The Quantity-Quality Trade-off Revisited: A Case for Taxing Children

Tomer Blumkin*   Yoram Margalioth♠ Efraim Sadka♦

April, 2011

Abstract

In this paper, we challenge the conventional wisdom that due to the negative correlation between family size and earning ability, family size can be used as a 'tagging' device, so that subsidizing children (via child allowances) enhances egalitarian objectives. We show that the case for subsidizing children crucially hinges on child allowances being provided on a universal basis. Notably, when child benefits are means-tested, taxing children at the margin (namely, setting total child benefits to decline with the number of children) is optimal under a broad class of egalitarian objectives.

JEL Classification: D6, H2, H5

Key Words: child allowance, re-distribution, means-testing, universal, tagging, optimal taxation

* Department of Economics, Ben-Gurion University, Beer-Sheba 84105, Israel, CesIfo, IZA. E-mail: tomerblu@bgumail.bgu.ac.il.

♠ The Buchman Faculty of Law, Tel-Aviv University, Tel-Aviv 69978, Israel. E-mail: margalio@post.tau.ac.il

♦ The Eitan Berglas School of Economics, Tel Aviv University, Tel-Aviv 69978, Israel, CesIfo, IZA. E-mail: sadka@post.tau.ac.il

1. Introduction

Family size is a key component in the determination of income tax liability in all OECD countries [see, e.g., Bradshaw and Finch (2002)]. Two major decisions affect family size: (i) marriage/cohabitation and (ii) fertility. In this paper we focus on the optimal fiscal treatment of children [for the fiscal treatment of the former see, for instance, two recent papers by Cremer et al (2009) and Kleven et al (2009)]. In practice, the existence of children generally reduces the household's tax liability. This may take a variety of forms, including: income splitting amongst (a standardized number of) family members (as in France); exemptions or standard deductions (as in the US); specific childcare deductions; tax credits; and the provision of child allowances, which could be either universal or means-tested.[1] In most countries the policy implemented is a mixture of some or all of the above measures.

The economic rationale underlying the preferential tax treatment of children is based on the following four key arguments. First, the existence of children raises the question of horizontal equity, which in the family size context implies that the tax liability of a household, which is determined based on its ability to pay (say, measured by the level of income), should also account for the (standardized) number of family members. According to this view, children are not a form of consumption good of their parents, but rather part of the tax-paying unit. As such, they reduce the

[1] Nearly all developed countries provide universal child allowances; namely, child allowances that do not depend on household's income (but may well vary with the number of children) with the notable exception of the US, where the child allowances system (which is embedded in the EITC program) is (partly) means-tested. Thus, for example, in 2009, for the income range of 0-5,970 USD, a household with no children is entitled to a wage subsidy of 7.65%, whereas, a household with one child is eligible for a wage subsidy of 34% (within the same income range households with 2 and 3 children are entitled to a wage subsidy of 40% and 45%, respectively). In this income range, as well as in the income phase-out ranges that are also structured with different rates for different numbers of children, child allowances are means-tested. However, parts of the program are universal. For example, the difference between the allowance of households with 3 children and that of households with 2 children (with the same level of income) is constant for all levels of income above 12,570 USD; and in the overlapping income ranges, in which the child allowance is fixed (the plateaus), the program is obviously universal, as the allowances depend only on the number of children and not on income.

disposable income per-capita, hence the ability-to-pay of the household [for incorporation of horizontal equity considerations into the design of optimal tax-transfer systems see, for instance, Balcer and Sadka (1982) and (1986)]. A second argument draws on demographic considerations (primarily, those related to the looming pension crisis in many countries). The sharp increase in dependency ratios, stemming both from the drop in fertility rates and the corresponding increase in life expectancy, is casting a shadow on the financial sustainability of many national pension systems. The economic rationale for providing child-related subsidies in this case is essentially *Pigouvian*: subsidies are aimed at internalizing fiscal externalities.[2] The third argument warrants the provision of child-related subsidies as a means to motivate women to participate in the labor market. This is achieved by the provision of subsidized child care services (as in Sweden), deduction of the costs of daycare,[3] and provision of child tax credits that are limited to mothers' income tax liability (as in Israel and the UK). The fourth argument, which is the focus of the current paper, justifies the use of child-related subsidies on re-distributive grounds. Family size is used as an efficient indicator ['tagging' device, a la Akerlof (1978)] for the earning capacity of the household. According to the quality/quantity paradigm [see the pioneering studies of Becker (1960) and Becker and Lewis (1973)], low-ability families may choose to 'specialize' in quantity, that is, to raise more children relative

[2] France, Sweden and Quebec, are notable examples of countries that have implemented policies with the explicit goal of enhancing fertility [see, e.g., Laroque and Salanie (2008)].

