Optimal Taxation and Fertility Policies

Joshua Bernstein*

September 19, 2018

Abstract

I analyze optimal taxation in an economy in which households make consumption, labor supply, and fertility choices. Applying the mechanism design approach, I derive sufficient statistics for the sign and shape of optimal wedges on child quantity, goods investment and time investment, and provide intuition for the main economic forces at play. Distorting fertility choices relaxes incentive constraints, which facilitates redistribution, but also may discourage households to earn income, thus hampering the redistributive strength of the income tax. I construct a family of tax functions that can implement the optimal allocation. A quantitative exercise demonstrates the welfare gains available from subsidizing child quantity and investment choices, the bulk of which can be obtained using feasible linear subsidies on child investment goods.

*Princeton University. I thank Mikhail Golosov, Mark Aguiar, and participants at the Princeton Public Finance, and Macro student seminars for many useful comments. Email: jmsb@princeton.edu

1 Introduction

The vast majority of people will make fertility decisions during their lifetimes, comprising of how many children to have, and how much to invest in them in terms of goods (e.g. food, clothing, and schooling) and time (e.g. early childcare, nighttime reading, and transportation).[1] Given that people make these decisions, any optimal tax system should take them into account. The core contribution of this paper is to tackle this task by extending a Mirrleesian income tax framework to include these fertility decisions in order to analyze the properties of optimal taxes (or subsidies) on child quantity and child investments. There are two key advantages to my approach. First, I am able to express optimal distortions in terms of estimable elasticities, thus connecting the optimal tax treatment of fertility choices to the growing literature that estimates the relationships between fertility choices and other variables such as consumption and income. Second, and related to this empirical link, I develop clear intuition for the properties of the optimal wedges, in particular their signs, by exploiting the interpretation of these elasticities. Neither of these have clear precedents in the existing literature on fertility taxation.

I begin by studying the case in which households have preferences that are linear in private consumption. This is a useful first step as it simplifies the analysis by focusing attention on the role that incentives play in the setting of optimal child-related taxes. Given linearity, the optimality of taxing or subsidizing child-related choices hinges on whether households with higher earnings ability prefer higher or lower amounts of each choice. Specifically, I show that the elasticity of the choice with respect to the earnings ability parameter is a sufficient statistic for the sign of the optimal marginal tax, where a negative sign means that a household receives a subsidy on the margin.

To grasp the intuition for this result, consider as an example the case of child quantity choice. Suppose that households with higher earnings ability prefer to have fewer children (the elasticity is negative) so that child quantity will be subsidized on the margin, and having an additional child will lower a household's total tax bill. The optimality of this subsidy stems from how it affects the household incentive constraint. In a standard income tax model incentive constraints bound the amount of redistribution that is possible since if income taxes are too high it becomes better for high earnings ability households to masquerade as lower ability households by choosing to earn less income and hence pay less income tax. However, since child quantity is observable, when households also make fertility choices, for a high ability household to successfully imitate a lower ability household, she would have to choose not only a lower income, but also a higher number of children (given that the elasticity is negative). Optimal taxes can then exploit the fact that there are now two margins of adjustment for each household. In particular, by subsidizing child quantity, lower income households will have even more children, thus making imitation more costly for higher ability households. The subsidy therefore relaxes the incentive constraints of each household since imitation is less desirable. Relaxation of the incentive constraint is valuable because it can then be tightened by increasing the amount of income tax that higher ability households have to pay, thus allowing more redistribution to take place.

I then extend the model to allow households to have preferences over consumption that feature diminishing marginal utility. In this case, the signs of the optimal wedges are jointly determined by two forces. The first is the incentive channel just described. The second stems from the effect that taxing fertility choices now has on a household's optimal choice of income. Intuitively, when fertility choices are taxed (or subsidized), they affect household consumption via the budget constraint. If consumption falls as a result of a tax on some fertility choice, a household may choose to earn more

[1] The very nature of pregnancy means that many children are of course "unplanned". However, I abstract from such uncertainty in my benchmark specification in order to focus on the trade off between incentive provision and redistribution that lies at the heart of optimal taxation problems.

income to offset this loss of consumption. To the extent that this income effect on labor supply occurs, the planner would prefer to tax fertility choices, since increasing household income increases tax revenue from the income tax thus enabling more redistribution to occur. Therefore, it is optimal to subsidize a fertility choice only when the benefits from relaxing incentives outweigh the costs associated with higher income. As in the linear case, I express this trade off in terms of the relevant estimable elasticities.

Having studied properties of optimal wedges, I then provide a result that characterizes the set of tax functions that will actually implement the optimal allocation. This is a non-trivial step since the tax function has to ensure that each household wants to choose the bundle prescribed to them in the optimal allocation designed by the planner, which requires careful treatment of all feasible bundles including those outside of the optimal allocation. The construction builds on Werning (2011) who considers the case of implementing a capital tax given a sequence of savings wedges.

Finally, I study a quantitative version of my model in order to assess the optimality of real tax policies aimed at families in the US. The key result is that there are large welfare gains from the optimal taxation of child-related choices, in particular goods investments, that the existing system does not exploit. I also show that most of these gains are achievable by introducing a simple linear subsidy on goods related to child investments. This is important, since in reality, observing these choices may be infeasible, thus ruling out non-linear taxation.

Related Literature There is a group of papers published in the early 2000s that study the optimal taxation of fertility choices. However, relative to my paper, the approaches taken either restrict the set of taxes available (e.g. linear taxes only), or consider non-linear taxation in the presence of only two levels of household ability.

Balestrino et al. (2002) study the optimal non-linear income and child quantity taxes, and the optimal linear child investment tax in an economy with households that can be differentiated along two dimensions: earnings ability, and child rearing ability. The multi-dimensional nature of the type-space makes the analysis complicated even in the two-by-two case that the authors consider. Restricting to only two sets of households also renders the formulas for optimal taxes hard to interpret since they are expressed in terms of primitives of the model rather than observable quantities. Finally, I find their conclusion about optimal child quantity taxes misleadingly emphasizes the role of child quantity taxes as an offset for the effect of other distortions, a result that relies on their specific assumptions.

Cigno and Pettini (2002) restrict attention to only linear taxes on consumption, child investments and child quantity and find that if higher wage parents invest more in their children, then it is optimal to tax consumption and investments, but to subsidize child quantity. This result has the flavor of results that I derive here. A key difference is that I do not restrict any taxes to be linear, which means that I can link each tax to its relevant elasticity thus making my results more general.

In each of these papers, it is assumed that households care about the quantity and quality of children. This assumption builds on the seminal contribution of Becker (1960) who first conceived of modeling children like other consumption goods, and of the trade off between the quality and quantity of children that a household chooses to have. While this mechanism has been used in the study of growth (e.g. de la Croix and Doepke, 2003), and earnings risk (Sommer, 2015), empirical evidence for the mechanism is somewhat limited. Black et al. (2005) study detailed Norwegian data and establish that family size effects on child quality outcomes are in fact almost entirely due to "birth order effects", where older siblings tend to do better than their younger siblings in later life outcomes such as earnings. Furthermore, while Rosenzweig and Zhang (2009) find evidence of a quantity-quality trade off in Chinese data, they conclude that it is a quantitatively small effect.

In addition to the lack of evidence for quantity-quality theory, there are a host of other potential mechanisms that can explain observed empirical relationships between fertility choices are other variables such as income. Jones et al. (2008) establish the negative correlation between income and quantity of children, and then explore a range of theories to try and explain it, from opportunity costs of time, to heterogeneous tastes for children. They conclude that no one theory can explain all the relevant facts simultaneously, and that more research is needed. This lack of consensus on the underlying mechanism motivates my theoretical approach, in which I avoid specifying a particular mechanism, and instead derive results that depend only on reduced form elasticities that are independent of underlying mechanisms.

Finally, to solve the optimal taxation problem, I build on the methodology first developed by Mirrlees (1971), and then further explored by Diamond (1998) and Saez (2001) for income taxation. In addition, I follow the recent trend in optimal taxation since Saez (2001) and derive formulas for optimal taxes that emphasize the roles of parameters that can potentially be estimated in the data, rather than relying on equations that are opaque combinations of primitives of the model.

The paper proceeds as follows. Section 2 introduces the household problem laying out the preferences of households that populate the economy. These are vital ingredients to the planning problem which is described in section 3. Section 4 explores the theoretical properties of the optimal wedges under quasi-linear preferences, while section 5 does so without this restriction. Given these results, I connect them to the literature on optimal commodity taxation in section 6. In section 7, I show how to design a tax function that can implement the optimal allocation. Sections 8 and 9 solve a quantitative version of the model, and compare the optimal tax policies to those observed in the US currently. In light of this, section 10 shows studies the potential welfare gains from implementing simpler policies that those required by the optimal allocation. Section 11 concludes.

2 A Model of Fertility Choice

There is a unit mass of households who each make choices over consumption, labor effort, child quantity, child goods investment, and child time investment, denoted by c, l, n, K and H respectively. K and H refer to the total amount of goods and time that are devoted to the household's n children. Each household is further characterized by a one dimensional parameter $\theta \in [\underline{\theta}, \infty)$ where $\underline{\theta} \geq 0$. Let the c.d.f. and p.d.f. of the distribution of θ be F and f respectively, where I assume that the p.d.f. exists. As is standard in the optimal taxation literature, I shall refer to θ as a household's ability. However, θ plays two distinct roles in my model, in contrast to the singular role it is usually given.[2]

First, θ captures a household's productivity in producing the single output good in the economy. Specifically, I adopt the usual linear production function and assume that if a household of ability θ supplies l units of labor effort, she earns income $y = \theta l$, where I have implicitly (and without loss of generality) normalized the wage rate to one. Second, θ directly impacts a household's optimal choices of child quantity n and child investments K and H. This interaction is born out in the utility function of households, which I now state (writing $l = y/\theta$):

$$U(c, l, n, K, H) = u(c) - \phi(y/\theta, H) + v(n, K, H, \theta)$$

where u is increasing and at least weakly concave, ϕ is increasing and at least weakly convex in each argument, and v is concave in n and increasing and at least weakly concave in K and H. Note

[2]It would of course be more general to let θ be a multi-dimensional parameter. However, as is well understood, solving mechanism design problems with multi-dimensional hidden types is prohibitively difficult. Therefore, I proceed under the assumption that the same parameter enters the utility function twice.

that both the benefits and costs of child quantity are subsumed in the v function. While this is somewhat different to other specifications (for example de la Croix and Doepke, 2003), doing so offers two benefits. First I am able to skip discussion of the non-convexities that arise when considering per-child investments $k = K/n$, $h = H/n$, by working with aggregate variables directly (Alvarez, 1999). Second, deriving taxes that depend on aggregate quantities is simpler since I do not have to account for how a tax on per-child investment may affect child quantity decisions, and vice versa.

