 children</th>
<th>2+ children</th>
</tr>
</thead>
<tbody>
<tr>
<td>total labor ($l_n$)</td>
<td>4258</td>
<td>4162</td>
<td>4122</td>
<td>4098</td>
</tr>
<tr>
<td>$b_n$</td>
<td>0</td>
<td>0.022</td>
<td>0.032</td>
<td>0.037</td>
</tr>
<tr>
<td>sample size</td>
<td>2,940</td>
<td>5,818</td>
<td>13,013</td>
<td>19,545</td>
</tr>
<tr>
<td>female labor ($l_n$)</td>
<td>2044</td>
<td>1937</td>
<td>1863</td>
<td>1834</td>
</tr>
<tr>
<td>$b_n$ (for females)</td>
<td>0</td>
<td>0.052</td>
<td>0.088</td>
<td>0.103</td>
</tr>
<tr>
<td>sample size</td>
<td>1,088</td>
<td>2,299</td>
<td>5,095</td>
<td>7,731</td>
</tr>
</tbody>
</table>

Table 2: Opportunity Cost of Childcare

Data: CPS 2005-2014. $l_n$ represent the weighted average hours per year devoted to the earnings by two spouses. The sample is restricted to married, 35-45 years aged, and working households. Opportunity cost of $n$ children is calculated as $b_n := \frac{l_n - l_0}{l_0}$. Since $b_{2+} \simeq b_2$, I work with $b_{2+}$.

One might consider opportunity cost across different income groups are lower, especially for the top income families. However, the opportunity cost is, interestingly, very high for high income families (see Table 3). This result can be attributed to positive correlation between childcare and education level of high income earners (see Ramey and Ramey (2009)).

<table>
<thead>
<tr>
<th>Quintiles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Income:</td>
<td>$45,726$</td>
<td>$69,416$</td>
<td>$89,452$</td>
<td>$113,703$</td>
<td>$157,960$</td>
</tr>
<tr>
<td>$b_n$</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 3: Opportunity Cost across Income Groups

### 4.3 Estimations of Productivity Distribution and $\rho$

This subsection finds estimates related to families’ non-observable characteristics. First, I estimate probability distribution of productivities. Second, I estimate a key parameter on the marginal utility of a child.
4.3.1 Estimation of the Distribution of Productivities: $f(\theta)$

Assuming the data was generated by suboptimal a tax equilibrium, the optimality condition (2) of the quasi-linear preference structure allows me to find productivities of families:

$$\theta = \frac{z_n}{(1 - T'(z_n))^\epsilon - b_n}. \quad (11)$$

Note that CPS has information about family structure and detailed sources of family income and taxes. I add (federal) earned income tax rates to the reported federal marginal taxes to find effective marginal rates.\(^{21}\) Using $b_n$ values from Table 2 and the weights assigned to the family (given by the data), I show the distribution of productivities in Figure 4.

![Figure 4: Productivity Distribution: $f(\theta)$](image)

4.3.2 Estimation of the Marginal Utility Parameter: $\rho$

An important contribution of this paper is to estimate family preferences on (extensive) number of children. I assume that $\beta \sim [1, \infty)$ 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.\(^{22}\) Unfortunately, the literature on (extensive) elasticity of fertility with respect to marginal cost of a child is

---

\(^{21}\)Earned income credits are reported, however, their marginal effects are not. I use the information on EITC for years 2005-2014 and impose its marginal effects.

\(^{22}\)Given a $\theta$, $1 - P(\beta_N(\theta))$ is the fraction of $N$ child families and the marginal cost of $N^{th}$ child is $\beta_N(\theta)$.
very sparse. In an interesting work, Cohen, Dehejia, and Romanov (2013) use an Israeli data and estimate a price elasticity of $-0.54$ for a child and in particular the estimate for the third child is $-0.42$. The elasticity with respect to the benefit of a child, however, has been studied by many works (see Table 4).

