Abstract

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

JEL-Classification: H21, H53, D82, J13

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

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

This paper focuses on the optimal design of income taxes and child tax credits. I make both theoretical and quantitative contributions. On the theoretical side, I explore the forces shaping optimal income taxes and child tax credits. The former is redistributive towards low-earning families. The latter reduces the income tax liabilities of those with children who are made monetarily worse off by child-rearing. On the quantitative side, I study the key forces behind the credits. While the goods cost reduces the welfare of low-earning families more relative to the high, the time cost reduces the welfare of the high-earning families more relative to the low. These impacts suggest that the goods cost is a motive for more provision to poor families, on the other hand, the time cost is a motive for more provision to the wealthier. As a result, the optimal child tax credits are U-shaped with respect to income.

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

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

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

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

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

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

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

The remainder of the paper is organized as follows. After a brief review of the literature, I provide an institutional background for taxation of families in Section 2. I introduce the model in Section 3. I derive the optimal tax schedule and also show why the conventional tax formula should be adjusted with new terms. 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 vast literature on fertility theories. The related works to my paper study well-known empirical evidence that fertility is negatively correlated with income. Schultz (1986) is an example of such works, which explicitly focuses on wages of spouses and relates the evidence with the changes in wage gap. Recently, Jones, Schoonbroodt, and Tertilt (2010) give a brilliant summary of fertility theories. They state that child-rearing costs are on the focus of many studies. Empirically, Haveman and Wolfe (1995) work on the goods and the time costs. They find that these costs incurred by parents and the government are around 14.5% of 1992 GDP. Two-thirds of the costs is financed by parents. In addition, 82% of parental costs is goods cost which includes expenditures on food, housing, health care, and clothes. There are many studies that work on the impact of the goods cost. Golosov, Jones, and Tertilt (2007) and Hosseini, Jones, and Shourideh (2013) are recent examples. The former paper studies the efficiency of the future allocations and the latter focuses on the consumption inequality in the long run.

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

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

My paper also contributes to the public finance literature, which is based on the trade-off between efficiency and equity. The trade-off arises because agents’ earning abilities are their private information. In my paper, not only earning abilities but also tastes for children are families’ private information, and hence the friction in the information is two dimensional. Because of the technical difficulties of multi screening problems, there are
few works study such an environment.\footnote{Baron and Myerson (1982) and Rochet and Chone (1998) provide some additional requirements to solve a multi screening problem 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 family taxation in which primary earner’s earning ability and secondary earner’s work cost are families’ private information. They show that marginal income tax rates of the primary earner should be smaller if his or her spouse works. The latter studies an environment in which workers have private information regarding their earning abilities and their taste of work and make a labor decision extensively and intensively. They provide a rationale for non-negative marginal rates. Unlike the studies above, I focus not only on the marginal rates but also on tax liabilities of families to study child tax credits. Moreover, both studies have only two categories of agents. In this paper, I derive optimal taxes for an arbitrary number of family sizes.

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

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

2 Institutional Background

There are around 80 mean-tested federal programs providing for different needs of families such as cash, food, housing, medical care, and social services. Almost 50% of the budget for welfare programs is spent for families with children.\footnote{Refer to Chart 3 of \url{http://budget.house.gov/uploadedfiles/rectortestimony04172012.pdf}} In this section, I give information about some of the cash programs: Child Tax Credit, Earned Income Tax Credit, and Child and Dependent Care Tax Credit. These are the main cash assistance programs provided to families with children.

2.1 Child Tax Credit

The Child Tax Credit (CTC) was enacted as a temporary provision in the Taxpayer Relief Act of 1997. A credit of $400 is given to families for each qualifying children and the
credit was refundable only for families with more than two children.\footnote{See \url{http://www.irs.gov/uac/Ten-Facts-about-the-Child-Tax-Credit} for eligibility tests.} The credit has gradually increased to $1,000 from 2001 to 2010 by the Economic Growth and Tax Relief Reconciliation Act of 2001. Moreover, the refundability is extended to all families. This refundable tax credit is called Additional Child Tax Credit. If a family has less tax liability than their child tax credit, they may get the minimum of unclaimed credits and 15\% of their income above $3,000. Because of the changes in the eligibility conditions and the credit amount, the federal spending for CTC increased from 0.2\% to 0.4\% of GDP between 2000 and 2010. Currently, the credit decreases for high income families. For example, the credit is reduced by $50 for each $1,000 when aggregate gross income is above $110,000 for married tax payers filing jointly. Finally, the credit has become permanent by the American Taxpayer Relief Act of 2012.

2.2 Earned Income Tax Credit

The Earned Income Tax Credit (EITC) is another program for working families. The literature on the EITC is voluminous and cannot be fully reviewed here. I refer to Hotz and Scholz (2003), and the references there. Here, I focus on how the credit differs with family size. The maximum credit and phase in and out rates drastically change with the number of children in families (see Table 5 in Appendix A.5). Figure 1 plots the EITC for 2014. Families with more children are given more credits.

\begin{center}
\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Earned Income Tax Credit in 2014}
\label{fig:eitc}
\end{figure}
\end{center}

\footnote{See \url{http://www.irs.gov/uac/Ten-Facts-about-the-Child-Tax-Credit} for eligibility tests.}
2.3 Child and Dependent Care Tax Credit

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

To conclude this section, I focus on the functionality of these welfare programs. According to the Tax Policy Center, which is a joint venture of the Urban Institute and the Brookings Institution, 6.6 million families are qualified for the CDCTC in 2010, more than 26 million taxpayers received the EITC in 2015, and 38 million families claimed from the CTC in 2013. More families benefit from the CTC because its eligibility requirement is more relaxed than the other welfare programs. Moreover, the CTC has become one of the most expensive welfare programs for the US government (see Figure 2).