[3] The cost of daycare is a business related expense of parents with young children who need someone to look after their children when they are at work. It is a business and personal (that is, consumption) mixed cost, as daycare provides value beyond the mere safekeeping of the children. According to basic income tax principles, the business part of it should be deducted in computing the taxable income of the parent (as is done, for example, in Canada and in Germany). Most countries do not allow an outright deduction due to the difficulty of separating the business and consumption elements, but instead reach some sort of a compromise such as the exclusion of employer provided child care (or daycare expenses reimbursed by the employer) from taxable income (as in the US, the Netherlands and Japan), or providing a credit in lieu of deduction (as in France).

to higher-ability households.[4] In such a case, in a second-best setting [a la Mirrlees, (1971)], where earning abilities are un-observed by the government; subsidizing larger families can promote a re-distributive goal.

Indeed, a relatively recent strand in the optimal income tax literature examines the potential supplementary re-distributive role of extending the tax base to account for the number of children in the household and child-related consumption (such as, education and daycare). For a comprehensive recent review of the literature, see Cigno (2009). This literature challenges some of the key results of the optimal income tax literature, such as, the desirability of a zero marginal tax rate levied on top-earners [see Phelps (1973) and Sadka (1976)] when the skill distribution is bounded, the redundancy of commodity taxation [see Atkinson and Stiglitz (1976) and Mirrlees (1976)], as well as, the conventional wisdom in tax policy design that the existence of children merits a reduction in tax liability (that is, children being a tax asset for their parents). The literature emphasizes a key distinction from the standard optimal tax setting, which derives from the unique characteristics of children: a crucial part of the process of rearing children may be viewed as consumption of a non-transferable domestically produced good (e.g., parental attention and affection), the production of which requires expertise (ability to nurture) that is fundamentally different from the ability to earn (the single source of variation across households in the standard optimal tax setting). The introduction of a second source of heterogeneity (alongside variation in earning ability) bears new re-distributive implications, affecting both policy goals and system design. In particular, it is shown that the direction of re-distribution is not necessarily in favor of the low earning-ability individuals, because the latter may enjoy some marked advantage in child-rearing, which may, all-in-all,

[4] For evidence of the existence of a quality-quantity trade-off see, e.g. Hanushek (1992).

compensate (in utility terms) for their low earning capacity. Moreover, the tax system design can employ observed family attributes to enhance target efficiency ('tagging'). The properties of the optimal integrated tax-transfer system (which allows the tax liability to depend on income level, family size as well as expenditure on child-related goods) are generally shown to depend on both comparative- and absolute-advantage (in domestic vis-à-vis market production) considerations.

In this paper, we address the key policy issue of the optimal tax treatment of children. Employing a continuum version of the two-household framework used by Cigno (1986) and (2001) we derive the properties of the general optimal income tax cum child benefit system set by an egalitarian government. We maintain the key assumption in this literature, that children are a form of consumption good which is time-consuming from the point of view of the parents. We follow the standard assumption in the optimal tax literature that households differ only in their earning abilities.[5] As this general system allows for the possibility of making the level of child benefits dependent on the household's level of income, we will henceforth refer to it as a means-tested system. The special case of an integrated system comprised of an income tax component, which does not depend on the household's number of children, and a child-benefit component, which does not depend on the household's level of income, will be henceforth referred to as a universal system.