<table>
<thead>
<tr>
<th>Work</th>
<th>Region of Data</th>
<th>Elasticity Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen, Dehejia, and Romanov (2013)</td>
<td>Israeli</td>
<td>0.19</td>
</tr>
<tr>
<td>Milligan (2005)</td>
<td>Canadian</td>
<td>0.11</td>
</tr>
<tr>
<td>Laroque and Salanie (2008)</td>
<td>French</td>
<td>0.20</td>
</tr>
<tr>
<td>Whittington, Alm, and Peters (1990)</td>
<td>US</td>
<td>0.12-0.23</td>
</tr>
</tbody>
</table>

Table 4: Benefit Elasticity of Having Children

Table 4 suggests benefit elasticities are similar across countries. Therefore, I assume cost elasticities are also similar and follow Cohen, Dehejia, and Romanov (2013) and set $\lambda = 0.4$.\(^{23}\)

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 (5) implies that the theoretical probability of having $n$ children is:

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

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

$$\rho \in \arg \max L = \prod_{n=0}^{2} \prod_{j=0}^{100} P_n(\theta_j)^{\pi_n(\theta_j)}. \quad \text{(ML)}$$

The solution of (ML) is $\hat{\rho} = 3.95$ where $\hat{\sigma}_\rho = 1.64$.

5 Numerical Analysis on the Optimal Tax System

In this section, I compute the optimal policy using the estimates derived in Section 4. To compare statutory and the optimal tax schedule, I set government spending as $G =$ $11,383, which is the statutory tax amount collected by the government from the sample

\(^{23}\)I do robustness analysis for $\lambda$ in Appendix A.4.
I numerically solve the government’s problem (PMDP), which is an optimal control problem and its Hamiltonian is stated in Appendix, at my selected and estimated parameters using the numerical solver GPOPS-II software.\textsuperscript{26}

Figure 6 compares optimal and statutory tax schedules of the 2011 tax year. Figure 5 shows that optimal income taxes are slightly higher for incomes between $70,000 - $130,000. On the other hand, lowest income households receive around $4,000 more under the optimal tax schedule.

Figure 5: Statutory versus Optimal Income Tax Schedules

Figure 6a and 6b show how optimal and statutory tax credits change across income. While the US tax credits are decreasing over income, the optimal credits are U-shaped.\textsuperscript{27} The main force behind the difference is the time cost of child-rearing. Increasing credits for high income families (with children) do not only rise their consumption but also decreases their marginal tax rates which reduces the distortion on their labor margin, and hence incentivizes them to work harder. The shape of statutory credits implies that the US government only focuses on the goods cost and does not deal with the time cost.

\textsuperscript{24}Statutory taxes are calculated by TAXSIM 9.2 version where TAXSIM is the NBER’s program which calculates US Federal income tax liabilities across years from individual data.

\textsuperscript{25}Per capita taxes reported in CPS is $10,384.

\textsuperscript{26}GPOPS-II is a flexible software program for solving optimal control problems, see Patterson and Rao (2014).

\textsuperscript{27}Small jumps in the statutory credits are owing to the effect of personal exemption when marginals change.
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 main source of this result is economies of scale on the impact of child-rearing costs 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 the optimal credit schedule.

These results suggest that the optimal tax schedule is strictly different than the current US tax system. Designing the tax schedule with U-shaped tax credits may improve social welfare. I find that the welfare gain from implementing the optimum is 1.3% in terms of equivalent increase in consumption for all families. In addition, 50% of families are better of with the optimal tax schedule. This suggests that the optimal tax schedule does not only increase social welfare but also is very attractive to families.

6 Robustness Analysis

In this section, I analyze robustness of U-shaped tax credits. Among many, I show four different robustness analysis here. First analysis is about government’s objective. Figure 7 shows the tax credit schedules for Utilitarian and Rawlsian government. Both schedules are U-shaped according to income. In contrast to the Utilitarian government solution, Rawlsian government provides more credits, especially to the lowest income

---

28 The other robustness analysis can be found in Appendix A.4.
29 Rawlsian government’s objective is to maximize lowest utility in the economy.
parents, because, the Rawlasian government has stronger redistribution motives at the bottom of income distribution.

![Figure 7: Government Preference Analysis](image)

Second, I put a constraint about child evolution over the productivities in the government’s problem. Let \( n(\theta) \) represent the average number of children for productivity \( \theta \). Under the optimal tax schedule stated in Figure 6, total number of children in the economy is \( \%9.7 \) less than the data. Because of the opportunity cost of childcare, the solution of (PMDP) implies \( n(\theta) \) is decreasing (see Figure 8a). On the other hand, \( n(\theta) \) of data circles around its average\(^{30}\). With the \( n(\theta) = 1.49 \) for all \( \theta \in \Theta \) constraint, the optimal tax credits are stated in Figure 8b. Credits are much higher and increasing over income. The reason is that the government has to provide more incentives for families to have more children. Also, the government enhances incentives towards higher income families because the opportunity cost of childcare becomes heavier towards them. With the same reason, high income families receive more credits for second child than the first.