In the following section, I introduce a static model, in which families face shocks on earning abilities and tastes for children and simultaneously decide how much income to generate and how many children to have. Child-rearing requires both goods and parental time, and these costs shape the optimal child tax credits.

---

3 Model

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

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

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

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

$$U(c, z, n, \theta, \beta) = u(c) - h(z, b_n, \theta) + m(n, \beta)$$

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

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

3.1 Family Problem

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

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

(2)

\[ c_n = z_n - T_n(z_n) - e_n. \]  

(3)

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

\[ V_n(\theta) := u(c_n) - h(z_n, b_n, \theta). \]  

(4)

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

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

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

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

(5)

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

In Figure 3, I illustrate the child choice graphically for \( n = 0, 1 \). For a particular ability level, the families with \( \beta \in (\beta_0, \beta_1(\theta)) \) choose to have no children because the marginal benefit of one child is less than its costs. For \( \beta = \beta_1(\theta) \), the benefit and cost having one child is equalized. When \( \beta \in (\beta_1(\theta), \beta_2(\theta)) \), the families decide to have one child because the marginal benefit of one child is higher than its cost and the marginal benefit of second

\(^8\text{I fix} \beta_0(\theta) = \underline{\beta} \text{ and } \beta_{N+1} = \bar{\beta}.\)
children is less than the marginal cost of second children.

\[ \beta \\
\beta_2(\theta) \\
\beta_1(\theta) \]

\[
\begin{align*}
&1 \text{ child families: } V(\theta, \beta, 1) = V_1(\theta) + m(1, \beta) \\
&0 \text{ child families: } V(\theta, \beta, 0) = V_0(\theta) + m(0, \beta)
\end{align*}
\]

Figure 3: Critical Child Taste Levels

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

### 3.2 The Government’s Problem

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

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

#### 3.2.1 Mechanism Design Problem

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

\[
\max_{c(\beta, \theta), n(\beta, \theta), z(\beta, \theta)} \int_{\Theta} \int_{B} \Psi(U((\beta, \theta))) p(\beta|\theta) f(\theta) d\beta d\theta \quad \text{(MDP)}
\]
subject to the incentive constraints

\[ U((\beta, \theta)) \geq \max_{(\beta, \theta') \in B \times \Theta} U((\beta, \theta), (\tilde{\beta}, \theta')) \quad \forall (\beta, \theta) \in B \times \Theta \quad (6) \]

and the resource constraint

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

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

Equation (6) states that the government should prevent mimicking via two channels, earning abilities and tastes for children. Consequently, \( \infty \times \infty \) possible deviations should be handled, which is hard to solve. To handle such deviations, I follow the arguments of the family problem solution and use the definitions of indirect utility and threshold tastes for children. First, for a given family size, the government uses a first-order approach to prevent deviation via earning abilities:

\[ V_n(\theta) := \frac{\partial V(\theta)}{\partial \theta} = -h \left( \frac{z_n}{\theta} + b_n \right) + h' \left( \frac{z_n}{\theta} + b_n \right) \frac{z_n}{\theta} \geq 0 \quad (7) \]

for all \( n = 0, 1, \ldots, N \). Second, the deviation in tastes are handled by (5). First of all, Equation (5) directly implies that mimicking the tastes of families with different sizes does not make families better off. In addition, for a particular family size, mimicking the tastes for children does not alter the children choice and hence does not change the family utility. As a result, the possibility of double deviation in the Equation (6) can be handled via Equation (4) and (5).

I use the Equation (4) and (5) and adjust the objective function and constraints. In addition, the constraint (6) of the problem MDP is replaced by Equation (4), (5), and (7). This new problem is a sophisticated version of the original mechanism design problem, and without loss of generality, I call the new problem as the “pseudo-mechanism design problem”.

3.2.2 Pseudo-Mechanism Design Problem

The government problem solves the following problem:

\[
\max_{T_n(\theta)} \int_{\Theta} \int_{\beta} \Psi (V_0(\theta) + m(0, \beta)) p(\beta|\theta)f(\theta)d\beta d\theta + \int_{\Theta} \int_{\beta_1(\theta)}^{\beta_2(\theta)} \Psi (V_1(\theta) - m(1, \beta)) p(\beta|\theta)f(\theta)d\beta d\theta \\
+ \ldots + \int_{\Theta} \int_{\beta_N(\theta)}^{\tilde{\beta}} \Psi (V_N(\theta) + m(N, \beta)) p(\beta|\theta)f(\theta)d\beta d\theta
\quad (PMDP)
\]
subject to Equation (4), (5), and (7) and the resource constraint:

\[
\int_\Theta \int_{\hat{\beta}} T_0(\theta) p(\beta|\theta) f(\theta) d\beta d\theta + \int_\Theta \int_{\hat{\beta}_1(\theta)} T_1(\theta) p(\beta|\theta) f(\theta) d\beta d\theta \\
+ \ldots + \int_\Theta \int_{\hat{\beta}_N(\theta)} T_N(\theta) p(\beta|\theta) f(\theta) d\beta d\theta \geq G 
\]  

(8)

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

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

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

**Proof.** See Appendix A.1.

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

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

\[
\frac{T_n'(\theta)}{1 - T_n'(\theta)} = \frac{1}{\varepsilon_n(\theta)} \times \frac{1}{\theta f(\theta) (P(\beta_{n+1}|\theta) - P(\beta_n|\theta))} \times \\
\int_\Theta \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} (P(\beta_{n+1}|\theta') - P(\beta_n|\theta')) \right] \\
+ \Delta T_{n-1}(\theta') p(\beta_n|\theta') \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta') p(\beta_{n+1}|\theta') \frac{\partial \beta_{n+1}}{\partial V_n} u'(c_n(\theta)) f(\theta') d\theta' 
\]

(9)

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

The formal proof is given by Appendix A.2. Here, to provide intuition, I follow Saez (2001) and show the heuristic proof of the Proposition 1. For simplicity, I focus on one-child families and assume \( u(c) = c \).