We start by examining the properties of the optimal general system. We show that, counter to conventional wisdom, it is desirable to tax children at the margin. That is, the total tax liability should rise with the number of children (for a given level of

[5] For a similar approach in the context of family taxation, see Cigno and Pettini (2002), who assume that market ability and domestic skills are uncorrelated. Assuming that households differ not only in their earning abilities but also in their nurturing capacities would require one to know the magnitude of, and the correlation between, the two types of ability in order to derive concrete policy conclusions. In the absence of substantiated empirical evidence, assuming no correlation seems to be a plausible benchmark for analysis.

income). The mechanism at work is associated with the nature of the quality-quantity trade-off faced by the household. In the absence of taxes, low-skill households are faced with a lower opportunity (time) cost of raising children relative to high-skill ones. Hence, they choose to 'specialize' in quantity (number of children), whereas high-skill households choose to 'specialize' in the quality of children (e.g., education). Therefore, (observed) family size may be employed as an indicator for the (unobserved) earning capacity of the household (a 'tagging' device). This negative correlation between family size and ability provides the rationale behind the conventional wisdom calling for subsidizing children on equity grounds. However, in a system in which child benefits can be made means-tested, the government can employ a more refined notion of correlation between ability and family size; namely, the correlation between these two variables which is conditional on income. For a given level of income, a high-skill household has more leisure than a low-skill one, as it has to work less in order to obtain the same level of income. Hence, conditional on income, a high-skill household has a comparative advantage in raising children over the low-skill household. Thus, conditional on income, the correlation between family size and ability is positive, thereby calling for taxing (rather than subsidizing) children at the margin.

Clearly, this somewhat surprising result hinges on the ability of the government to set child benefits that are means-tested. If the government is restricted to a universal system, the conditional correlation between ability and family size can no longer be of use. The relevant correlation then becomes the unconditional one. With taxes in place, the latter correlation cannot be unambiguously signed; hence, one cannot determine unequivocally whether children should be taxed (or subsidized) at the margin. Nonetheless, we are able to provide some plausible numerical examples, in which subsidizing children at the margin is socially desirable, in sharp contrast to the general (means-tested) case.

Naturally, a universal system can never do better than a general (means-tested) one. In fact, we are able demonstrate the strict dominance of the means-tested system when the skill distribution is discrete (with any arbitrary finite number of skill levels).

The structure of the remainder of the paper will be as follows. In the following section we introduce the analytical framework. In section 3 we formulate the government problem and derive the properties of the general (means-tested) income tax cum child benefit system. In section 4, we compare the general (means-tested) system with the restricted (universal) one. The universal case is discussed in section 5. Section 6 concludes.

2. The Model

Consider an economy with a continuum of households. The number of households is normalized to unity, with no loss in generality. We assume that the production technology employs labor only, and exhibits constant returns to scale and perfect substitution across the various skill levels. Following the standard assumption in the optimal tax literature, we assume that households differ in a single attribute: their earning ability/skill level (equaling the wage rate, assuming a competitive labor market). We let w denote the wage rate and assume that w is distributed over some, possibly unbounded, support $[\underline{w}, \overline{w}]$, with a cumulative distribution function $F(w)$ and corresponding density $f \equiv F'$. We follow Mirrlees (1971) by assuming that abilities (wage rates) are unobserved by the government, thus constraining the latter to second-best re-distributive policies.[6]

All households share the same preferences, represented by the following additively separable utility function:

$$(1)\quad V(c,l,n,e) = c + h(l) + [v(n) + u(e) + \Phi(\alpha)];$$

where c denotes (parental) consumption, n denotes the number of children, e denotes the (per-child) amount of child-specific market goods/services provided by the parents (education, day-care etc.), α denotes the amount of time (per-child) dedicated by parents to child rearing/nurturing activities (parental attention) and l denotes (parental) leisure.[7] We assume that v, u Φ and h are strictly concave and strictly increasing, and further assume the Inada conditions hold, so that interior solutions are guaranteed throughout.

Several remarks are in order. Note first that our setting captures the fundamental quantity-quality trade-off [a la Becker (1960)] faced by the household, whether to increase the quantity (number of children) or invest in their quality (human capital/education and/or parental attention).[8,9] We assume that parents derive utility from their child rearing/nurturing activities. The quasi-linear specification rules out

---

[6] As differences in earning ability are assumed to be the single source of heterogeneity in the economy, we refrain from introducing horizontal equity considerations into the analysis.

[7] Notice that e is measured per-capita, for simplicity; that is, there are no economies of scale embodied in the consumption of children. In reality, some economies of scale are likely to exist and are often addressed by reference to equivalence-scales. Ignoring economies of scale does not affect the qualitative nature of our key arguments. In fact, assuming economies of scale could even strengthen our argument.