The function v permits a general specification of how θ interacts with the variables n, K, and H. This is advantageous for two reasons. Firstly, more structure is unnecessary since I shall derive optimal taxes whose key features depend only on reduced form elasticities rather than primitives of the model such as v. Second, it is unclear what sort of structural relationship to specify since there is little consensus on what mechanisms govern the way households make fertility choices. As surveyed by Jones et al. (2008), there are a wide range of potential explanations for the observed empirical relationships between income and fertility choices, from standard substitution effects, to models with heterogeneous tastes for children. As such, it would be premature for me to specify an exact mechanism of fertility choice since more research is needed to fully understand which kinds of mechanisms are likely to be the dominant forces at play.

3 The Planning Problem

In the baseline model specification, the planner is welfarist and so only cares about the households' utilities. For concreteness, I consider an additive welfare function, with Pareto weights $\{\alpha(\theta)\}_{\Theta}$ (normalized so that $\mathbb{E}[\alpha(\theta)] = 1$) capturing the planner's preference for redistribution beyond the implicit desire for insurance captured by the risk averse household preferences. I assume that $\partial\alpha/\partial\theta < 0$ so that the planner would like to redistribute resources from higher ability households to lower ability ones. This reflects three potential motivations. First, there could be purely normative reasons for wanting to redistribute. Second, there is the usual idea that the planner would like to insure households against the risk of having low earnings ability in the absence of private insurance markets for this risk. Finally, and novel to this framework, the planner would like to insure the children born to each household against the risk of having parents with low ability.

If the planner could observe θ, her desired level of redistribution could be achieved by a system of lump sum taxes and transfers conditional on θ. However, as in all taxation models, I assume that the planner cannot observe θ and in fact can only observe a household's choices of income, child quantity and investments $\{y, n, K, H\}$, which prevents the planner from even inferring θ from household choices[3].

Given this asymmetric information environment, I follow the recent taxation literature and use mechanism design to solve the planning problem (Golosov et al., 2003). Specifically, the planner chooses an allocation $A = \{c(\theta), y(\theta), n(\theta), K(\theta), H(\theta)\}_{\Theta}$ that maximizes social welfare subject to resource feasibility and incentive-compatibility:

$$\max_{A} \int_{\Theta} \alpha(\theta) U(\theta) f(\theta) d\theta$$

[3]It is certainly questionable whether some kinds of investment are really observable, especially those involving time. However, from a pedagogical perspective, it makes sense to begin with my assumption of full observability, since relaxing this assumption is a simple extension. Intuitively, when a choice becomes unobservable, taxes on observable goods change to reflect how well targeting them also allows the planner to target the unobserved choice. For an example in the context of human capital, see section 7 of Stantcheva (2015).

subject to

$$\begin{aligned}
U(\theta) &= u(c(\theta)) - \phi\left(\tfrac{y(\theta)}{\theta}, H(\theta)\right) + v(n(\theta), K(\theta), H(\theta), \theta) \\
\theta &= \arg\max_{\theta'} u(c(\theta')) - \phi\left(\tfrac{y(\theta')}{\theta}, H(\theta')\right) + v(n(\theta'), K(\theta'), H(\theta'), \theta) \\
\int_\Theta (c(\theta) + K(\theta)) f(\theta) d\theta &\leq \int_\Theta y(\theta) f(\theta) d\theta
\end{aligned}$$

The first constraint simply defines household utility under truth-telling in the allocation A. The second constraint ensures that each household truthfully reports their type when faced with the allocation A, as is required in a direct mechanism. The last constraint ensures that the allocation is resource-feasible, i.e. that total consumption and goods investment is less than or equal to total output.

As is standard, I simplify the set of incentive constraints (ICs) by using the First Order Approach (Farhi and Werning, 2013), which replaces each household's global incentive constraint with a necessary envelope condition from their private optimization. I then verify ex-post that the global ICs are indeed satisfied. Therefore, the planner's problem that I solve can be stated as

$$\max_A \int_\Theta \alpha(\theta) U(\theta) f(\theta) d\theta$$

subject to

$$\begin{aligned}
U(\theta) &= u(c(\theta)) - \phi\left(\tfrac{y(\theta)}{\theta}, H(\theta)\right) + \beta v(n(\theta), K(\theta), H(\theta), \theta) \\
U_\theta(\theta) &= \tfrac{y(\theta)}{\theta^2}\phi_l\left(\tfrac{y(\theta)}{\theta}, H(\theta)\right) \\
&\quad + \beta v_\theta(n(\theta), K(\theta), H(\theta), \theta) \\
\int_\Theta (c(\theta) + K(\theta)) f(\theta) d\theta &\leq \int_\Theta y(\theta) f(\theta) d\theta
\end{aligned}$$

This problem can be solved using standard optimal control techniques. Details of the first order conditions can be found in the appendix.

4 Optimal Fertility Policies Under Quasi-linearity

4.1 Elasticity Concepts

A contribution of this paper is to express optimal taxes in terms of quantities that are estimable, much like Saez (2001) does for income taxation. As such, I now introduce the relevant elasticities that enter the formulas for the optimal taxes.

First, define the compensated elasticity of labor supply with respect to ability by

$$\epsilon^c(\theta) = \frac{\partial \log l^c(\theta)}{\partial \log \theta}$$

where $l^c(\theta)$ is the compensated labor supply function of a household with ability θ. Note that $l^c(\theta)$ is conditional on all other household choices, so that $\epsilon^c(\theta)$ measures the percent change in labor supply for a one percent change ability holding n, K, and H fixed.

Next define the conditional demand elasticity for choice $j \in \{n, K, H\}$ with respect to ability by

$$\epsilon_\theta^j(\theta) = \frac{\partial \log j(\theta)}{\partial \log \theta}$$

where $j(\theta)$ is the optimal choice of j for a given θ, holding other choices fixed. Hence, $\epsilon_\theta^j(\theta)$ measures the percent change in demand for j for a one percent change in ability holding l and all $j' \in \{n, K, H\} \setminus \{j\}$ fixed.

Finally, define the conditional demand semi-elasticity for choice $j \in \{n, K, H\}$ with respect to the tax rate τ_j by

$$e_{\tau_j}^j(\theta) = \frac{\partial \log j(\theta)}{\partial \tau_j}$$

where $j(\theta)$ is the optimal choice of j for a given θ, holding other choices fixed. Hence, $e_\theta^j(\theta)$ measures the percent change in demand for j for a one unit change in τ_j, holding l and all $j' \in \{n, K, H\} \setminus \{j\}$ fixed.

4.2 Wedges in the Optimal Allocation

In order to understand features of the optimal allocation, it is intuitive to analyze the distortions it creates relative to the choices each household would make in the absence of intervention. Such comparisons can be made concrete via the concept of wedges, which can be thought of as locally linear marginal tax rates. More precisely, consider the following household problem when the household is subject to linear tax rates on income, children and child investments, τ_y, τ_n, τ_K and τ_H:

$$\max_{y,K,H,n} u\left(y\left(1-\tau_y\right) - K\left(1+\tau_K\right) - \tau_H H - \tau_n n\right) - \phi\left(\frac{y}{\theta}, H\right) + v\left(n, K, H, \theta\right)$$

From the first order conditions, it is simple to show that the household's optimal choices are such that the following equations hold:

$$\begin{aligned}
\tau_y &= 1 - \tfrac{\phi_l(y/\theta, H)}{\theta u'(c)} \\
\tau_n &= \tfrac{v_n(n,K,H,\theta)}{u'(c)} \\
\tau_K &= \tfrac{v_K(n,K,H,\theta)}{u'(c)} - 1 \\
\tau_H &= \tfrac{v_H(n,K,H,\theta) - \phi_H(y/\theta,H)}{u'(c)}
\end{aligned}$$

These wedges provide a convenient way to quantify the distortions created by the planner's optimal allocation, as I now demonstrate. To avoid cluttered notation, all functions and their derivatives are implicitly evaluated at the optimal allocation, e.g. $v_n(\theta) = v_n(n(\theta), K(\theta), H(\theta), \theta)$. All proofs of results are contained in the appendix.

4.3 Optimal Wedges

It is instructive to begin by assuming that the household has quasi-linear preferences, so that $u(c) = c$. This is reminiscent of Diamond (1998) who analyzed the income taxation problem under risk neutrality, and was able to provide clear intuition for results as a consequence. I start by discussing the income wedge itself, and then move onto the main wedges of interest on child quantity, and goods and time investment.

4.3.1 The Optimal Income Wedge, $\tau_y^*(\theta)$

Proposition 1. *The optimal income wedge is given by*

$$\frac{\tau_y^*(\theta)}{1-\tau_y^*(\theta)} = \eta(\theta)\frac{1-F(\theta)}{\theta f(\theta)}\left(1+\frac{1}{\epsilon^c(\theta)}\right)$$

This expression is standard in the optimal taxation literature (see Saez (2001) for a detailed discussion), and highlights the usual three forces that determine the optimal income wedges. $\eta(\theta) = \frac{1}{1-F(\theta)}\int_\theta^\infty (1-\alpha(v)) f(v)\, dv \geq 0$ captures the planner's preference for redistribution via the Pareto weights. $\frac{1-F(\theta)}{\theta f(\theta)}$ is a standard hazard rate term, and trades off the gain in revenue from the wedge (numerator) against the total distortion it creates (denominator). Finally, the elasticity term captures the usual behavioral effects of income taxation, where higher taxation leads to lower labor supply and hence less income to tax. The key difference is that the elasticity $\epsilon^c(\theta)$ implicitly accounts for the fact that labor maybe more or less elastically supplied depending on the fertility choices of the household.

4.3.2 The Optimal Child Quantity Wedge, $\tau_n^*(\theta)$

Proposition 2. *The optimal child quantity wedge is given by*

$$\tau_n^*(\theta) = \eta(\theta)\frac{1-F(\theta)}{\theta f(\theta)}\left(-\frac{1}{e_{\tau_n}^n(\theta)}\right)\epsilon_\theta^n(\theta)$$

The first two terms have the same interpretation as before. Focusing on the novel term $\left(-\frac{1}{e_{\tau_n}^n(\theta)}\right)\epsilon_\theta^n(\theta)$ indicates that the shape and sign of the optimal child quantity wedge depends on how households adjust their choice of child quantity with respect to changes in both the wedge and ability. Noting that $e_{\tau_n}^n(\theta) < 0$ since it captures a pure substitution effect, it is intuitive that the wedge is declining in the absolute size of this semi-elasticity since if households' child quantity choices are more sensitive to the wedge, then it is a blunter instrument for redistribution, and has larger efficiency costs. This intuition closely resembles the role of ϵ^c in the optimal income wedge expression, as discussed by Saez (2001).