Suppose that the government increases the taxes of one-child families with \( \theta' \geq \theta \) earning abilities by \( dT \) (see Figure 4). This change creates three effects. First, since the one-child families consume less, there will be a welfare loss for the society by \( g_1(\theta') \) for \( \theta \)

\(^9\)I let \( \Delta T_{-1}(\theta) = 0 \) when \( n = 0 \) and \( \Delta T_N(\theta) = 0 \) when \( n = N \).
each dollar of $dT$ for all $\theta' > \theta$ where

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

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

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

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

Second effect is on the income decision of families whose abilities are in $[\theta, \theta + d\theta]$. To increase taxes by $dT$, the government should increase the marginal rates of families with $[\theta, \theta + d\theta]$ by $\tau = \tilde{\tau}_1\frac{z_1}{\theta}$, where $\tilde{\tau}$ represents the change in the marginal tax rates on income (see Figure 4).\(^{11}\) This increment creates a behavioral effect, i.e., the families in the small band decrease their income by $dz = \frac{z_1\epsilon_1(\theta)\tau}{1 - T_1'(\theta)}$, where $\epsilon_1(\theta) := \frac{\partial \log z_1}{\partial \log (1 - T_1'(z_1))}$ is the elasticity of income with respect to marginal tax rates. Combining the terms gives the first

\(^{10}\)See Equation (12) for a general definition.

\(^{11}\)To change the marginal rates over abilities by $\tau$, the marginal rates on income should increase by $\tilde{\tau}$. 

---

Figure 4: Increase in $T_1(\theta)$ by $dT$
behavioral effect is:

\[ dB_1 = -T'_1(\theta)dzf(\theta)d\theta = -dT_1(\theta) \frac{T'_1(\theta)}{1 - T'_1(\theta)} \epsilon_1(\theta)\theta [P(\beta_1|\theta) - P(\beta_2|\theta)] f(\theta). \]

density of 1 child-families

If the number of children was exogenously given, there would not be any extra effect. Hence, if the original mechanism was optimal, these effects should sum up to zero: \( dG + dB_1 = 0 \). In this situation, the optimal tax formula would then be very similar to that of Mankiw and Weinzierl (2010). However, the number of children to have is a choice in my set up, and \( dT \) affects the optimal number of children of families whose tastes for children are in the neighborhood of \( \beta_1(\theta) \) and \( \beta_2(\theta) \) (see Figure 5).

The one-child families whose tastes for children are in the neighborhood of \( \beta_1(\theta) \) prefer to have no children after the increase in their taxes. As a result, their tax liabilities are changed by: \( \Delta T_0(\theta') := T_0(\theta') - T_1(\theta') \) for all \( \theta' \geq \theta \). For a particular \( \theta' \), the effective change is: \( \Delta T_0(\theta') \frac{\partial \beta_1(\theta')}{\partial V_1} p(\beta_1|\theta') f(\theta') \) where \( \frac{\partial \beta_1(\theta')}{\partial V_1} \) is the mechanical effect of \( V_1(\theta') \) on \( \beta_1(\theta') \) and \( p(\beta_1|\theta') f(\theta') \) 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: \( \Delta T_1(\theta') \frac{\partial \beta_2(\theta')}{\partial V_1} p(\beta_2|\theta') f(\theta') \).

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

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

Together with the second behavioral effect, the original mechanism is optimal if \( dG + dB_1 + dB_2 = 0 \). This equality gives the equation in Proposition 1 when \( u(c) = c \) for \( n = 1 \).

---

12 Taxes for different categories can be considered as taxes for families with different sizes.

13 The new threshold tastes are represented by \( \tilde{\beta}_1(\theta) \) and \( \tilde{\beta}_2(\theta) \). See Figure 5.
Note that, these procedures can be applied for any \( n = 0, 1, \ldots, N \) to find the optimal marginal tax rates of families with \( n \) children.

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

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

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

**Proof.** Define elasticity as \( \varepsilon_n := \frac{\log \frac{\partial z_n}{\partial (1-T_n')}}{\log \frac{\partial z_n}{\partial (1-T_n')}} = \frac{1-T_n'}{z_n} \frac{\partial z_n}{\partial (1-T_n')} \). The first-order condition for income is: \( (1-T_n') = h'(\frac{z_n}{\theta} + b_n) \). Taking the derivative with respect to \( (1-T_n') \) and rewriting yields: \( \varepsilon_n = \frac{h'(\frac{z_n}{\theta} + b_n)}{h'(\frac{z_n}{\theta} + b_n)} = \varepsilon(1 + \frac{b_n}{z_n\theta}) \).

It is straightforward to see that \( \varepsilon_n(\theta) \) depends on \( z_n \), and hence the elasticity of income of parents is endogenous. Moreover, \( \varepsilon_1(\theta) > \varepsilon_0(\theta) \) because childless families do not spend time on child care, i.e. \( b_0 = 0 \). This result is in line with Blundell, 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: \( f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta)) \). This term is endogenous because the family size is a choice. The term provides information about the underlying tastes for children. The government knows that families with tastes in \((\beta_n(\theta), \beta_{n+1}(\theta))\) will generate same income and will have same number of children if they face same marginal tax rates.