[8] See, for example, Moav (2005), for a similar setting.

[9] The variable e is henceforth interpreted as the level of parental provision of child related consumption goods, such as education (Becker, 1991), but may well take additional plausible interpretations, such as the maximized lifetime utility of each child (Becker and Barro, 1988) or old age support expected by parents from each of their children (Cigno, 1993). One possible way to interpret the specific functional form given in equation (1) is to define the quality of (per-child) level of education provided by the parents (E) as depending on both the time inputs contributed by the parents (α) and the amount of market goods or services purchased by the parents denoted by (e), say, private-tutor services. Then, assuming that the quality of education is given by $E = f(e,\alpha)$ and further assuming that the 'production technology', f, is taking an additively separable form, one obtains the functional form given in equation (1).

income effects, and is assumed for tractability purposes [as in Diamond (1998) and Salanie (2003) in the optimal tax literature]. It is worth noting that Becker (1960) conjectured that the elasticity of family size (quantity/number of children) with respect to income would be rather small, which is consistent with some of the empirical evidence [see, e.g., Hotz, Klerman and Willis (1997), and more recently Cohen, Dehejia and Romanov (2007)]. Note, further, that we follow the standard approach in the endogenous fertility literature and assume that the household can deterministically choose the number of children.[10] Finally note, that in our setting there is no difference between the household and the individual, as we adopt a unified approach to household decision making.

Each household is faced with the following budget constraint:

$$(2)\quad \begin{aligned} &c + n \cdot e = z(y,n); \\ &y = w \cdot (1 - l - n \cdot \alpha), \end{aligned}$$

where y and z denote gross and net income levels, respectively. Several remarks are in order. First notice, that we normalize each household's time endowment as well as the price levels of both c and e to unity, with no loss in generality. Notice further that wealthier households find it more costly to raise children, due to the larger opportunity cost they incur (forgoing time in the labor market).[11] Finally, note that we consider a general non linear tax schedule, which depends both on the number of children and on the level of gross income. This tax schedule is implicitly defined by the difference between the gross and net income levels, $t(y,n) \equiv y - z(y,n)$. Note that $t(y, n)$ denotes an integrated income tax and child benefit system. From an economic point of view, this system, referred to as a means-tested system, cannot be

[10] For models assuming exogenous fertility see, for instance, Cremer, Dellis and Pestieau (2003).

[11] Wealthier households will partially mitigate this by replacing their own time inputs with relatively cheaper outsourced day-care services (via choosing to increase e and decrease α). We will further discuss this point below.

decomposed into separate income tax and means-tested child benefit components, except in the special case where $t(y, n)$ takes the form: $t(y,n) = a(y) + b(n)$. The latter is referred to as a universal system, with $a(y)$ denoting an income tax component and $b(n)$ denoting a non means-tested (universal) child benefit system.

The typical household seeks to maximize the utility function in equation (1), subject to the budget constraint in (2). Substituting from the budget constraint in (2) into the utility function in equation (1) to eliminate c and l, we obtain the indirect utility function $U(w)$ given by:

(3) $U(w) = \max_{n,y,e,\alpha} \{[z(y,n) - n \cdot e)] + h(1 - y/w - n \cdot \alpha) + [v(n) + u(e) + \Phi(\alpha)]\}$

The first-order-conditions for the typical w-household's optimal choice are given by:

(4) $v'(n) + z_n(y,n) - \alpha \cdot h'(1 - y/w - n \cdot \alpha) - e = 0$,

(5) $z_y(y,n) - (1/w) \cdot h'(1 - y/w - n \cdot \alpha) = 0$,

(6) $u'(e) - n = 0$,

(7) $\Phi'(\alpha) - n \cdot h'(1 - y/w - n \cdot \alpha) = 0$,

where z_n and $1 - z_y$ denote, respectively, the marginal subsidy provided to an additional child, and, the marginal tax rate levied on labor income.[12]

By differentiating the first-order conditions in (4)-(7) with respect to w, one can show (see appendix A for details) that in the absence of any form of government intervention; namely, when $z \equiv y$, hence, $z_y = 1$ and $z_n = 0$:

(8) $$n'(w) = \frac{\alpha \cdot \Phi''(\alpha) + \Phi'(\alpha)}{[v''(n) - e'(n)] \cdot \Phi''(\alpha) - w^2},$$

[12] We will henceforth assume that the second order conditions are always satisfied, thus employ first-order conditions only to characterize the individual incentive constraints when formulating the government problem. This latter assumption will ensure no 'bunching' in the optimal solution of the government problem [see Ebert (1992), for a rigorous treatment of 'bunching' in the context of optimal non-linear labor income tax in the continuum case; notice that in the two-type case bunching (that is, pooling) will never be part of the optimal solution as shown by Stiglitz (1982)].

and

(9) $e'(w) = n'(w)/u''(e)$.

By virtue of the second-order condition for the household's optimization problem, the denominator of the expression on the right-hand side of equation (8) is positive. It thus follows that $n'(w)<0$ [and hence $e'(w)>0$, by virtue of (9) and the concavity of u] if-and-only-if: $-\alpha \cdot \Phi''(\alpha)/\Phi'(\alpha) > 1$. Thus, when the degree of concavity of Φ, as measured by the familiar coefficient of relative-risk aversion, is sufficiently large, the model yields the plausible result suggested by the quantity-quality paradigm: poor families will 'specialize' in quantity and hence choose to have a larger number of (less educated) children. The opposite will hold true for wealthy families: they will 'specialize' in quality (educating their offspring). We will henceforth make the following assumption:

(A) $-\alpha \cdot \Phi''(\alpha)/\Phi'(\alpha) > 1$.

This 'calibrating' assumption will later play a crucial role in the design of the welfare system.

In order to motivate assumption (A), suppose first that the amount of time devoted by parents to nurturing activities, α, is a fixed parameter, rather than an endogenous choice variable optimally set by the household. In this case, one can show that $n'(w)<0$ and $e'(w)>0$, unambiguously.[13] In such a case, as all households invest the same amount of time per child, the opportunity cost of raising children would be higher for households with higher earning ability, yielding hence, unambiguously, the expected patterns of specialization. In the more general model employed here, the household can, however, adjust the time dedicated to nurturing activities. Fixing the number of children, n, and fully differentiating the expression in (7) with respect to w

[13] See Blumkin et al (2010) for details.

(assuming no taxes in place, as before), it follows that $\alpha'(w) = n / \Phi''(\alpha) < 0$ because Φ is concave. Thus, conditional on the number of children, poor (low-skill) households will tend to invest more time in their children than will wealthy (high-skill) households (the latter will resort to outsourcing services such as tutoring and day-care, due to the relatively high opportunity cost incurred). The opportunity cost of raising children given by $\alpha(w) \cdot w$ may either rise or fall in *w*. Here comes into play our 'concavity' assumption [given in (A)] which essentially implies that the adjustment effect [$\alpha'(w) < 0$] would be relatively small, and hence, the patterns of specialization would remain as in the case where α is a fixed parameter.

We next turn to characterize the properties of the integrated income tax cum child benefit system.

3. The General (Means-Tested) System

The government seeks to maximize an egalitarian social welfare function given by:

(10) $$W = \int_{\underline{w}}^{\overline{w}} G[U(w)]dF(w);$$

where *G* is strictly increasing and strictly concave,[14] by choosing the tax schedule, *t*(*y*,*n*), subject to a revenue constraint:

(11) $$\int_{\underline{w}}^{\overline{w}} t[y(w), n(w)]dF(w) = R;$$

[14] In the formulation of the welfare function in (10), we take *U*(*w*) as the argument; namely, the utility driven by the parent. This utility includes an altruistic component derived from providing consumption to the offspring [a type of altruism a la Barro (1974) rather than joy-of-giving as in Andreoni (1990)]. One could also include the utility derived by the offspring per-se in the welfare calculus in addition to that of the altruistic parent. This type of double counting would create a positive externality, justifying the subsidization of children. However, as this paper focuses on the re-distributive motive for taxing/subsidizing children, we set aside this alternative motive, without discounting its importance, by 'laundering out' the child utility component.