The appearance of $\epsilon_\theta^n(\theta)$ is novel to the literature. I begin by exploiting the fact that all other terms are positive, to establish that the elasticity $\epsilon_\theta^n(\theta)$ is a sufficient statistic for the sign of $\tau_n^*(\theta)$.

Corollary 1. $\tau_n^*(\theta)$ *is negative if and only if*

$$\epsilon_\theta^n(\theta) < 0$$

To grasp the intuition for this result suppose that households with higher earnings ability prefer to have fewer children (the elasticity is negative) so that child quantity will be subsidized on the margin, and having an additional child will lower a household's total tax bill. The optimality of this subsidy stems from how it affects the household incentive constraint. In a standard income tax model incentive constraints bound the amount of redistribution that is possible since if income taxes are too high it becomes better for high earnings ability households to masquerade as lower ability households by choosing to earn less income and hence pay less income tax. However, since child quantity is observable, when households also make fertility choices, for a high ability household to successfully imitate a lower ability household, she would have to choose not only a lower income, but also a higher number of children (given that the elasticity is negative). Optimal taxes can then exploit the fact that there are now two margins of adjustment for each household. In particular, by subsidizing child quantity, lower income households will have even more children, thus making imitation more costly for higher ability households. The subsidy therefore relaxes the incentive constraints of each household since imitation is less desirable. Relaxation of the incentive constraint is valuable because it can then be tightened by increasing the amount of income tax that higher ability households have to pay, thus allowing more redistribution to take place.

While the sign of ϵ_θ^n determines the sign of the optimal child quantity wedge, its magnitude also plays a distinct role. Specifically, the larger the absolute value of ϵ_θ^n, the larger is the absolute value of the optimal wedge. This relationship stems from the fact that the planner would like to redistribute resources as efficiently as possible, and that the elasticity ϵ_θ^n directly measures how efficient a child quantity distortion is relative to an income distortion. To see this, note that by dividing the expressions for the optimal income and child quantity wedges, I obtain

$$\tau_n^*(\theta) / \frac{\tau_y^*(\theta)}{1-\tau_y^*(\theta)} = \epsilon_\theta^n(\theta)\left(-\frac{1}{e_{\tau_n}^n(\theta)}\right) / \left(1+\frac{1}{\epsilon^c(\theta)}\right)$$

so that for fixed responses to tax rates, $e_{\tau_n}^n(\theta)$ and $\epsilon^c(\theta)$, the relative size of the child quantity wedge is increasing in the magnitude of $\epsilon_\theta^n(\theta)$. Intuitively, when $\epsilon_\theta^n(\theta)$ is larger in absolute value, distorting child quantity is more efficient than distorting income because the incentive constraint is less binding in that direction. Put a different way, as $|\epsilon_\theta^n(\theta)|$ increases, the difference in child quantity for households with different abilities rises for any given tax system. Therefore, distorting a household's child quantity choice is relatively more efficient as imitation is more costly for households along this dimension.

4.3.3 The Optimal Child Goods Investment Wedge, $\tau_K^*(\theta)$

Proposition 3. *The optimal child goods investment wedge is given by*

$$\tau_K^*(\theta) = \eta(\theta)\frac{1-F(\theta)}{\theta f(\theta)}\left(-\frac{1}{e_{\tau_K}^K(\theta)}\right)\epsilon_\theta^K(\theta)$$

Focusing on the novel term $\left(-\frac{1}{e_{\tau_K}^K(\theta)}\right)\epsilon_\theta^K(\theta)$, I note that $e_{\tau_K}^K(\theta) < 0$ by the substitution effect, and that the optimal wedge declines as this tax response increases in magnitude. Given this, it is clear that ϵ_θ^K is a sufficient statistic for the sign of $\tau_K^*(\theta)$.

Corollary 2. $\tau_K^*(\theta)$ *is positive if and only if*

$$\epsilon_\theta^K(\theta) > 0$$

The intuition for this result is similar to the case of child quantity. Suppose that households with higher earnings ability prefer to invest more goods in their children (the elasticity is positive) so that goods investment will be taxed on the margin, and additional investing will raise a household's total tax bill. The optimality of this tax again stems from how it affects the household incentive constraint. Since goods investment is observable, for a high ability household to successfully imitate a lower ability household, she would have to choose not only a lower income, but also a lower investment level. Hence, by taxing goods investment, lower income households will invest less, thus making imitation more costly for higher ability households. The tax therefore relaxes the incentive constraints of each household since imitation is less desirable. Relaxation of the incentive constraint is valuable because it can then be tightened by increasing the amount of income tax that higher ability households have to pay, thus allowing more redistribution to take place.

Like the child quantity wedge, the optimal child goods investment wedge also depends on the size of the elasticity $\epsilon_\theta^K(\theta)$. Again, this link captures the pure efficiency of taxing goods investment relative to taxing any other margin. For example, suppose that $\epsilon_\theta^K(\theta) > 0$ so that $\tau_K^*(\theta) > 0$. Then, $\tau_K^*(\theta)$ increases with $\epsilon_\theta^K(\theta)$ because the steeper the gradient of K with respect to θ, the less the incentive constraint binds in that direction, making taxation of this choice more efficient.

4.3.4 The Optimal Child Time Investment Wedge, $\tau_H^*(\theta)$

Proposition 4. *The optimal child time investment wedge is given by*

$$\tau_H^*(\theta) = \eta(\theta)\frac{1-F(\theta)}{\theta f(\theta)}\left(-\frac{1}{e_{\tau_H}^H(\theta)}\right)\epsilon_\theta^H(\theta)$$

Noting that $e_{\tau_H}^H < 0$ by the substitution effect, I again conclude that the optimal wedge is decreasing in the magnitude of the tax response, and that ϵ_θ^H is a sufficient statistic for the sign of $\tau_H^*(\theta)$.

Corollary 3. $\tau_H^*(\theta)$ *is positive if and only if*

$$\epsilon_\theta^H > 0$$

The intuition for this result is similar to the case of goods investment. Suppose that households with higher earnings ability prefer to invest more time in their children (the elasticity is positive) so that time investment will be taxed on the margin, and additional investing will raise a household's total tax bill. The optimality of this tax again stems from how it affects the household incentive constraint. Since time investment is observable, for a high ability household to successfully imitate a lower ability household, she would have to choose not only a lower income, but also a lower investment level. Hence, by taxing time investment, lower income households will invest less, thus making imitation more costly for higher ability households. The tax therefore relaxes the incentive constraints of each household since imitation is less desirable. Relaxation of the incentive constraint is valuable because it can then be tightened by increasing the amount of income tax that higher ability households have to pay, thus allowing more redistribution to take place.

Like the previous wedges, the optimal child time investment wedge also depends on the size of the elasticity $\epsilon_\theta^H(\theta)$. Again, this link captures the pure efficiency of taxing goods investment relative to taxing any other margin. For example, suppose that $\epsilon_\theta^H(\theta) > 0$ so that $\tau_H^*(\theta) > 0$. Then, $\tau_H^*(\theta)$ increases with $\epsilon_\theta^H(\theta)$ because the steeper the gradient of H with respect to θ, the less the incentive constraint binds in that direction, making taxation of this choice more efficient.

5 Optimal Fertility Policies Without Quasi-linearity

5.1 Elasticity Concepts

I now relax the assumption that households have quasi-linear preferences over private consumption, so that $u'(c) > 0$, and $u''(c) < 0$. While this is certainly more realistic, the cost is that the expressions for the optimal wedges become substantially more complicated. In order to maintain a sufficient statistics approach, I must define the following elasticities of consumption.

Define the elasticity of consumption with respect to ability θ holding all choices except $j \in \{n, K, H\}$ fixed by

$$\epsilon_\theta^{c,j} = \frac{\partial \log c}{\partial \log \theta}$$

Define the semi-elasticity and elasticity of consumption with respect to the tax rate τ_j holding all choices except $j \in \{n, K, H\}$ fixed by

$$e_{\tau_j}^{c,j} = \frac{\partial \log c}{\partial \tau_j}, \; \epsilon_{\tau_j}^{c,j} = \frac{\partial \log c}{\partial \log \tau_j}$$

As shown in the appendix, these elasticities can be expressed using the elasticities already defined in the risk neutral case. However, for brevity and intuition, I use the consumption notation in the analysis that follows.

Finally, it is useful to define the elasticity of marginal utility of consumption,

$$\sigma(\theta) = -\frac{u''(c(\theta))\, c(\theta)}{u'(c(\theta))}$$

5.2 Optimal Wedges

5.2.1 The Optimal Income Wedge, $\tau_y^*(\theta)$

Proposition 5. *The optimal income wedge is given by*

$$\frac{\tau_y^*(\theta)}{1 - \tau_y^*(\theta)} = \eta(\theta) \frac{1 - F(\theta)}{\theta f(\theta)} u'(\theta) \left(\frac{1 + \epsilon^u(\theta)}{\epsilon^c(\theta)} \right)$$

With a slight abuse of notation, I redefine $\eta(\theta) = \frac{1}{1-F(\theta)} \int_\theta^\infty \left(\frac{1}{u'(v)} - \frac{\alpha(v)}{\lambda} \right) f(v)\, dv \geq 0$ to capture redistribution preferences with household risk aversion. $\lambda = 1/\mathbb{E}\left[1/u'(\theta)\right]$ is the social cost of public funds (equal to unity in the risk neutral case). Finally, the elasticity term now captures both the usual behavioral cost of income taxation (ϵ^c) and the benefit via the income effect $(1 + \epsilon^u)$ since households facing higher taxes will actually supply more labor. Again, these elasticities must be estimated taking fertility choices into account, in contrast to the existing literature that tends to ignore these choices.

5.2.2 The Optimal Child Quantity Wedge, $\tau_n^*(\theta)$

Proposition 6. *The optimal child quantity wedge is given by*

$$\tau_n^*(\theta) = \eta(\theta) \frac{1 - F(\theta)}{\theta f(\theta)} u'(\theta) \left(-\frac{\epsilon_\theta^n}{e_{\tau_n}^n} \left(1 - \sigma(\theta)\, \epsilon_{\tau_n}^{c,n}\right) - \frac{\epsilon_\theta^{c,n}}{e_{\tau_n}^{c,n}} \sigma(\theta)\, \epsilon_{\tau_n}^{c,n} \right)$$

I begin by noting that this expression nests the risk neutral case in which $u' = 1$ and $\sigma = 0$. Therefore, risk aversion introduces an additional term that crucially plays a role in determining the sign of $\tau_n^*(\theta)$.