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

Next, I provide an interpretation of the conventional and novel terms that appear in the tax formula.
**Interpretation of the terms:** The interaction of the terms in Equation (13) is complex. Here, I go over term by term and provide a basic interpretation of each term. First, the elasticity, $\varepsilon_n$, is reciprocally correlated with the marginal taxes. Note that the marginal taxes create distortions on income decision and the distortions are higher for families with higher elasticity of income. The distortions create a deadweight loss for the economy and the government considers this loss and reduces the marginal taxes of those with higher income elasticity.

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

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

Finally, I focus on the tax difference terms. The first term, $\Delta T_{n-1}$, tightens the incentive constraints (see Equation (5)). The government relaxes such constraints by decreasing the marginal rates of $n$– children families. On the other hand, $\Delta T_n$ relaxes the incentives, and hence government increases marginal rates.$^{14}$

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

### 4 Quantitative Analysis

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

According to the empirical labor market literature, the effect of non-labor income on labor is small (see Blundell and Macurdy (1999)). In addition, to understand the relationship between labor income and number of children, it is natural to eliminate the non-labor income effect on labor. Therefore, I assume that families have a quasi-linear preference in consumption: $u(c) = c$. Moreover, I assume that childless families a constant elasticity of income with respect to marginal rates: $h(z, \theta) = \frac{\varepsilon}{1 + \varepsilon} \left( \frac{z}{\theta} \right)^{\varepsilon + 1} \theta$ where $\varepsilon := \frac{\partial \log z}{\partial \log (1 - T')}$ is the elasticity of the total family income with respect to marginal taxes. The estimate for the elasticity of family income requires attention, because the literature on elasticity of income is based on individual levels. I study individual elasticities to figure out the family elasticity in Appendix A.3. In the benchmark, I use $\varepsilon = 0.56$. Note that this number is quite close to the elasticity estimates in Chetty (2012), who creates a common confidence

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

17
interval for the elasticities of different studies.

### 4.1 Sample Selection

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

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

Second, I put lower and upper bound on the age of each spouse. The spouses are 35-45 years old. The age restriction helps in three ways: First, the age effect on income and children is eliminated. Empirical evidence suggests that earnings increase in the early ages (16-35) and become stabilized after the age of 35.\(^\text{16}\) 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.

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

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

---


\(^{16}\)See [http://www.bls.gov/news.release/wkyeng.t03.htm](http://www.bls.gov/news.release/wkyeng.t03.htm)
I plot the relationship between family labor income and the number of children in the family in Figure 6. The figure implies the well-known empirical evidence that the fertility rate is negatively correlated with family labor income.

![Figure 6: 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 converted to 2014$ using CPI deflator. The age of the spouses is between 35 and 45. The sample size is 37,165 after all restrictions.

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

### 4.2 Parental Time

The assumption on the cost of earnings and child care suggests that time is normalized to one: $h(0) = 0$ and $h'(1) = 1$ (see Kleven, Kreiner, and Saez (2009)). Hence, $b_n$ is the fraction of child care to the total labor time (market work and child care). To capture an estimate for $b_n$, I use the 2003 wave of American Time Use Survey (ATUS) sample which is also used by Aguiar and Hurst (2007). I restrict the sample using the criteria stated in Section 4.1. I show the time devoted to child care and market work in Table 1.

Table 1 shows that there is an economics of scale in the time cost of child-rearing. More analysis on $b_n$ can be found in Appendix A.4. In the benchmark, I assume one-child families devote 9% of their time to child care and two-child families spend 14% of their time for child-rearing.
<table>
<thead>
<tr>
<th>Category of labor:</th>
<th>0 child family</th>
<th>1 child families</th>
<th>≥ 2 children families</th>
</tr>
</thead>
<tbody>
<tr>
<td>child care</td>
<td>0.2</td>
<td>5.7</td>
<td>9.4</td>
</tr>
<tr>
<td>market</td>
<td>57</td>
<td>54.9</td>
<td>54.9</td>
</tr>
<tr>
<td>(b_n \simeq)</td>
<td>0</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>sample size</td>
<td>158</td>
<td>247</td>
<td>547</td>
</tr>
</tbody>
</table>

Table 1: Time Devoted to Market and Childcare

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

### 4.3 Cost of Goods

Haveman and Wolfe (1995) use Consumer Expenditure Survey (CEX) data and suggest that the goods cost is $12,788 per child (in terms of 2014$). Examples of such costs include expenditures on food, housing, transportation, clothing, and health care. More recently, a publication of the US Department of Agriculture, Lino (2014), analyzes the goods cost of child-rearing for families with different wealth.\(^{17}\) 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 2.

<table>
<thead>
<tr>
<th>Category of Families</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>39,989</td>
<td>11,597-13,211</td>
<td>19,847-21,137</td>
<td>23,710-24,683</td>
</tr>
<tr>
<td>Middle Income</td>
<td>84,114</td>
<td>16,442-19,014</td>
<td>28,149-30,426</td>
<td>33,444-35,273</td>
</tr>
<tr>
<td>High Income</td>
<td>189,443</td>
<td>27,095-32,646</td>
<td>47,793-52,234</td>
<td>56,104-58,022</td>
</tr>
<tr>
<td>(e_n \simeq)</td>
<td>$12,000</td>
<td>$20,000</td>
<td>$24,000</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Expenditures on child-rearing

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

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

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

---

4.4 Estimation of the Distribution of Earning Abilities

The quasi-linear preference structure allows me to find earning abilities of families (see Equation (2)):

\[ \theta = \frac{z}{(1 - T')\epsilon} - b_n. \]  

(10)

Note that CPS has information about family structure and detailed family income and taxes.\(^\text{18}\) Given the complexity of state rates, I focus only on federal tax rates. In addition, I add earned income tax rates to the federal marginal tax rates.\(^\text{19}\) Using \(b_n\) values from Table 1 and the weights of families given in the data, I derive the distribution of earning abilities and show it in Figure 7.