where $y(w)$ and $n(w)$ are the optimal individual choices of the gross income level and number of children, respectively, given by the first-order-conditions in (4)-(7); and R denotes the (pre-determined) level of government revenue needs. Notice that we start by analyzing the most general (means-tested) setting in which taxes/benefits may vary across income levels as well as family size. Below, we also consider a universal system (as is often the case in many countries) in which the tax function takes an additively separable form: $t(y,n) = a(y) + b(n)$.[15]

Following Mirrlees (1971) and (1976) and Salanie (2003), we reformulate the government optimization problem (see appendix B for details) as choosing the functions $U(w), n(w)$ and $y(w)$, so as to maximize the social welfare function in equation (10), subject to the revenue constraint:

(12)

$$\int_{\underline{w}}^{\bar{w}} \begin{bmatrix} y(w) - U(w) - n(w) \cdot e[n(w)] + h\big[1 - y(w)/w - \alpha[n(w), y(w)] \cdot n(w)\big] + v[n(w)] \\ + u\big[e[n(w)]\big] + \Phi\big[\alpha[n(w), y(w)]\big] \end{bmatrix} dF(w) = R,$$

and the incentive compatibility constraint:

(13) $$U'(w) = h'\big[1 - y(w)/w - n(w) \cdot \alpha[n(w), y(w)]\big] \cdot y(w)/w^2, \text{ for all } w,$$

where $e[n(w)]$ and $\alpha[n(w), y(w)]$ are implicitly defined by the first-order conditions in (6)-(7).

[15] It is implicitly assumed that the government cannot observe the household's expenditure on education, so the latter cannot be subsidized or taxed. Our results will not change if we allow for such taxation, as long as we plausibly assume that *e* is a form of anonymous transaction and hence cannot be observed on the individual level. It implies that only linear taxation of *e* can be used. In such a case, the price of *e* would be given by 1+*t*, where *t* denotes the unit tax on *e*. Such linear taxation of *e* will not affect the redistributive role of non-linear and potentially means-tested child benefits. For example, if means-tested benefits are allowed for, then conditional on income, high-skill families have a larger time endowment. As all households are still faced with the same (after-tax) prices of *e*, high-skill families will find it relatively cheaper to raise children and will choose to specialize in quantity, as in the no-tax case. Hence, taxing children at the margin (as shown below to be the optimal solution for the no-tax case) would still be warranted. For incorporating taxation of child-specific commodities in an optimal tax setting with endogenous fertility, see Cigno (2009).

It is useful to point out that we do not directly derive the integrated net-income function *z*(*y*, *n*). We rather derive the optimal functions *y*(*w*), *n*(*w*), *e*(*w*), $\alpha(w)$ and *U*(*w*); and then calculate $z[y(w), n(w)]$, the net income, employing condition (3):

$$z[y(w), n(w)] \equiv U(w) + n(w) \cdot e(w) - h[1 - y(w)/w - \alpha(w) \cdot n(w)] - v[n(w)] - u[e(w)] - \Phi[\alpha(w)]. \tag{14}$$

Note that in this way, we can only define the net income function, *z*, and the tax function, $t \equiv y - z$, at bundles [*y*(*w*), *n*(*w*), *e*(*w*), $\alpha(w)$ and *U*(*w*)] that are actually chosen by individuals in the optimal solution. Thus, *z*(*y*, *n*) is not well defined elsewhere. Therefore, strictly speaking, one cannot directly derive the marginal income tax rate, $1 - z_y$, and the marginal child benefit, z_n. Instead, as is common in the literature, we define these rates through the relevant individual marginal rates of substitution. Indeed, z_n is defined as the marginal rate of substitution of net income (*c*+*ne*) for children (*n*) and z_y is defined as the marginal rate of substitution of net income for gross income (*y*). Formally, using the individual first-order conditions in (4) and (5), yields:

$$z_n = -v' + \alpha \cdot h' + e \tag{15}$$

and

$$z_y = h'/w. \tag{16}$$

We next turn to solve the optimization program as an optimal control problem employing *Pontryagin's* maximum principle. We choose *n*(*w*) and *y*(*w*) as the two control variables and *U*(*w*) as the state variable. Formulating the *Hamiltonian* then yields:

$$(17) \quad H = \left[G(U) + \lambda \cdot \left[\begin{array}{l} y - U - n \cdot e(n) + h[1 - y/w - n \cdot \alpha(n,y)] + v(n) \\ + u[e(n)] + \Phi[\alpha(n,y)] - R \end{array} \right] \right] \cdot f$$

$$+ \mu \cdot \left[h'[1 - y/w - n \cdot \alpha(n,y)] \cdot y/w^2 \right],$$

where $\mu(w)$ denotes the co-state multiplier, λ is the multiplier associated with the government revenue constraint and $e(n)$ and $\alpha(n,y)$ are given, respectively, by the implicit solutions to the first-order conditions in (6)-(7).