Corollary 4. *The optimal child quantity wedge is negative if and only if*

$$-\frac{\epsilon_\theta^n}{e_{\tau_n}^n} \left(1 - \sigma(\theta)\, \epsilon_{\tau_n}^{c,n}\right) < \left(\sigma(\theta)\, \epsilon_{\tau_n}^{c,n}\right) \frac{\epsilon_\theta^{c,n}}{e_{\tau_n}^{c,n}}$$

Although more complicated than the risk neutral case, intuition for this condition can be obtained by considering each the role of each term separately. First, note that $-\epsilon_{\theta}^{n}/e_{\tau_n}^{n}$ captures exactly the same incentive and efficiency considerations as in the risk neutral case: ϵ_{θ}^{n} measures how much distorting n relaxes the incentive constraint, while $e_{\tau_n}^{n} < 0$ measures the behavioral response of a household's child quantity choice. Second, consider the term $\epsilon_{\theta}^{c,n}/e_{\tau_n}^{c,n}$. As defined, this term simply measures the mechanical effects that a change in child quantity choice has on consumption via the budget constraint, $c = \theta l \left(1 - \tau_y\right) - K \left(1 + \tau_K\right) - \tau_H H - \tau_n n$, holding all other choices fixed. I stress that these elasticities can hence be computed as functions of elasticities already studied. They are not independent parameters. In isolation, this term is irrelevant from the planner's point of view. However, what matters is how these changes in consumption affect the household's optimal choice of labor supply and hence income.

In the risk neutral case, a household's optimal choice of income depends only on ability θ, and the income wedge τ_y. As such, changes in consumption due to changes in child quantity have no effect on household income. Mathematically, this is represented by $\sigma = 0$, so that the above condition becomes the one discussed in the risk neutral section. However, when the household is risk averse, optimal income choices also depend on all the other choices and wedges, in particular the child quantity choice and wedge. In this case, changes in consumption due to changes in child quantity directly affect the income choice of a household. For example, if child quantity is subsidized, then having another child offers an alternative source of income to the household. Therefore, a household might optimally choose to work less and have another child instead. But if a household works less, then there will be less tax revenue from the income tax, and therefore the subsidies to child quantity will have to decrease to respect the aggregate resource constraint faced by the planner.

The strength of this effect is measured by the term $\sigma\left(\theta\right)\epsilon_{\tau_n}^{c,n}$. $\epsilon_{\tau_n}^{c,n}$ measures by what percentage consumption changes as a result of a 1% change in the child quantity wedge τ_n (acting through the choice of child quantity in the budget constraint). $\sigma\left(\theta\right)$ is an inverse measure of how much a household is willing to substitute children for consumption. In the risk neutral case, $\sigma\left(\theta\right) = 0$ so that a household is infinitely willing to substitute consumption for children and so does not adjust her income at all in the face of a child quantity wedge. As $\sigma\left(\theta\right)$ increases, the household becomes less willing to substitute, and so adjusts her income accordingly to maintain the same level of consumption in the face of a child quantity wedge.

Therefore, when choosing whether to tax of subsidize child quantity, the planner must trade-off the benefits of relaxing incentives against this new cost of discouraging household labor effort. The above condition shows that when the gradient of child quantity with respect to ability is steep (and negative) enough, the benefits of subsidization outweigh the costs, and $\tau_n^* < 0$ is optimal.

5.2.3 The Optimal Child Goods Investment Wedge, $\tau_K^*\left(\theta\right)$

Proposition 7. *The optimal child goods investment wedge is given by*

$$\tau_K^*\left(\theta\right) = \eta\left(\theta\right)\frac{1 - F\left(\theta\right)}{\theta f\left(\theta\right)}u'\left(\theta\right)\left(-\frac{\epsilon_{\theta}^{K}}{e_{\tau_K}^{K}}\left(1 - \sigma\left(\theta\right)\epsilon_{1+\tau_K}^{c,K}\right) - \frac{\epsilon_{\theta}^{c,K}}{e_{\tau_K}^{c,K}}\sigma\left(\theta\right)\epsilon_{1+\tau_K}^{c,K}\right)$$

Rearranging the final term makes it clear that a similar result holds concerning the sign of $\tau_K^*\left(\theta\right)$.

Corollary 5. *The optimal child goods investment wedge is positive if and only if*

$$-\frac{\epsilon_{\theta}^{K}}{e_{\tau_K}^{K}}\left(1 - \sigma\left(\theta\right)\epsilon_{1+\tau_K}^{c,K}\right) > \left(\sigma\left(\theta\right)\epsilon_{1+\tau_K}^{c,K}\right)\frac{\epsilon_{\theta}^{c,K}}{e_{\tau_K}^{c,K}}$$

The intuition for this condition is very similar to the child quantity case. A positive wedge on child goods investment will relax incentives when $\epsilon_{\theta}^{K} > 0$, but also mechanically reduces consumption via the budget constraint. To the extent that households are unwilling to substitute consumption for goods investment, this will lead households to actually earn more income to make up for the increased cost of K. In this sense, there is actually a larger force for taxing goods investment on the margin since it has both an incentives and tax revenue benefit.

5.2.4 The Optimal Child Time Investment Wedge, $\tau_H^*(\theta)$

Proposition 8. *The optimal child time investment wedge is given by*

$$\tau_H^*(\theta) = \eta(\theta)\frac{1-F(\theta)}{\theta f(\theta)}u'(\theta)\left(-\frac{\epsilon_{\theta}^{H}}{e_{\tau_H}^{H}}\left(1-\sigma(\theta)\epsilon_{\tau_H}^{c,H}\right)-\frac{\epsilon_{\theta}^{c,H}}{e_{\tau_H}^{c,H}}\sigma(\theta)\epsilon_{\tau_H}^{c,H}\right)$$

Again, the sign of $\tau_H^*(\theta)$ is determined by the final term in brackets.

Corollary 6. *The optimal child time investment wedge is positive if and only if*

$$-\frac{\epsilon_{\theta}^{H}}{e_{\tau_H}^{H}}\left(1-\sigma(\theta)\epsilon_{\tau_H}^{c,H}\right) > \left(\sigma(\theta)\epsilon_{\tau_H}^{c,H}\right)\frac{\epsilon_{\theta}^{c,H}}{e_{\tau_H}^{c,H}}$$

Once again, the intuition is similar to that of the goods investment case.

6 Connection to Optimal Commodity Taxation

One of Becker's original insights was that fertility choices could be modeled using standard consumer theory. It is therefore unsurprising that the optimal tax treatment of these choices bares some similarities to the theoretical results on optimal commodity taxation. However, I now emphasize that although similar in flavor, my results stem from different underlying mechanisms. Further, I also describe ways in which my approach is somewhat richer than the usual treatment of commodity taxation.

Since Corlett and Hague (1953), it has been well understood that it is optimal to tax those commodities that are substitutable for labor in the utility function. This stems from the simple logic that by encouraging labor supply, the planner can achieve more redistribution via the income tax. More recently, Boadway and Jacobs (2014) have derived explicit formulas for these optimal commodity taxes that spell out this intuition, and feature relevant elasticities, much like I do here. However, there are two key differences between the usual approach to commodity taxation and the method I adopt in this paper.

First, commodity taxes are almost always assumed to be from a restricted class. The fact that the planner cannot in general observe how much of a particular commodity each household purchases means that taxation must be anonymous. Therefore, commodity taxes are restricted to be linear. By contrast, fertility choices can much more plausibly be assumed to be observed by the planner. In reality, many governments collect information on household births for social security reasons, and surveys such as the Consumer Expenditure Survey and American Time Use Survey in the USA contain special sections devoted to investments in children. Given this observability, non-linear taxation of fertility choices becomes feasible and motivates the approach I have taken in this paper. As a result of this non-linearity, the forces determining optimal fertility taxes and their relationships to the optimal income tax schedule differ markedly from the case of commodity taxation. As I have shown, the optimal taxes on fertility choices depend on social welfare weights and the distribution of abilities in the economy; both of these forces are missing in Boadway and Jacobs (2014) when linear commodity taxes are considered. Perhaps surprisingly, the non-linearities I consider actually lead to simpler (indeed closed-form) expressions for the optimal taxes since I can express everything in terms of ability, while linear taxes lead to complicated expressions involving the so-called "index of discouragement" (Mirrlees, 1976) and averages over all other taxes being chosen.

When analyzing the optimal taxes on fertility choices, a key insight I develop is the role that these taxes play in relaxing incentive constraints. On a superficial level, this result seems very similar to the original point made by Corlett and Hague (1953). However, the actual mechanism underlying my result is very different to the usual mechanism from commodity taxation, which relies on non-separabilities in the household utility function between commodity choices and labor supply (Atkinson and Stiglitz, 1976). In particular, my utility function is separable so that the usual mechanism for commodity taxation is shut down. The key difference, and the underlying force for taxation of fertility choices in my model, is that the ability parameter θ enters the utility function twice. The first is in the disutility of labor supply function ϕ, reflecting the standard interaction between income and ability. The second is in the sub-utility function v, which then generically directly links ability to the fertility choices. As described earlier, this interaction reflects a wide range of potential mechanisms governing the patterns of fertility choices in the population, from quantity-quality trade-off to heterogeneous tastes for children. It is this link that drives the optimality of distorting household fertility choices. An advantage of my analysis is that the exact mechanisms behind the link is irrelevant for the properties of the optimal taxes. All that matters are the elasticities that the mechanisms generate.

7 Implementation

While the wedges just analyzed are useful to build intuition for how households' choices are distorted by the planner in the optimal allocation, they do not give insights into how taxes might be used to implement the allocation itself. The design of such taxes is the topic of this section.

Formally, a tax system is a function $T : \mathbb{R}^4_+ \to \mathbb{R}$ that maps household choices (y, K, H, n) into a tax payable to the planner $T(y, K, H, n)$, where $T(y, K, H, n) < 0$ means that the household receives funds from the planner. A tax system implements the optimal allocation if and only if

$$\begin{aligned}(c(\theta), y(\theta), K(\theta), H(\theta), n(\theta)) \quad = \quad & \arg\max_{c,y,K,H,n} \quad u(c) - \phi\left(\tfrac{y}{\theta}, H\right) + v(n, K, H, \theta) \\ & \text{s.t.} \qquad\qquad c = y - K - T(y, K, H, n)\end{aligned}$$

for all household types $\theta \in \Theta$, i.e. if the tax system induces each household to choose the bundle prescribed to them in the optimal allocation chosen by the planner.

In order to define the set of tax systems that will implement the optimal allocation, I adapt a procedure described by Werning (2011). The idea is to find the "smallest" tax system that will implement the allocation, where "smallest" refers to the size of the taxes specified by the function T.