![Figure 7: Earning Ability Distribution](image)

4.5 Estimation of the Distribution of Child Taste

An important contribution of this paper introduces a distribution of tastes for children to the literature. I assume that \(\beta_i \sim [0, \bar{\beta}]\) is distributed according to a power function, i.e. the cumulative density is \(P(\beta) = \left(\frac{\beta}{\bar{\beta}}\right)^\eta\). The density of childless families equals \(P(\beta_1(\theta))\) for each \(\theta\). Hence, I interpret \(\eta := \frac{\partial \log P(\beta_1)}{\partial \log \beta_1}\) as the non-participation elasticity of

\(^{18}\)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 the child tax credits, federal marginal tax rates, and federal and state tax liabilities.

\(^{19}\)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.
zero-child families with respect to their tastes for children. Moreover, I assume \( m(n, \beta) = -(N - n)^p \beta \). Within this framework, I need to estimate \( \eta \) and \( p \).

I use Bernoulli maximum likelihood estimation to find the estimates. First, I derive percentiles of \( \theta \) distribution and calculate \( V_n(\theta_j) \) for each \( j-th \) percentile. Second, I can calculate the fraction of \( n- \) child families: \( \pi_n(\theta_j) \). In addition, I calculate the average number of children for each \( \theta_j: n(\theta_j) \). I plot \( V_n(\theta_j), \pi_n(\theta_j) \) and \( n(\theta_j) \) in Figure 8.

![Figure 8: Figures for \( V_n(\theta_j), \pi_n(\theta_j), \) and \( n(\theta_j) \) respectively. Also within the sample restriction, the average number of children in the environment is 1.492. Note that families with more than two children are considered as they have two children.]

Note that the probability of having \( n = 0, 1, 2 \) children are represented by \( P_0(\theta_j) := P(\beta_1(\theta_j)), P_1(\theta_j) := P(\beta_2(\theta_j)) - P(\beta_1(\theta_j)), \) and \( P_2(\theta_j) := 1 - P(\beta_2(\theta_j)) \), respectively. The Equation (5) provides their values for each \( \theta_j \). Next, I fix the upper bound with \( \beta = 300 \) and derive the Bernoulli maximum likelihood function:

\[
\max_{\eta, p} \mathcal{L} = \prod_j P_0(\theta_j)^{\pi_0(\theta_j)} P_1(\theta_j)^{\pi_1(\theta_j)} P_2(\theta_j)^{\pi_2(\theta_j)}. \tag{11}
\]

The estimates are given in Table 3 and I plot the distribution of child taste in Figure 9.

### 4.6 Deriving the Optimal Tax System

To solve the problem numerically, I find that the government collects $13,412 per capita taxes from the sample I use, while the population produces $109,421 per capita income.

---

20Note that CPS has information on how much taxes a family pays. Hence, I can calculate \( V_n(\theta) \).
Table 3: Estimation of Child Taste Distribution

<table>
<thead>
<tr>
<th>η</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.31</td>
<td>4.82</td>
</tr>
<tr>
<td>(0.83)</td>
<td>(13.50)</td>
</tr>
</tbody>
</table>

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

Figure 9: Probability distribution of child taste: $p(\beta)$

Hence, I set $G = $13,412 in my calculations. Before solving the optimal system, first I show the taxes of families under the current tax system in Figure 10. Panels A and B show the tax liabilities for low and high income families, respectively. Panels C and D suggest that the child tax credits are constant for the first seven quantiles. The credits decline after that and reach zero around the ninth quantile. Moreover, the marginal tax rates are higher for families with children at the first two quantiles. This is because the earned income tax credits fall from the plateau and increase the marginal rates (see Figure 1).

I solve the government’s problem (i.e. PMDP) at my selected and estimated parameters using the GPOPS-II software.\(^{21}\) Note that the government problem is an optimal control problem and the Hamiltonian of the problem is stated in Appendix A.2.

First I plot the optimal indirect utilities and optimal family sizes in Figure 11. It is clear that the current and the optimal indirect utilities of families are similar. However, the density of family sizes has distinct patterns across abilities (see Figure 8). The density of two child families decreases with earning abilities. The situation is reversed for childless

\(^{21}\)GPOPS-II is a flexible software for solving optimal control problems. For additional details see Patterson and Rao (2013).
Figure 10: Current Tax System

Left (right) panel is for the left (right) side of the income distribution. The first row shows the actual taxes paid by families. The second row shows how much child tax credits they get. Note that $k_1$ ($k_2$) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families. See Figure 19 in Appendix A.5 for $S$ base taxes.
families. The main reason behind this result is the time cost of child-rearing. The cost is relatively higher for families with higher earning abilities and the higher earning families produce fewer children. Interestingly, the density of the one-child family also increases. This result stems from economies of scale in child-rearing. Because of this scale, the credit received by one child families is higher than the average child credit received by two children families. As a result, the density of one child families increases with earning abilities.

Next, I numerically solve for the optimal tax system. First of all, the transversality conditions of Hamiltonian satisfy the conditions in Sadka (1976) and Seade (1977). As a result, the bottom and top of the incomes for each family sizes face zero marginal rates (see Panels E and F of Figure 12).