Formulating the necessary first-order conditions, employing the *transversality* conditions and following some re-arrangements (see appendix C for details) yields the following expression for the optimal marginal child benefit:

$$(18) \quad z_n = \left[1 - \frac{D(w)}{D(\underline{w})} \right] \cdot \frac{1 - F(w)}{f(w)} \cdot h''\left[1 - y/w - n \cdot \alpha\right] \cdot \frac{y}{w^2} \left[\alpha + n \cdot \frac{\partial \alpha}{\partial n} \right],$$

where $D(w) = \frac{1}{1 - F(w)} \int_{w}^{\bar{w}} G'[U(t)] dF(t).$

Strikingly, one can show (see appendix D for details) that the optimal condition in (18) implies, unambiguously, that $z_n < 0$. That is, children should be taxed at the margin. Formally,

Proposition 1: In the optimal integrated tax/benefit system, total tax liability rises with the number of children (for a given level of pre-tax income).

We obtain a fairly strong result. In a system of child allowance, many may advocate reducing the allowance for each additional child on the grounds of economies of scale in child rearing.[16] Formally, in our setting this would imply that the allowance per additional child; namely, z_n, would decline with *n*, that is $z_{nn} < 0$. Proposition 1 suggests that z_n itself (not z_{nn}) should be negative. Moreover, suppose that statutorily, the tax/benefit system is separated into an income tax component,

[16] This is essentially the rationale underlying the common use of equivalence scales.

$a(y)$, and a means-tested per-child allowance, $k(y,n)$. That is, $t(y,n) = a(y) - k(y,n) \cdot n$. The standard argument of economies of scale in child rearing calls for the average child allowance, *k*, to decline with the number of children, *n*. The proposition is in fact stronger, as it calls for total child allowance, kn, to decline with *n*. This implies that *k* must decline at a faster rate than the rise in *n*. That is, the elasticity of the per-child allowance, *k*, with respect to the number of children, *n*, is higher than one (in absolute value). We emphasize that we obtain this result even though there are no economies of scale in child rearing in our setting.

The rationale for this result is as follows.[17] In the absence of taxes, low-skill households are faced with a lower opportunity (time) cost of raising children relative to high-skill ones. Hence, they choose to 'specialize' in quantity (number of children), whereas high-skill households choose to 'specialize' in quality (e.g., education). In a second best setting, (observed) family size may be employed as an indicator of the (unobserved) earning capacity of the household (a 'tagging' device).[18] The negative correlation between family size and ability provides the rationale behind the conventional wisdom calling for subsidizing children on equity grounds. However, in a system in which child benefits can be made means-tested, the government employs a more refined concept of correlation between ability and family size for 'tagging' purposes; namely, the correlation between these two variables, which <u>is conditional on income</u>. To see this, note that for a given level of income, a high-skill household has more leisure than a low-skill one, as it has to work less in order to obtain the same level of income. Hence, conditional on income, a high-skill household has a

[17] For a graphical illustration of result given in proposition 1, calling for taxing children at the margin, see Blumkin et al (2010).
[18] Note that conditioning transfers on family size serves as a *second-best* 'tagging' device because fertility is an endogenous variable in our setting, which responds to financial incentives offered by the government [for recent empirical attempts to estimate the effect of financial incentives on fertility, see Cohen, Dehejia and Romanov (2007) and Laroque and Salanie (2008)].

comparative advantage in raising children over the low-skill household. Thus, conditional on income, the correlation between family size and ability is positive (and not negative as conventional wisdom suggests). In light of the positive correlation between family size and ability (conditional on income), taxing (rather than subsidizing) children at the margin would be socially desirable.[19] [20]