\(^{30}\)The average of number of children of the sample is 1.77. Families with more than two children are considered as families with two children. Therefore, the average is 1.49.
Third analysis focuses on households’ preferences. I work with \( U = \frac{(c-h(\tilde{z}+b_n)\theta)^{1-\sigma}}{1-\sigma} + m(n, \beta) \). I let \( \sigma = 0.5 \). Figure 9a and 9b show the tax credit schedule for Utilitarian and Rawlasian government, respectively. The credit schedules are U-shaped and very similar with Figure 7.

In the fourth analysis, I study tax credits for single mothers. I adjust estimation parameters: \( \varepsilon = 0.8 \) (see Blundell, Pistaferri, and Saporta-Eksten (2012)), \( e_1 = $5,000 and
$e_2 = 8,500$, and $b_1 = 0.035$ and $b_2 = 0.039$.  \[^{31,32}\] Figure 10a shows the productivity distribution for single females. The solution of (PMDP) with the adjusted parameters implies that the optimal credits are U-shaped (see Figure 10b).

These four analysis suggest that U-shaped tax credit schedule is very robust. In addition to these analysis, I make more robustness analysis in Appendix A.4. For different values of \(\lambda, e_n, b_n,\) and \(N\) as well as for different range of ages (of spouses in data), the optimal tax credits are U-shaped.

7 Conclusion

This paper studies optimal income taxation and child tax credits in a static Mirrlees model in which potential parents privately observe their characteristics: child tastes and productivities. 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 informational friction. 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.

\[^{31}\]According to Table 7 of Lino (2012), low income single mothers spend $10,000 for one child and $17,000 for two children.

\[^{32}\]A single woman without children works 2196 hours per year on average. A single mother with one child and two children work 2118 and 2111 hours, respectively.
Income taxes are designed to redistribute from high to low income families and child tax credits decrease tax liabilities of parents who bear 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 them. 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 in income. Quantitatively, I find that the optimal credits are decreasing in the first three quarters 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 costs.

This paper sheds light on the optimal income taxation including the child benefits for families who have multidimensional privately observed characteristics. 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 productivities and child tastes, can be linked with the age of the parents and, therefore, the effect of optimal taxes on the time of fertility 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 parents 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 on designing the optimal provisions via costs. For example, policies that provide a high-quality childcare in return of goods might be tempting for high income families. This extension can also examine the current debate in the US on universal childcare 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)

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

The first-order conditions are:

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

Also, the co-states are:

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

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:

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

where

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

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

\[33\text{Let } \beta_0 = \bar{\beta}, \text{ and } \beta_{N+1} = \bar{\beta}.
\]

\[34\text{T}_n(\theta) \text{ is the taxes paid by } \theta-\text{productivity families with } n \text{ children. Let } \Delta T_{-1}(\theta) = 0 \text{ and } \Delta T_N(\theta) = 0.
\]
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.

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

$$
\frac{T'_n(\theta)}{1 - T'_n(\theta)} = \frac{1}{\epsilon_n(\theta)} \times \frac{1}{\theta f(\theta)(P(\beta_{n+1}(\theta)|\theta) - P(\beta_n(\theta)|\theta))} \times \\
\int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} (P(\beta_{n+1}(\theta')|\theta') - P(\beta_n(\theta')|\theta')) \right]
$$

$$+ \sum_{j=n}^{n+1} \Delta T_{j-1}(\theta') p(\beta_j(\theta')|\theta') \frac{\partial \beta_j(\theta')}{\partial V_n(\theta')} ] u'(c_n(\theta)) f(\theta') d\theta' \quad \forall n \in N$$

where $\epsilon_n(\theta)$ is the compensated elasticity of income of $n$–child families with respect to marginal tax rates.$^{35}$

---

$^{35}$Let $\Delta T_{-1}(\theta) = 0$ and $\Delta T_N(\theta) = 0.$
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.\textsuperscript{36} 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 childcare 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 childcare 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 a 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 (see Table 5).\textsuperscript{37} 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 ≤ $6,070</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>$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


\textsuperscript{36}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.