Proceeding constructively, first fix a type $\theta \in \Theta$, and consider a type specific tax system that ensures that a household of type θ always wants to choose her prescribed bundle $(c(\theta), y(\theta), K(\theta), H(\theta), n(\theta))$. In other words, given any bundle (y, K, H, n), set the tax on that bundle such that the household is indifferent between that bundle and her prescribed choice:

$$u\left(y - K - T^{\theta}(y, K, H, n, \theta)\right) - \phi\left(\frac{y}{\theta}, H\right) + v(n, K, H, \theta) = U(\theta)$$

where $U(\theta)$ is the household's utility from choosing her prescribed bundle in the optimal allocation. Rearranging this equation yields an expression for the "smallest" type-specific tax that will induce the type θ household to choose the correct bundle, where "smallest" refers to the indifference condition:

$$T^{\theta}(y,K,H,n,\theta) = y - K - u^{-1}\left(U(\theta) + \phi\left(\frac{y}{\theta},H\right) - v(n,K,H,\theta)\right)$$

Note that for $(y,K,H,n) = (y(\theta),K(\theta),H(\theta),n(\theta))$, the tax also satisfies the condition

$$c(\theta) = y(\theta) - K(\theta) - T^{\theta}(y(\theta),K(\theta),H(\theta),n(\theta),\theta)$$

Having defined a tax system for each type of household, the dependence on θ can be removed by taking the upper envelope of taxes at every bundle (y,K,H,n):

Proposition 9. *Given a set of "smallest" type specific tax systems $\{T(y,K,H,n,\theta)\}_{\theta\in\Theta}$, the tax system defined by*

$$T^{*}(y,K,H,n) = \sup_{\theta'} T^{\theta}\left(y,K,H,n,\theta'\right)$$

will implement the optimal allocation.

The proof builds on the fact that at any bundle outside of the optimal allocation, the tax is constructed to be large enough to deter any household from wanting to choose it, while at any $(y(\theta'),K(\theta'),H(\theta'),n(\theta'))$ in the optimal allocation,

$$T^{*}\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right)\right) = T^{\theta}\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right),\theta'\right) = y\left(\theta'\right) - K\left(\theta'\right) - c\left(\theta'\right)$$

, which together with incentive compatibility of the optimal allocation ensures that each household type chooses her prescribed bundle.

It is clear that T^{*} itself is the "smallest" tax system that will implement the optimal allocation since it was constructed from a set of "smallest" type specific tax systems. Therefore, any tax system that is at least as large at T^{*} everywhere and is equal to T^{*} on the optimal allocation will also implement it.

Proposition 10. *If T is a function of (y,K,H,n) such that $T(y,K,H,n) \geq T^{*}(y,K,H,n)$ everywhere, and $T(y,K,H,n) = T^{*}(y,K,H,n)$ for all (y,K,H,n) in the optimal allocation, then T will implement the optimal allocation.*

This result provides a clean way to assess the optimality of real policies that redistribute resources based on income, child quantity, and child investments, to which I now turn.

8 Quantitative Exercise

In this section, I implement my model numerically, in order to quantitatively analyze the optimal tax policies. I first construct a baseline economy and calibrate it to U.S. data, and then proceed to analyze the optimal taxes under this calibration.

8.1 Baseline Calibration

In the baseline economy, there is no social planner, and households simply face linear taxes and subsidies on income, child investments, and child quantity. For simplicity, I abstract from the child time-investment decision. Using the following functional form for utility, a household of ability θ solves

$$\max_{c,l,K,n} \frac{c^{1-\upsilon}}{1-\upsilon} - \eta \frac{l^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} + \psi \frac{(\theta K)^{1-\sigma_K}}{1-\sigma_K} + \frac{(\theta n)^{1-\sigma_n}}{1-\sigma_n} - \phi n$$

subject to

$$c + (1+\tau_K) K + \tau_n n = (1-\tau_y) \theta l$$

The main advantage of this specification of utility is that it provides clean expressions for the key elasticities mentioned in the theoretical analysis. In particular, it is simple to show the following:

Proposition 11. *Using this utility specification, the relevant elasticities can be computed as*

$$\epsilon^c(\theta) = \epsilon$$

$$\epsilon^K_\theta(\theta) = -\frac{\sigma_K - 1}{\sigma_K}$$

$$\epsilon^n_\theta(\theta) = -\frac{\sigma_n - 1}{\sigma_n}$$

This tight link between preference parameters and elasticities yields a transparent calibration method that I now describe.

8.1.1 Exogenously Calibrated Parameters

I set the coefficient of relative risk-aversion, $\upsilon = 1$, and the Frisch elasticity of labor supply, $\epsilon = 1$. I also set the linear tax on labor income $\tau_y = 0.2$ and the linear tax on child investment goods $\tau_K = 0$ so that it has the same relative price as the consumption good since a majority of goods are consumed by both parents and children within the household. All exogenously set parameters are summarized in Table 1.

Parameter	Value	Target
υ	1	CRRA = 1
ϵ	1	Frisch elasticity = 1
τ_y	0.2	
τ_K	0	

Table 1: Exogenously Set Parameters

8.1.2 Endogenously Matched Parameters

I assume that the distribution of ability is from the Pareto-lognormal family, and choose its parameters (mean μ, standard deviation Σ, and Pareto tail parameter α) so that the income distribution matches the income distribution of households with children from the US Census Bureau (2014). Relatedly, I set η so that on average, households spend a third of their total time endowment (normalized to 1) working.

I set the remaining parameters so that the baseline economy replicates salient features of the data related to fertility choices. Since ϕ scales the utility cost of having children, I set it so that the mean number of children born to a household is 1.9 (Sommer, 2016). Since σ_n uniquely determines the relationship between fertility and ability (and hence income), I set σ_n to match an income elasticity of fertility $\frac{\partial \log n}{\partial \log y} = -0.1$, which is in line with the evidence discussed in Jones et al. (2008) and the US Census Bureau (2017). In a similar manner to ϕ and σ_n, the parameters ψ and σ_K determine the scale and gradient of child expenditures. Therefore I set these parameters to match two moments documented by Lino (2014): households in the the bottom third of the income distribution spend 25% of their income on each child while households in the middle third spend 16% of their income per child. Finally, the linear tax on children is set at $\tau_n = -0.02$ (i.e. it is a subsidy) so that the fraction of GDP spent by the government on households with children is 0.5% in the baseline economy (OECD, 2017). All endogenously set parameters are summarized in Table 2.

Parameter	Value	Target	Source
μ	0.0011	Income distribution	Census (2014)
Σ	0.82		
α	40		
η	8	$\mathbb{E}[l] = 0.33$	
ϕ	0.23	$\mathbb{E}[n] = 1.9$	Sommer (2016)
σ_n	1.25	$\frac{\partial \log n}{\partial \log y} = -0.1$	Census (2017)
ψ	0.6	Mean per child exp. of bottom third	Lino (2014)
σ_K	1.09	Mean per child exp. of middle third	Lino (2014)
τ_n	-0.02	$\mathbb{E}[\tau_n n] / \mathbb{E}[y] = 0.5\%$	OECD (2017)

Table 2: Endogenously Matched Parameters

8.1.3 Pareto Weights

In order to compute optimal taxes, I need to choose a set of Pareto Weights. In my main results, I choose utilitarian weights, so that $\alpha(\theta) = 1$ for all $\theta \in \Theta$. This means that all redistribution is driven by the planner's desire to insure households against the risk of having low ability in earning and child rearing.

8.2 Results

8.2.1 The Optimal Wedges on Child Quantity, $\tau_n^*(\theta)$

Figure 1 shows the optimal wedges for child quantity, where, to ease interpretation, I have plotted the wedge as a fraction of median income. Three features stand out. First, the wedges are always negative which means that the optimal allocation features marginal subsidies to child quantity. The optimality of subsidies is justified by appealing to the intuition behind corollary 4. In the calibration, $\sigma_n > 1$ so that $\epsilon_\theta^n < 0$, i.e. higher ability households have fewer children. Furthermore, $\upsilon = 1$ so that income effects on labor supply are relatively weak. Therefore, the benefit of loosening incentive constraints by subsidizing child quantity outweighs the cost of lower labor supply. Second, the wedges have a pronounced U-shape, which is determined by the pattern of the term $\eta(\theta) \frac{1-F(\theta)}{\theta f(\theta)}$ in the expression for the optimal wedge: the subsidy is largest where the benefit from redistribution most outweighs the cost of distorting household choices. Finally, the wedges are quantitatively meaningful and take an average value of 3% of median annual income, which translates to \$2400/year using my empirical income distribution. This magnitude is comparable to real child credit policies in the US and other countries.

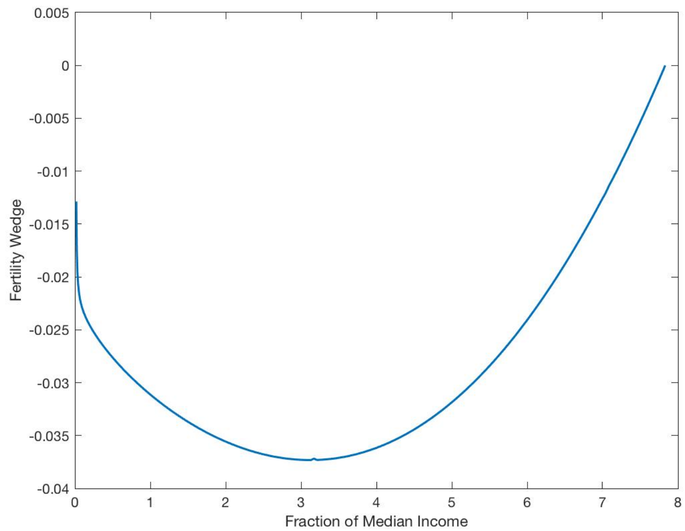

Figure 1: Optimal Wedges on Child Quantity

8.2.2 The Optimal Wedges on Child Investment, $\tau_K^*(\theta)$

Figure 2 shows the optimal wedges for child investment, where again, I have plotted the wedge as a fraction of median income. Similar to the child quantity wedge, there are three key features. First, the wedges are always negative which means that the optimal allocation features marginal subsidies to child investment. The optimality of subsidizing child investment on the margin follows a similar logic to the child quantity wedge. In the calibration, $\sigma_K > 1$ so that $\epsilon_\theta^K < 0$, i.e. higher ability households spend less in total on their children. Combining this with the weak income effect on labor supply implies that subsidizing child investment is optimal, since the benefit of loosening incentive constraints outweighs the cost of reducing labor supply. Second, and unlike the quantity wedges, the child investment wedges are predominantly increasing in household income, so that richer households receive a smaller marginal subsidy than poorer households. In this sense, the child investment wedges are progressive. This progressivity is driven by the fact that income effects matter more to the planner for households with higher ability and hence income: discouraging the labor supply of a high ability household has a larger effect on the aggregate resource constraint (and hence total redistribution) than discouraging the labor supply of a lower ability household. Therefore it is optimal to set smaller subsidies of child investment for higher ability households. Finally, the wedges are quantitatively significant, and average 9% of median annual income, which translates to \$7,300/year using my empirical income distribution.