Panel E of Figure 12 shows that the government distorts the labor decision of two-child families more at the bottom. In return, these families receive a high subsidy via tax credits (see Panel A of Figure 12). This mainly stems from the effect of goods costs. The government provides enough goods to low-income families to raise their children. This also provides incentives for the low-income families to produce more children and less income, because the cost of generating income is relatively higher than the cost of child-rearing for the bottom. On the other hand, the distortion is relatively less for families with children at the top of the income distribution (see Panel F of Figure 12). This is mainly because the government does not want to increase the distortion on the income decision of the families with children whose income is more elastic because of the time cost of

Figure 11: Figures for $V_n(\theta_j)$, $\pi_n(\theta_j)$, and $n(\theta_j)$ respectively. Average number of children is 1.539 which is relatively 3% more than the current number children in the data.
Figure 12: Optimal Tax System

Left (right) panel is for the left (right) side of the income distribution. The first row shows the actual taxes paid by families. The second row shows how much child tax credits they get. Note that $k_1$ ($k_2$) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families. See Figure 20 in Appendix A.5 for $\$ base taxes.
the child-rearing. In addition, the government subsidizes these families to relieve their income loss due to the time cost of child-rearing (see Panel B of Figure 12).

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

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

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

4.7 Proposal

In this subsection, I propose a simple tax schedule. I let that the income taxes are determined by the optimal taxes of childless families. In addition, I propose a linear income dependent tax credits. The credits for the first (second) child linearly decrease in the first quartile, and are constant for the 25-65% (25-75%) of the income distribution. For the rest of income distribution, the credits linearly increase. I state the linear rates in Table 4. In addition, the minimum credits is determined by using the minimum values of the optimal credits.

I plot the optimal and the proposed credits in Figure 13. With this proposal, almost 87% of the welfare gain attained by the optimum is captured. To understand how the income dependent taxes improve social welfare, I also consider a proposal in which the
<table>
<thead>
<tr>
<th>Credit</th>
<th>Credit Rate</th>
<th>Phase In</th>
<th>Phase Out</th>
<th>Credit Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>-2%</td>
<td>$46,700$</td>
<td>$110,000$</td>
<td>3.7%</td>
</tr>
<tr>
<td>$k_2$</td>
<td>-2.8%</td>
<td>$46,700$</td>
<td>$139,000$</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Table 4: Credit Rates in Phase In and Phase Out

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

credit per child is same for all children and constant across income. This proposal only captures 70% of the welfare gain. As a result, the income dependent child tax credits can improve social welfare significantly.

![Figure 13: Proposal Tax Credits](image)

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014$. $k_n$ is the optimal tax credits and $\hat{k}_n$ is the proposed tax credit for $n = 1, 2$. See Figure 21 in Appendix A.5 for $ base credits.

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

5 Robustness

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

In this subsection, I relax age restriction of the sample. The minimum age of a spouse in a two-spouse family is relaxed to 25. I calculate the optimal credits and plot them in Figure 14. The tax credits are U-shaped for this sample. As a result, the credit shapes are quite robust without a restriction on ages of spouses.

![Figure 14: Tax Credits: Age Analysis](image)

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014$. See Figure 22 in Appendix A.5 for $ base credits.

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

5.2 Detailed Cost Analysis

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

Next, I change the set of expenditures on the goods cost. Some might consider that not all types of goods cost are on the basic needs for children. For example, moving a bigger house, which is around 30% of the goods cost, can be considered as a non-required cost for child-rearing. I modify the set of goods costs for basic needs. The new set consists of the expenditure on food, clothing, health care and education. These costs are around 50% of the original goods costs (see Lino (2014)). As a result, I set $e_1 = $6,000 and $e_2 = $10,000.
In Figure 15, I show that the optimal tax credits for this environment. Note that the credit amount are reduced due to the reduction in the child-rearing costs (see Figure 13). However, the credits are still U-shaped, and my results are robust.

![Figure 15: Tax Credits: Costs Analysis](image)

$k_n$ is the optimal tax credits for $n = 1, 2$. The amounts are 1,000 in 2014$. See Figure 23 in Appendix A.5 for $^\text{\$} base credits.

In the next subsection, I study if the marital status of households matters for the shape of tax credits. Since there are few single fathers, I focus only on single mothers.\(^{22}\)

### 5.3 Marital Status Analysis

In this subsection, I study the optimal taxation of single females. I make adjustment on child-rearing costs. The time costs for single mothers are set by $b_1 = 10\%$ and $b_2 = 19\%$, and goods cost are set by $e_1 = $11,500 and $e_2 = $18,500 (see Table 8 in Lino (2014)). I pick the elasticity of income as $\epsilon = 0.8$ (see Blundell, Pistaferri, and Saporta-Eksten (2012)). The optimal tax credits are U-shaped (see Figure 16). Note that the credit amounts are larger than the case for married families. The main reason is that the time cost for singles is bigger than the time cost for marrieds.

### 6 Conclusion

This paper studies optimal income taxation and child tax credits in a static Mirrlees model with heterogeneous shocks of child tastes and earning abilities. By facing these risks, fam-

\(^{22}\)See [https://www.census.gov/hhes/families/files/graphics/CH-1.pdf](https://www.census.gov/hhes/families/files/graphics/CH-1.pdf).
$k_n$ is the optimal tax credits for $n = 1, 2$. The amounts are 1,000 in 2014$. See Figure 24 in Appendix A.5 for $^*$ base credits.

families decide how much income to generate and how many children to have by considering child-rearing costs. The government aims to provide insurance against the shocks, which are families’ private information. To do so, the government designs an optimal tax system which combines income taxes of childless families and child tax credits. The sufficient statistics for labor wedges and their relationship with child tax credits are derived.