\textsuperscript{37}According 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. 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 \( \gamma = (\beta, \theta) \in \mathcal{B} \times \Theta = \Gamma \) 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 \((c(\gamma), z(\gamma), n(\gamma))\) 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 \( \mathcal{U}(c(\gamma'), z(\gamma'), n(\gamma')) \) be the utility of the \( \Gamma \) family who report \( \gamma' \) and gets \((c(\gamma'), z(\gamma'), n(\gamma'))\).

Family characteristics are twofold, 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 (4) and critical tastes for children (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 3.** Any truthful mechanism \((c(\gamma), z(\gamma), n(\gamma))\) can be replaced by a new truthful mechanism \( \{c_n(\theta), z_n(\theta)\}_{\theta \in \Theta} \) such that \( \forall \theta \in \Theta \) and \( \forall n \in \mathcal{N} \), there is a \( \beta_n(\theta) \) such that if \( \beta \in (\beta_n(\theta), \beta_{n+1}(\theta)) \),\(^{38}\) then \( \mathcal{U}(c_n(\theta), z_n(\theta), n, \gamma) \geq \max \mathcal{U}((c(\gamma'), z(\gamma'), n(\gamma')) \), \( \gamma' \)). The new mechanism generates same utility as the original mechanism and the government collects as much taxes as the original mechanism.

**Proof.** For each \( \theta \), partition the set \( \mathcal{B} \) into \( N + 1 \) sets such that if \( \beta \in \mathcal{B}_j \) then \( n(\beta, \theta) = j \)

\(^{38}\) I let \( \beta_0 = \beta \) and \( \beta_{N+1} = \beta \).
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 \( \beta, \beta' \in \mathcal{B}_j \), the truthfulness of the original mechanism implies:

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

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 \( u(c(\beta, \theta)) - h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta \) 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 \( t_j = \sup_{z \in Z_j(\theta)} z - u^{-1}(V_j(\theta) + h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta) \). Note that \( z - u^{-1}(V_j(\theta) + h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta) \) is a weakly concave function in \( z \) and reaches maximum for a \( z \) value and goes to \(-\infty\) when \( z \rightarrow \infty \). So there is a \( z_j(\theta) \in \bar{Z}_j(\theta) \) such that \( t_j = z_j(\theta) - u^{-1}(V_j(\theta) + h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta) \).\(^{39}\) Define \( c_j(\theta) := u^{-1}(V_j(\theta) - h\left(\frac{z(\beta, \theta)}{\theta} + b_j\right) \theta) \).

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 \( j \in \mathcal{N} \).

Finally, I define \( \beta_n(\theta) := M_{n-1}(V_n(\theta) - V_{n+1}(\theta)) \) where \( M_n(\beta) := m(n + 1, \beta) - m(n, \beta) \) for all \( n \in \mathcal{N} \).\(^{40}\) \( \beta_n(\theta) \) are the critical tastes for children for each \( \theta \) and for each \( n \).

Note that truthfulness of original mechanism implies: for all \( \beta \in \mathcal{B}_j(\theta) \) 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_{\bar{n}}^{-1}(V_{j+1}(\theta) - V_j(\theta)).\(^{41}\) Therefore \( \mathcal{B}_j(\theta) = (\beta_{j-1}(\theta), \bar{\beta}_j(\theta)). \)

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 \( \theta, \theta', \beta \in \mathcal{B}_j(\theta), \beta' \in \mathcal{B}_j(\theta') \):

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

\(^{39}\)\(Z_j(\theta)\) is the closure of the \( Z_j(\theta)\)

\(^{40}\)Let \( \beta_0 = \beta \) and \( \beta_{N+1} = \bar{\beta} \).