9 Implementation and Policy Comparisons

9.1 Implementation: computing a version of $T^*(y, K, n)$

As discussed earlier, the wedges alone are insufficient to understand what taxes can implement the optimal allocation. Furthermore, while the function T^* provides a kind of lower bound on how large

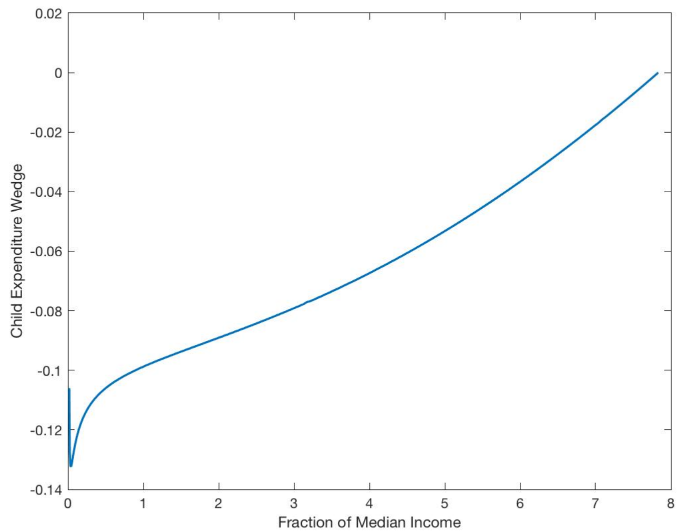

Figure 2: Optimal Wedges on Child Investment

taxes must be at every possible bundle (y, K, n), it is a fairly abstract object with little real-world motivation. Therefore, I now consider a tax function T that is everywhere equal to T^*, but can much more easily be connected to policies we observe in reality.

In the USA, a household's total tax bill is computed by subtracting any available credits away from the income tax bill. Therefore I include as the first component of my tax function T, an income tax function $T_0(y)$, that matches the real income tax function used for Head of Household filers in the US in 2016 (Tax Foundation). The table below shows how these income taxes are computed.

Lower Bracket ($)	Upper Bracket ($)	Marginal Tax Rate
0	13,250	0.1
13,250	50,400	0.15
50,400	130,150	0.25
130,150	210,800	0.28
210,800	413,350	0.33
413,350	441,000	0.35
441,000	$+\infty$	0.396

Table 3: US Income Tax Parameters (2016)

The second component, T_1, is simply computed as the residual tax bill,

$$T_1(y, K, n) = T^*(y, K, n) - T_0(y)$$

and hence represents total credits received by the household in the case $T_1 < 0$. Note that T_1 must depend on all three choices (y, K, n) to ensure that equality can be obtained for all possible bundle choices.

In sum, I consider a tax function

$$T(y, K, n) = T_0(y) + T_1(y, K, n)$$

that is everywhere equal to $T^*(y, K, n)$ and hence implements the optimal allocation.

9.2 A Review of Tax Policy Towards Families in the USA

Tax policy towards families in the USA can broadly be split into two key categories: income tax credits, and health care subsidies. In each of these categories, the most prominent programs are the EITC and CTC, and Medicaid.

EITC and CTC As described in the introduction, the EITC has significant dependence on both income and child quantity within a household. Analysis by Hoynes and Rothstein (2016) shows that the program is targeted at low income households, with credits increasing in the number of children. The CTC is similar to the EITC in design, but has much larger reach in terms of income eligibility - households with income in excess of $100,000/year are still eligible for CTC payments. In the same paper, Hoynes and Rothstein comment that this scope seems puzzling and perhaps goes against the redistributive aims of the government.

Medicaid Medicaid provides subsidized healthcare coverage for families whose incomes are below 138% of the Federal Poverty Line for the relevant household size. If a household is eligible, I model Medicaid as a payment of $2577 per child to the household, which was the estimated expenditure per child in the USA in 2014 (Henry J Kaiser Family Foundation).

9.3 Results

Figure 3 compares the tax credits implied by the function T_1 to the tax credits implied by the the real policies just described. The figure plots the credits as a function of household income, holding the number of children fixed at $n = 2$. Since the function T_1 also depends on child expenditures, I fix K at its optimal value for each income. For completeness, other cases are considered in the appendix. The top panel shows results for all households, while the bottom panel shows only the bottom 90% of households in the income distribution.

The key result is that for all income levels, credits in the model are substantially larger than the credits implied by real policies. While the large credits at the top of the distribution are an artifact of the exogenous income tax function I imposed as part of my implementation, the large credits at the bottom are robust to other income tax functions since income taxes are necessarily close to zero when income is close to zero.

The large magnitude of credits in the optimal allocation follows from the logic outlined in the analysis of the optimal wedges. Subsidizing child quantity and expenditures on the margin not only mechanically increases the total credits received by poorer households, but also loosens the incentive constraints of higher ability households, which allows the planner to increase marginal income tax rates, and afford even more distribution towards lower ability households. Quantitatively, $\sigma_K > \sigma_n$ implies that subsidizing child expenditure is a relatively more efficient way to achieve this redistribution, though both choices are distorted in the optimal allocation.

This logic suggests that real credits are too small relative to the optimum because they do not fully exploit the gains from subsidizing both child quantity and child investment. In particular, none of the real policies mentioned have any dependence on child expenditure, and thus completely forgo the associated gains from subsidization, while the marginal subsidies of child quantity seem too small.

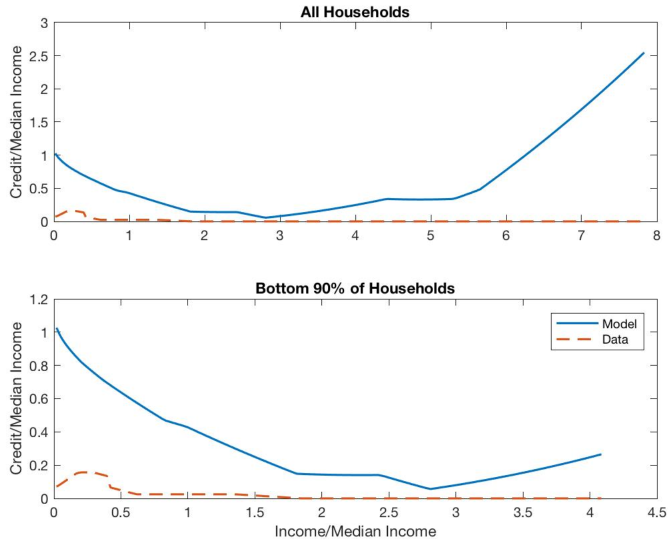

Figure 3: Tax Credits for households with two children

10 Welfare Gains From Simple Policies

The numerical exercises suggest that there are large welfare gains from subsidizing child-related investments and child quantity choices. While child quantity is arguably verifiable by a tax authority, it is much harder to make the same case for child expenditures, which in reality may span a wide range of goods and services, some of which will naturally overlap with personal expenditures (e.g. food and housing). When child expenditures are treated as anonymous by the social planner, non-linear taxation becomes infeasible, and linear taxes or subsidies must be considered instead. In this section, I investigate how much of the welfare gain from non-linear taxation of child expenditures can be achieved using a linear tax rate.

10.1 Augmented Planning Problem

The planner chooses an allocation $\{c(\theta), y(\theta), K(\theta), n(\theta)\}_{\theta\in\Theta}$ subject to both the same information constraints as before, and also behavioral constraints that describe how a household makes choices when confronted with a linear tax on child expenditures. Following Boadway and Jacobs (2014), it is useful to consider a household choosing consumption and child expenditures for given choices of income and child quantity,

$$\max_{c,K} u(c) - \phi\left(\frac{y}{\theta}\right) + v(n, K, \theta)$$

subject to

$$c + (1+\tau_K) K = y - T(y, n)$$

where T is a tax function that depends only on income and child quantity, and arises from the implementation of the solution to the planning problem.

Solving this problem yields conditional demand functions, $\tilde{c}(\theta, y, n, \tau_K)$ and $\tilde{K}(\theta, y, n, \tau_K)$, that specify how a household will choose her consumption and child expenditures for given levels of labor productivity, income, child quantity, and the tax rate on child expenditures. The planner must take this behavior into account when solving the planning problem. Specifically, the planner chooses an allocation $A = \{c(\theta), y(\theta), n(\theta), K(\theta)\}_\Theta$ and a tax rate τ_K that solve

$$\max_{A,\tau_K} \int_\Theta \alpha(\theta) U(\theta) f(\theta) d\theta$$

subject to

$$\begin{aligned}
U(\theta) &= u(c(\theta)) - \phi\left(\frac{y(\theta)}{\theta}\right) + \beta v(n(\theta), K(\theta), \theta) \\
\theta &= \arg\max_{\theta'} U(\theta) \\
c(\theta) &= \tilde{c}(\theta, y, n, \tau_K) \\
K(\theta) &= \tilde{K}(\theta, y, n, \tau_K) \\
\int_\Theta (c(\theta) + K(\theta)) f(\theta) d\theta &\leq \int_\Theta y(\theta) f(\theta) d\theta
\end{aligned}$$

10.2 Quantitative Results

The optimal linear tax is $\tau_K^* = -0.3$, so that child expenditure is subsidized, as in the non-linear case. This number is also very close to the mean subsidy rate in the unrestricted model, which turns out to be -0.29. Figure 4 plots the optimal allocations in the cases when taxes are left unrestricted and when the tax on child expenditures is constrained to be linear. It is clear that the allocations are very similar even when non-linear taxation of child expenditures is ruled out, indicating that a large fraction of the welfare gains from taxing these choices can be achieved through the simple policy of a linear subsidy on child expenditures, accompanied by non-linear taxation of income and child quantity. Indeed, comparing social welfare under each regime indicates that the restricted taxes obtain 99% of the welfare gain from non-linear taxation.