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

This paper sheds light on the optimal income taxation including the child benefits for families who have multidimensional private information. I conclude by describing three extensions that I leave for future research. First, the paper abstracts from a dynamic setting. Such a setting can explain how the child benefit should be characterized by the age of the children. Moreover, two heterogeneous risks, the earning abilities and child tastes, can be linked with the age of the parents and, therefore, the effect of optimal taxes on the fertility age can be studied. Second, the paper abstracts from the child quality decision, which is positively correlated with parental time according to Boca, Flinn, and Wiswall...
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


A  Appendix:

A.1 Mechanism Design: Two-Dimensional Private Information

In this section, I show the implementability conditions for a two-dimensional private information problem. I approach it similarly to Jacquet, 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 shock. 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 reports \( \gamma \) as their type, the government chooses optimal \( c(\gamma), n(\gamma), z(\gamma) \). This mechanism should satisfy the revelation principle, by which any government mechanism can be decentralized by a truthful mechanism \( (z(\gamma), n(\gamma), c(\gamma)) \gamma \in \Gamma \) such that

\[
U(\gamma, \gamma) \geq U(\gamma, \gamma').
\]

In this setup, a strategy has two dimensions, and hence a possible mimicking strategy has two dimensions. However, the possibility of double deviation in the mimicking strategy can be eliminated and double deviation can be reduced to single deviation by two constraints: indirect utility of \( n \) child families (4) and threshold 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 \( (z(\gamma), n(\gamma), c(\gamma)) \gamma \in \Gamma \) can be replaced by a new mechanism \( (c_n(\theta), z_n(\theta)) n \in \{0,1,\ldots,N\}, \theta \in \Theta, \) such that

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

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

\[23\text{I let } \beta_0 = \overline{\beta} \text{ and } \beta_{N+1} = \overline{\beta}.\]
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:

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

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

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 \( u(c(\beta, \theta)) - h \left( \frac{z(\beta, \theta)}{\theta} + b_j \right) \) 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 \( t = \sup_{z \in Z_j(\theta)} z - u^{-1}(V_0(\theta) + h(\frac{z}{\theta} + b_j) \theta) \). Note that \( z - u^{-1}(V_j(\theta) + h(\frac{z(\theta)}{\theta} + b_j) \theta) \) 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 \( z_j(\theta) \in Z_j(\theta) \) such that \( t = z_j(\theta) - u^{-1} \left( V_j(\theta) + h \left( \frac{z(\theta)}{\theta} + b_j \right) \theta \right) \). \(^{24}\)

Define \( c_j(\theta) := u^{-1} \left( V_j(\theta) - h \left( \frac{z_j(\theta)}{\theta} + b_j \right) \right) \). 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 \( j = 0, 1, \ldots, N \).

Finally, I define \( \beta_n(\theta) := M^{-1}(V_n(\theta) - V_{n+1}(\theta)) \) where \( M(\beta) := m(n + 1, \beta) - m(n, \beta) \) for all \( n = 0, 1, \ldots, N \). \(^{25}\) \( \beta_n(\theta) \) are the threshold tastes for children for each \( \theta \) and for each \( n \). Note that truthfulness of original mechanism implies: for all \( \beta \in B_j(\theta) \) 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 \( M(V_j(\theta) - V_{j+1}(\theta)) \geq \beta_j(\theta) \geq M(V_{j-1}(\theta) - V_j(\theta)) \). \(^{26}\) Therefore \( B_j(\theta) = (\beta_{j-1}(\theta), \beta_j(\theta)) \).

All is left to show the new mechanism \((c_n(\theta), z_n(\theta))_{n \in \{0,1,\ldots,N\}, \theta \in \Theta} \) is truthful. First I show it is truthful within families with the same number of children: For all \( \theta, \theta', \beta \in B_j(\theta), \beta' \in B_j(\theta'):\n
\[
U(c_j(\theta), z_j(\theta), (\beta, \theta)) = V_j(\theta) + m(j, \beta) \geq u(c(\beta', \theta')) - h \left( \frac{z(\beta', \theta')}{\theta} \right) - m(j, \beta)
\]

\(^{24}\) \( Z_j(\theta) \) is the closure of the \( Z_j(\theta) \)

\(^{25}\) Let \( \beta_0 = \beta \) and \( \beta_{N+1} = \beta \).

\(^{26}\) Note that I 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.\(^{27}\) 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 \(\theta, \theta', \beta \in B_j(\theta), \beta' \in B_j'(\theta'):\)

\[
U(c_j(\theta), z_j(\theta), (\beta, \theta)) = V_j(\theta) + m(j, \beta) \geq V_j'(\theta) + m(j', \beta')
\]

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

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

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

This procedure can be followed for any \(j = 0, 1, \ldots, N\). As a result, the proof is completed. \(\Box\)

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

**Definition 1.** \(z_n(\theta)_{n \in \{0,1,\ldots,N\}}\) is implementable if and only if there exist transfer functions \(c_n(\theta)_{n \in \{0,1,\ldots,N\}}\) such that \((c_n(\theta), z_n(\theta))_{n \in \{0,1,\ldots,N\}, \theta \in \Theta}\) 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 \(z_n(\theta)_{n \in \{0,1,\ldots,N\}}\) for all \(\theta \in \Theta\) is implementable if and only if \(z_n \geq 0\).