\(^{41}\)Note that let \( m \) to be concave in its first dimension and therefore \( \beta_{j-1}(\theta) < \beta_j(\theta) \).
where the inequality is from the truthfulness of the initial mechanism. 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 \( j \in \mathcal{N}. \)

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

**Definition 2.** \( \{z_n(\theta)_{\theta \in \Theta}\}_{n \in \mathcal{N}} \) is implementable if and only if there exist transfer functions \( \{c_n(\theta)_{\theta \in \Theta}\}_{n \in \mathcal{N}} \) such that \( \{(c_n(\theta), z_n(\theta))_{\theta \in \Theta}\}_{n \in \mathcal{N}} \) 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 \( \{z_n(\theta)_{\theta \in \Theta}\}_{n \in \mathcal{N}} \) is implementable if and only if \( z_n(\theta) := \frac{\partial z_n(\theta)}{\partial \theta} \geq 0. \)

**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 \( \theta, \theta'. \)

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

\[^{42}\text{Note that } (c_j(\theta'), z_j(\theta')) \text{ is in the closure of the set } (c(\beta, \theta), z(\beta, \theta))_{\beta \in B_j(\theta)}.\]
Now let \( z_n \geq 0 \). Similarly, using the one-dimensional result, there is \( c_n(\theta) \) such that
\[
\begin{aligned}
u(c_n(\theta)) - h\left(\frac{z_n(\theta)}{\theta} + b_n\right)\theta &\geq u(c_n(\theta')) - h\left(\frac{z_n(\theta')}{\theta} + b_n\right)\theta \quad \text{for all } \theta, \theta'.
\end{aligned}
\]

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.

### 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 last 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 individual elasticities.

To figure out family income elasticity, I need \( \varepsilon_f, \varepsilon_m \), 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.\(^4\)

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 \( \varepsilon_f = 0.8 \).

\(^4\)Blundell, Pistaferri, and Saporta-Eksten (2012) finds that the Frisch elasticity of married men is 0.4. For different models the value goes up to 0.6.
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 11). Hence, \( \varepsilon = 0.6 \times 0.4 + 0.4 \times 0.8 = 0.56 \). Note that this ratio 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).

![Figure 11: Female Share of Family Income](image)

**Figure 11: Female Share of Family Income**

### A.4 Additional Robustness Analysis

This subsection provides additional robustness analysis for U-shaped tax credits. Figure 12a and 12b show the optimal tax credits for different values of elasticity of having second child with respect to her marginal cost, \( \lambda = 0.35 \) and \( \lambda = 0.5 \), respectively. The optimal child tax credits are U-shaped in income.

Figure 12c shows that the optimal tax credits are decreasing with income when the child-rearing requires only goods and the goods amount are set to \( e_1 = $11,000 \) and \( e_2 = $19,000 \). On the other hand, when child-rearing requires only 5% of parental working time for each child as Knowles (1999) suggests, the credits are increasing in income almost everywhere.

When I relax the age restriction of spouses in families from 35-45 to 25-45, optimal tax credits are still U-shaped (see Figure 12e).

Finally, I consider a case where families can have up to three children. Figure 12f shows that optimal tax credits are U-shaped in income and decreasing by family size.\(^{44}\)

\(^{44}\)I numerically solve the government’s problem (i.e. PMDP), using the numerical solver DIDO version 7.3.7. For details on the solution algorithm, refer to Ross and Fahroo (2003).
(a) Elasticity of Having Children

(b) Elasticity of Having Children

(c) Only Goods Cost

(d) Only Time Cost

(e) Age Restriction of Spouses

(f) $N = 3$ Children
A.5  NUMERICAL SOLUTION

Hamiltonian requires calculations for inner integrals. I first theoretically calculate a version of an integral:

$$\int \frac{(V - A\beta)^{1-\sigma}}{1-\sigma} \lambda \beta^{-\lambda-1} d\beta = \frac{\beta^{-\lambda} V^{-\sigma}}{(1-\sigma)(1-\lambda)} \times \left( (\lambda - 1) {}_2F_1(-\lambda, \sigma; 1 - \lambda; \frac{A\beta}{V}) - A\lambda {}_2F_1(1 - \lambda, \sigma; 2 - \lambda; \frac{A\beta}{V}) \right)$$

where $\, _2F_1$ is Gauss hypergeometric function. Matlab’s command for hypergeometric functions is very slow. Instead, I use $\, _2F_1$’s definition for calculations:

$$\, _2F_1(a, b; c; y) = 1 + \frac{ab}{1!c} y + \frac{a(a+1)bb(b+1)}{2!c(c+1)} y^2 + \ldots = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n y^n}{(c)_n n!}.$$  

When the additional terms are less than $10^{-15}$, I stop the loop. This calculation is very accurate and quick that has been used many times to calculate social welfare in the government’s problem. Note that $A$ changes by family size. For example, $A = 2^\rho$ for non-parents.