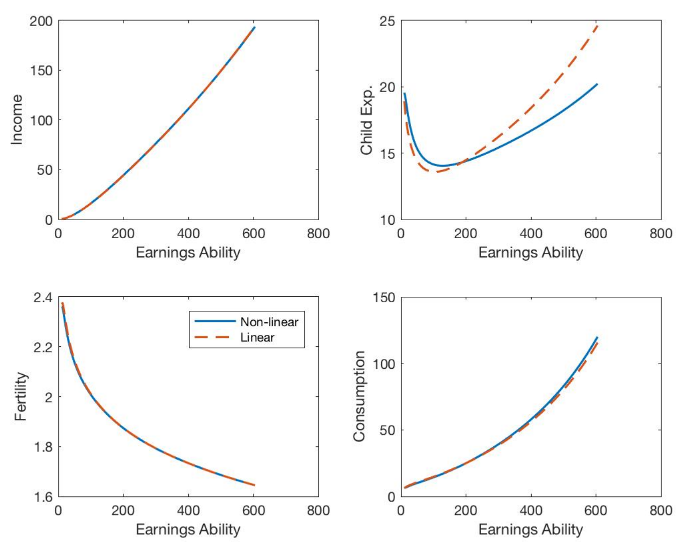

Figure 4: Optimal Allocations

11 Conclusion

I have presented a general framework for analyzing the optimal tax treatment of choices related to fertility. My model is the first to adopt a truly Mirrleesian approach to fertility taxation, and as such offers multiple advantages over the existing literature. First, in the face of a plethora of potential underlying mechanisms governing the link between fertility choices and other household choices such as income, my model delivers expressions for optimal wedges that are robust to a wide range of mechanisms by expressing wedges in terms of reduced form elasticities that are estimable given the relevant data. Second, my expressions give clear intuition for the shape and sign of the optimal distortions on child quantity and investments in goods and time, and how these properties relate to the elasticities. Finally, the quantitative exercise indicates that large welfare gains are achievable by taxing child-related choices in an optimal manner, and that most of these gains are obtainable using feasible policies.

A natural limitation of this study stems from the Mirrleesian approach I take towards analyzing optimal taxes. In the model I present, all choice variables are deterministic and monotonic functions of the household type parameter θ. This means that child quantity is a deterministic and decreasing function of household income. While this relationship captures the general negative correlation between income and fertility well, it imposes too much structure on the data that the model generates, since it becomes impossible for some high income households to have as many children as some lower income households, an outcome that is certainly observed in the data. Further, reinterpreting the variable n as the average number of children in households of a fixed type is unsatisfactory since it would make the corresponding taxes a function of average child quantity, which then becomes difficult to reconcile with real policies. A possible solution to this issue would be to let child quantity be a random variable that the household can only partially control, for example by choosing the mean value. This might reflect the uncertainty associated with child conception and would certainly allow the model to generate more realistic relationships between child quantity and other outcomes. The taxes could also then be adjusted so that taxes are a function of number of children actually born rather than the household's choice of the mean value. This is an interesting extension that I hope to pursue in later work.

References

[1] Alvarez, F. (1999). Social Mobility: The Barro-Becker Children Meet the Laitner-Loury Dynasties. *Rev. Econ. Dyn.*, 2(1):65–103.

[2] Angrist, J. D., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *J. Labor Econ.*, 28(4):773–824.

[3] Attanasio, B. O., Low, H., and Virginia, S. (2008). Explaining Changes in Female Labor Supply in a Life-Cycle. *Am. Econ. Rev.*, 98(4):1517–1552.

[4] Balestrino, A., Cigno, A., and Pettini, A. (2002). Endogenous Fertility and the Design of Family Taxation. *Int. Tax Public Financ.*, 9(2):175–193.

[5] Baughman, R. and Dickert-Conlin, S. (2009). The earned income tax credit and fertility. *J. Popul. Econ.*, 22(3):537–563.

[6] Becker, G. S. and Lewis, H. G. (1974). Interaction between Quantity and Quality of Children. *NBER Chapters*, I:81–90.

[7] Becker, G. S., Murphy, K. M., and Tamura, R. (2016). Human Capital, Fertility, and Economic Growth. 98(5).

[8] Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education*. *Q. J. Econ.*, 120(2):669–700.

[9] Blundell, R., Pistaferri, L., and Saporta-Eksten, I. (2015). Children, Time Allocation and Consumption Insurance. 284024(March):1–38.

[10] Boca, D. D. and Wiswall, M. (2010). Household Choices and Child Development. *Work. Pap.*, (5155):1–56.

[11] Boca, D. D. and Wiswall, M. (2014). Transfers to Households with Children and Child Development. (8393).

[12] Cigno, A. (2001). Comparative advantage, observability, and the optimal tax treatment of families with children. *Int. Tax Public Financ.*, 8(4):455–470.

[13] Cigno, A., Luporini, A., and Pettini, A. (2003). Transfers to families with children as a principal-agent problem. *J. Public Econ.*, 87(5-6):1165–1177.

[14] Cigno, A. and Pettini, A. (2002). Taxing family size and subsidizing child-specific commodities? *J. Public Econ.*, 84(1):75–90.

[15] Córdoba, J. C. and Ripoll, M. (2014). The Elasticity of Intergenerational Substitution, Parental Altruism, and Fertility Choice. (15).

[16] Cremer, H., Dellis, A., and Pestieau, P. (2003). Family size and optimal income taxation. *J. Popul. Econ.*, 16(1):37–54.

[17] Cremer, H., Gahvari, F., and Pestieau, P. (2006). Pensions with endogenous and stochastic fertility. *J. Public Econ.*, 90(12):2303–2321.

[18] Cremer, H., Gahvari, F., and Pestieau, P. (2011). Fertility, human capital accumulation, and the pension system. *J. Public Econ.*, 95(11-12):1272–1279.

[19] Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

[20] D'Addio, A. C. and Mira d'Ercole, M. (2005). Policies, Institutions and Fertility Rates: a Panel Data Analysis for OECD Countries. (41):7–43.

[21] Dahl, B. G. B. and Lochner, L. (2012). The Impact of Family Income on Child Achievement : Evidence from the Earned Income Tax Credit. 102(5):1927–1956.

[22] De La Croix, D., Doepke, M., Azariadis, C., Cole, H., Farmer, R., Hansen, G., Ohanian, L., and Schneider, M. (2002). Inequality and Growth: Why Differential Fertility Matters.

[23] Department of Health & Human Services (2016). 2016 Federal Poverty Level Chart. Technical report.

[24] Diamond, P. A. (1998). Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *Am. Econ. Rev.*, 88(1):83–95.

[25] Guryan, J., Hurst, E., and Kearney, M. (2008). Parental Education and Parental Time with Children. *Source J. Econ. Perspect.*, 22(3):23–46.

[26] Hanushek, E. (1992). The Trade-Off between Child Quality and Quantity.

[27] Ho, C. and Pavoni, N. (2016). Efficient child care subsidies. *Work. Pap.*

[28] Hotz, V. J., Klerman, J. A., and Willis, R. J. (1997). The Economics of Fertility in Developed Countries: a Survey. *Handb. Popul. Fam. Econ.*, (7):275–347.

[29] Hoynes, H. and Rothstein, J. (2016). Tax Policy Toward Low-Income Families.

[30] Jacobs, B. and Boadway, R. (2014). Optimal linear commodity taxation under optimal non-linear income taxation. *J. Public Econ.*, 117:201–210.

[31] Jones, L. E. and Schoonbroodt, A. (2008). Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?

[32] Lino, M. (2014). Expenditures on Children by Families, 2013. (1528).

[33] Marr, C., Huang, C.-c., Sherman, A., and Debot, B. (2015). EITC and Child Tax Credit Promote Work, Reduce Poverty and Support Children's Development. 000:23–27.

[34] Mirrlees, J. (1971). Exploration in optimum income the theory. *Rev. Econ. Stud.*, 38(2):175–208.

[35] OECD (2017). Public spending on family benefits Public. Technical report.

[36] Rosenzweig, M. R. and Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy. *Rev. Econ. Stud.*, 76(3):1149–1174.

[37] Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Rev. Econ. Stud.*, pages 205–229.

[38] Sommer, K. (2016). Fertility choice in a life cycle model with idiosyncratic uninsurable earnings risk. *J. Monet. Econ.*, 83:27–38.

[39] Stantcheva, S. (2015). Optimal Taxation and Human Capital Policies over the Life Cycle. *Work. Pap.*

[40] Thevénon, O. (2011). Family Policies in OECD countries: A comparative Analysis. *Popul. Dev. Rev.*, 37(1):57–87.

[41] Werning, I. (2011). Nonlinear Capital Taxation. 2011:1–20.

A The Planning Problem

Recall the baseline planning problem using the First Order Approach:

$$\max_{A} \int_{\Theta} \alpha(\theta) U(\theta) f(\theta) d\theta$$

subject to

$$\begin{aligned}
U(\theta) &= u(c(\theta)) - \phi\left(\frac{y(\theta)}{\theta}, H(\theta)\right) + \beta v(n(\theta), K(\theta), H(\theta), \theta) \\
U_\theta(\theta) &= \frac{y(\theta)}{\theta^2}\phi_l\left(\frac{y(\theta)}{\theta}, H(\theta)\right) \\
&\quad +\beta v_\theta(n(\theta), K(\theta), H(\theta), \theta) \\
\int_\Theta (c(\theta) + K(\theta)) f(\theta) d\theta &\leq \int_\Theta y(\theta) f(\theta) d\theta
\end{aligned}$$

This is a standard optimal control problem and so can be solved using Hamiltonian techniques: substitute the first constraint into the aggregate resource constraint replacing the $c(\theta)$ term, and attach multipliers $\mu(\theta)$ to the differential equation constraint, and λ to the resource constraint. The Hamiltonian then reads

$$\begin{aligned}
\mathcal{H} &= \alpha(\theta) U(\theta) f(\theta) \\
&+ \lambda\left(y(\theta) - u^{-1}\left(U(\theta) + \phi\left(\frac{y(\theta)}{\theta}, H(\theta)\right) - \beta v(n(\theta), K(\theta), H(\theta), \theta)\right) - K(\theta)\right) f(\theta) \\
&+ \mu(\theta)\left(\frac{y(\theta)}{\theta^2}\phi_l\left(\frac{y(\theta)}{\theta}, H(\theta)\right) + \beta v_\theta(n(\theta), K(\theta), H(\theta), \theta)\right)
\end{aligned}$$

Adopting abbreviated notation for derivatives, the First Order Conditions (FOCs) are as follows:

$$\begin{aligned}
U(\theta): &\quad \alpha(\theta) f(\theta) - \lambda\frac{1}{u'(\theta)} f(\theta) &= -\mu_\theta(\theta) \\
y(\theta): &\quad \lambda\left(1 - \frac{\phi_l(\theta)}{\theta u'(\theta)}\right) f(\theta) + \mu(\theta)\left(\frac{1}{\theta^2}\phi_l(\theta) + \frac{y(\theta)}{\theta^3}\phi_{ll}(\theta)\right) &= 0 \\
n(\theta): &\quad \lambda\left(\frac{\beta v_n(\theta)}{u'(\theta)}\right) f(\theta) + \mu(\theta)\beta v_{n\theta}(\theta) &= 0 \\
K(\theta): &\quad \lambda\left(\frac{\beta v_K(\theta)}{u'(\theta)} - 1\right) f(\theta) + \mu(\theta)\beta v_{\theta K}(\theta) &= 0 \\
H(\theta): &\quad \lambda\left(\frac{\beta v_H(\theta) - \phi_H(\theta)}{u'(\theta)}\right) f(\theta) + \mu(\theta)\beta v_{\theta H}(\theta) &= 0
\end{aligned}$$