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

For the "if" side of the lemma, I directly apply the one-dimensional implementability condition: for all \(n = 0, 1, \ldots, N\), let \(z_n(\theta)_{n \in \{0,1,\ldots,N\}}\) is implementable. Then for a particular \(n\), truthfulness implies \(u(c_n(\theta)) - h \left(\frac{z_n(\theta)}{\theta} + b_n\right) \geq u(c_n(\theta')) - h \left(\frac{z_n(\theta')}{\theta} + b_n\right)\) for all

\(^{27}\) Note that \((c_j(\theta'), z_j(\theta'))\) is in the closure of the set \((c(\beta, \theta), z(\beta, \theta))_{\beta \in B_j(\theta)}\).
As a result, the one-dimensional result suggests that for each \( n = 0, 1, \ldots, N \) income is non-decreasing: \( \dot{z}_n \geq 0 \).

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

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

### A.2 Proof of Proposition 1

**Proof.** The Hamiltonian of the problem is:

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

for all \( n = 0, 1, \ldots, N \). Also, the co-states of the system are:

\[
\dot{\mu}_n = \lambda \left( 1 - \frac{\partial c_n}{\partial z_n} \right) (P(\beta_{n+1}|\theta) - P(\beta_n|\theta)) f(\theta) = -\frac{\mu_n}{\theta} h'' \left( \frac{z_n(\theta)}{\theta} + b_n \right) \frac{z_n(\theta)}{\theta}
\]

for all \( n = 0, 1, \ldots, N \). Also, the co-states of the system are:

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

for \( n = 0, \ldots, N \), where \( T_n(\theta) = z_n(\theta) - c_n(\theta) \) are the optimal income taxes collected from

---

28 Numeric exercises show that there is no bunching.

29 I assume that the implementability constraint does not bind and show ex-post is the case.

30 I derive co-state for \( V_0(\theta) \) and \( V_N(\theta) \) separately.
the θ—ability families with n children and for \( n = 0, \ldots, N - 1 \), \( \Delta T_n(\theta) = T_n(\theta) - T_{n+1}(\theta) \) is the tax credit for an extra child for θ—ability families.\(^3\)

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

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

where

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

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

(12)

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

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

\[
\frac{T'_n}{1 - T'_n} = \frac{1}{\varepsilon_n} \times \frac{1}{\theta f(\theta)(P(\beta_{n+1}|\theta) - P(\beta_n|\theta))} \times \int_{\theta}^{\bar{\theta}} \left[ \frac{(1 - g_n(\theta'))}{u'(c_n(\theta'))} \left( P(\beta_{n+1}|\theta') - P(\beta_n|\theta') \right) + \Delta T_{n-1}(\theta) p(\beta_n|\theta) \frac{\partial \beta_n}{\partial V_n} + \Delta T_n(\theta) p(\beta_{n+1}|\theta) \frac{\partial \beta_{n+1}}{\partial V_n} \right] u'(c_n(\theta)) f(\theta') d\theta' \quad (13)
\]

for \( n = 0, 1, \ldots, N \).

\(^3\)When \( n = 0 \), let \( \Delta T_{-1}(\theta) = 0 \) and similarly when \( n = N \), let \( \Delta T_N(\theta) = 0 \).

\(^3\)If \( \Delta T_{-1}(\theta) = 0 \) when \( n = 0 \) and \( \Delta T_N(\theta) = 0 \) when \( n = N \).
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.\(^{33}\)

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 39% of the family income (see Figure 17). Hence, $\varepsilon = 0.61 \times 0.4 + 0.39 \times 0.8 \simeq 0.56.$

\(^{33}\)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.
The ratio is very close to 0.39. Note that this graph suggests that the gender gap for this sample is 0.64, which is quite close to the actual gender gap in the US (0.7).

### A.4 Evolution of $b_n$

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

![Graph showing the evolution of $b_n$ for different income groups.]

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

---

"Refer to Aguiar and Hurst (2007) for further details."
A.5 Figures and Tables

Figure 19: Current Tax System

Left (right) panel is for families with less (more) than $50,000. The first row shows the actual taxes paid by families. The second row shows how much tax credits they get. Note that $k_1$ ($k_2$) represents the tax credits for one (two) child families. The last row shows the federal marginal tax rates of families.
Figure 20: Optimal Tax System

Left (right) panel is for families with less (more) than $50,000. The first row shows the actual taxes paid by families. The second row shows how much tax credits they get. Note that $k_1$ ($k_2$) represents the tax credits for two (one) child families. The last row shows the federal marginal tax rates of families.
Figure 21: Proposed Tax Credits

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

Figure 22: Tax Credits: Age Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014$. $k_n$ is the optimal tax credits for $n = 1, 2$. 
Figure 23: Tax Credits: Costs Analysis

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

Figure 24: Tax Credits: Marital Status Analysis

Left and right panel show the credits for the first child and the second child, respectively. The amounts are 1,000 in 2014$. $k_n$ is the optimal tax credits for $n = 1, 2$. 
Table 5: Earned Income Tax Credit Phase in and Phase out regions for 2014
Credits are in terms of $. The numbers in cells present values for married couples who fill taxes jointly. First column is for number of children. Second column shows maximum earnings to be eligible for the credit.

<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>6,480</td>
<td>0.08</td>
<td>496</td>
<td>13,540</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>9,720</td>
<td>0.34</td>
<td>3,305</td>
<td>23,260</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>13,650</td>
<td>0.40</td>
<td>5,460</td>
<td>23,260</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>13,650</td>
<td>0.45</td>
<td>6,143</td>
<td>23,260</td>
<td>0.21</td>
</tr>
</tbody>
</table>