Noting that $\lim_{\theta\to+\infty}\mu(\theta) = \lim_{\theta\to 0}\mu(\theta) = 0$, and recalling that $\mathbb{E}[\alpha(\theta)] = 1$, integrating the first condition over all of Θ yields $\frac{1}{\lambda} = \mathbb{E}\left[\frac{1}{u'(c(\theta))}\right]$. Similarly, integrating from any θ to $+\infty$ yields $\mu(\theta) = -\int_\theta^\infty \left(\frac{\lambda}{u'(c(v))} - \alpha(v)\right) f(v) dv$. Given this, define $\eta(\theta) = -\mu(\theta)/(1 - F(\theta))\lambda$. To derive the elasticity terms, consider a household of ability θ facing wedges τ_y, τ_n, τ_K, and τ_H. From the main text, I have that

$$\begin{aligned}\tau_y &= 1-\tfrac{\phi_l(y/\theta,H)}{\theta u'(c)}\\ \tau_n &= \tfrac{\beta v_n(n,K,H,\theta)}{u'(c)}\\ \tau_K &= \tfrac{\beta v_K(n,K,H,\theta)}{u'(c)}-1\\ \tau_H &= \tfrac{\beta v_H(n,K,H,\theta)-\phi_H(y/\theta,H)}{u'(c)}\end{aligned}$$

In each case, I now consider the partial derivative of variable $j \in (y, n, K, H)$ with respect to θ and τ_j, holding all other terms fixed.

For income, it is useful to recall that $y = \theta l$, and treat l as the variable of interest. Following Stantcheva (2014), I then define

$$\epsilon^u(\theta) = \frac{\phi_l(l,H)/l + \theta^2 u''(c)}{\phi_{ll}(l,H) - \theta^2 u''(c)}$$

$$\epsilon^c(\theta) = \frac{\phi_l(l,H)/l}{\phi_{ll}(l,H) - \theta^2 u''(c)}$$

which immediately lead to the expressions in the text, noting that $\epsilon^u = \epsilon^c$ when $u'' = 0$.

In the case of n, I differentiate the household FOC with respect to θ,

$$\beta v_{nn}(\theta)\frac{\partial n}{\partial \theta} + \beta v_{n\theta}(\theta) = \tau_n u''(c)\frac{\partial c^n}{\partial \theta}$$

where

$$\frac{\partial c^n}{\partial \theta} = l(1-\tau_y) - \tau_n \frac{\partial n}{\partial \theta}$$

Similarly for τ_n,

$$\beta v_{nn}(\theta)\frac{\partial n}{\partial \tau_n} = u'(c) + \tau_n u''(c)\frac{\partial c^n}{\partial \tau_n}$$

where

$$\frac{\partial c^n}{\partial \tau_n} = -n - \tau_n \frac{\partial n}{\partial \tau_n}$$

Combining these expressions yields

$$\beta\theta v_{n\theta}(\theta) = \theta\tau_n u''(c)\frac{\partial c^n}{\partial \tau_n}\left(\left(\frac{\partial c^n}{\partial \theta}/\frac{\partial c^n}{\partial \tau_n}\right) - \left(\frac{\partial n}{\partial \theta}/\frac{\partial n}{\partial \tau_n}\right)\right) - \theta u'(c)\frac{\partial n}{\partial \theta}/\frac{\partial n}{\partial \tau_n}$$

which can be expressed in terms of the relevant elasticities,

$$\beta\theta v_{n\theta}(\theta) = u'(c)\left(-\epsilon^n_\theta/e^n_{\tau_n} - \left(-\frac{u''(c)\,c}{u'(c)}\right)\epsilon^{c,n}_{\tau_n}\left(\left(\epsilon^{c,n}_\theta/e^{c,n}_{\tau_n}\right) - \left(\epsilon^n_\theta/e^n_{\tau_n}\right)\right)\right)$$

Substituting into the FOC of the program yields the expression for the optimal wedge.

In the case of K, first differentiate with respect to θ,

$$\beta v_{KK}(\theta)\frac{\partial K}{\partial \theta} + \beta v_{\theta K}(\theta) = (1+\tau_K)\,u''(c)\frac{\partial c^K}{\partial \theta}$$

and then with respect to $1+\tau_K$,

$$\beta v_{KK}(\theta)\frac{\partial K}{\partial(1+\tau_K)}=u'(c)+(1+\tau_K)u''(c)\frac{\partial c}{\partial(1+\tau_K)}$$

Combining yields

$$\beta\theta v_{\theta K}(\theta)=u'(c)\left(-\frac{\epsilon_\theta^K}{e_{\tau_K}^K}-\left(\frac{-u''(c)c}{u'(c)}\right)\epsilon_{1+\tau_K}^{c,K}\left(\frac{\epsilon_\theta^{c,K}}{e_{\tau_K}^{c,K}}-\frac{\epsilon_\theta^K}{e_{\tau_K}^K}\right)\right)$$

which can be substituted into the relevant FOC of the program to yield the expression for the optimal wedge.

Finally, consider H. Differentiating with respect to θ,

$$\beta v_{HH}(\theta)\frac{\partial H}{\partial\theta}+\beta v_{\theta H}(\theta)-\phi_{HH}(\theta)\frac{\partial H}{\partial\theta}=\tau_H u''(c)\frac{\partial c^H}{\partial\theta}$$

and then with respect to τ_H,

$$\beta v_{HH}(\theta)\frac{\partial H}{\partial\tau_H}-\phi_{HH}(\theta)\frac{\partial H}{\partial\tau_H}=u'(c)+\tau_H u''(c)\frac{\partial c^H}{\partial\tau_H}$$

Combining yields

$$\beta\theta v_{\theta H}(\theta)=u'(c)\left(-\frac{\epsilon_\theta^H}{e_{\tau_H}^H}-\left(-\frac{u''(c)c}{u'(c)}\right)\epsilon_{\tau_H}^{c,H}\left(\frac{\epsilon_\theta^{c,H}}{e_{\tau_H}^{c,H}}-\frac{\epsilon_\theta^H}{e_{\tau_H}^H}\right)\right)$$

which can be substituted into the relevant FOC of the program to yield the expression for the optimal wedge.

B Proof of Proposition 9

Proof. First consider a bundle (y,K,H,n) not in the optimal allocation. By construction, the tax payable by a household of type θ is at least the tax payable under the type specific tax system so that the household is weakly better off by choosing her bundle in the optimal allocation. Now, consider a bundle in the optimal allocation, $(y(\theta'),K(\theta'),H(\theta'),n(\theta'))$ for some $\theta'\in\Theta$. I claim that

$$T^*\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right)\right)=T\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right),\theta'\right)$$

i.e. that the θ' type household faces the largest tax bill at the θ'-bundle in the optimal allocation. By incentive compatibility of the optimal allocation, I know that for another type $\theta''\neq\theta'$

$$\begin{aligned}&u\left(y(\theta')-K(\theta')-T\left(y(\theta'),K(\theta'),H(\theta'),n(\theta'),\theta'\right)\right)\\&-\phi\left(\tfrac{y(\theta')}{\theta''},H(\theta')\right)+v\left(n(\theta'),K(\theta'),H(\theta'),\theta''\right)\qquad\leq\quad U(\theta'')\end{aligned}$$

where

$$T\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right),\theta'\right)=y\left(\theta'\right)-K\left(\theta'\right)-c\left(\theta'\right)$$

by construction of the type specific tax system for θ'. Therefore, to make any θ'' type household indifferent between this bundle and her optimal allocation bundle, it must be that

$$T\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right),\theta''\right)\leq T\left(y\left(\theta'\right),K\left(\theta'\right),H\left(\theta'\right),n\left(\theta'\right),\theta'\right)$$

thus establishing the claim. In particular, this means that any household of a type $\theta''\neq\theta'$ is weakly better off choosing her own bundle in the optimal allocation, which completes the implementation proof. □

C Policy Comparisons

Figures 5 and 6 compare the optimal credits from the model to the policies in the US for the cases of three children and 1 child respectively. The patterns are similar to the case of two children, as described above.

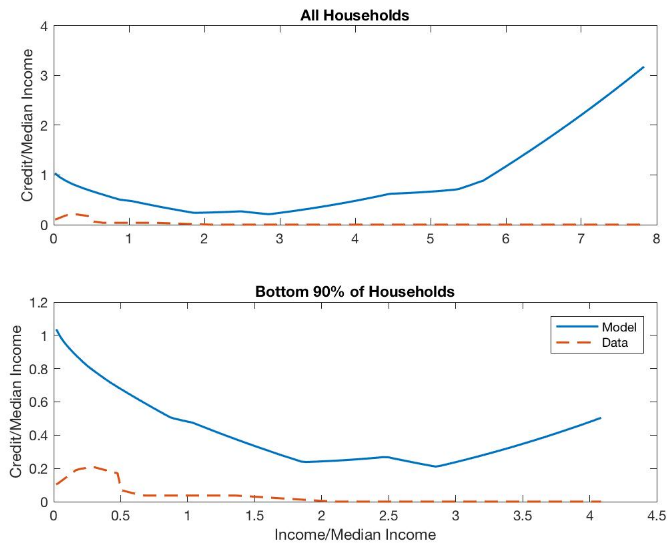

Figure 5: Tax Credits for households with two children

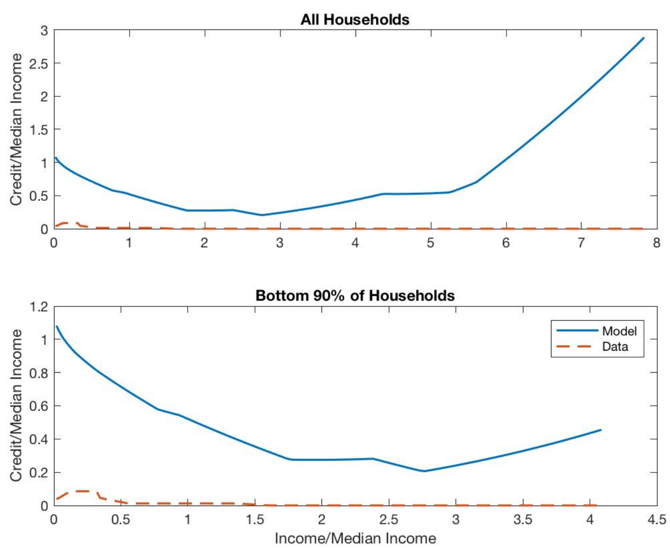

Figure 6: Tax Credits for households with two children