Turning back to the optimal condition for the marginal tax in (18), it is straightforward to verify that the standard zero tax at the top [Phelps (1973) and Sadka (1976)], when the distribution of skills is bounded from above, and at the bottom [Seade (1977)] apply. Notice also that our prediction differs from the classic result of the redundancy of commodity taxation [Atkinson and Stiglitz (1976)]. This derives from the fact that, even though utility is separable between leisure and the set of consumption goods, households are faced with different costs of raising children (due to the fact that raising children is time-consuming and households differ in their earning skills).[21]

4. Means-Tested versus Universal Systems

Naturally, a general (means-tested) tax/benefit system cannot do worse than a universal one, which is confined to a separable form (between an income tax and child benefit components). An interesting question is whether a universal system

[19] It is important to emphasize that in equilibrium, high ability households will choose to spend more hours in the labor market and raise a lower number of children, relative to low-ability households. However, our argument suggests that if they mimic the low ability households (an out-of-equilibrium strategy which will not be incentive compatible by construction of our optimal policy rule), then by choosing the same level of income, they will find it relatively cheaper to raise children.

[20] Cigno and Pettini (2002) demonstrate the 'tagging' role played by child benefits in the context of a linear tax system, showing that in the presence of a full range of policy instruments, a tax on the number of children may well be socially justifiable, even when the wage-fertility correlation is negative. In Cigno and Pettini (2002) the 'tagging' argument implies that children should be a tax asset for low-wage households and a tax liability for high-wage ones (taking into account the overall effect of child benefits and the taxes levied on child related consumption). This may still reconcile with taxing the number of children if child related consumption is sufficiently subsidized.

[21] This observation was first made by Cigno (1986).

(which is fairly prevalent in many OECD countries) can nevertheless suffice in certain conditions to attain the social optimum. We are able to show that this is never the case when the distribution of skill levels is discrete. Formally, we state and prove the following proposition:

Proposition 2: When the distribution of skills is discrete, any universal system can be replaced by a means-tested system that attains a higher level of social welfare.

Proof: See appendix E.

5. The Universal Case

In section 3 we have demonstrated, counter to conventional wisdom, that taxing children at the margin would be socially desirable for re-distributive purposes, when child benefits are allowed to be means-tested. However, in many countries (in fact, in most developed countries,) benefits are offered on a universal basis and are not subject to means testing. That is, the net income/benefit schedule essentially takes an additively separable form: $z(y,n)=a(y)+b(n)$. Therefore it is of interest and policy relevance to see under what conditions, a universal system can justify subsidizing children at the margin. We attempt to address the following question: starting from any given income tax system, under what conditions will a universal system of child allowances with marginal subsidies be desirable?[22]

To address this issue we must first re-formulate the government optimization program. In this case $y(w)$ is no longer a control variable, but is rather implicitly defined [along with $\alpha(w)$] by the first-order conditions of the household's utility maximization problem (as a function of family size, n):

(5') $$a_y(y)-(1/w)\cdot h'(1-y/w-n\cdot\alpha)=0,$$

[22] We will examine, in particular, the desirability of providing a marginal child subsidy, when the labor income tax is set at the optimum.

(7') $\Phi'(\alpha) - n \cdot h'(1 - y/w - n \cdot \alpha) = 0.$

The government then chooses *U*(*w*) and *n*(*w*) so as to maximize the social welfare function given by (10), subject to the revenue constraint:

$$(12') \int_{\underline{w}}^{\bar{w}} \begin{bmatrix} y[n(w)] - U(w) - n(w) \cdot e[n(w)] + h[1 - y[n(w)]/w - \alpha[n(w)] \cdot n(w)] \\ + v[n(w)] + u\big[e[n(w)]\big] + \Phi\big[\alpha[n(w)]\big] \end{bmatrix} dF(w) = R,$$

and the incentive-compatibility constraint:

(13') $U'(w) = h'\big[1 - y[n(w)]/w - n(w) \cdot \alpha[n(w)]\big] \cdot y[n(w)]/w^2,$ for all *w*,

where $y[n(w)]$, $\alpha[n(w)]$ and $e[n(w)]$ are implicitly defined by the first-order conditions in (5'), (6) and (7').

The *Hamiltonian* in this case becomes:

(17')

$$H = \left[G(U) + \lambda \cdot \begin{bmatrix} y(n) - U - n \cdot e(n) + h[1 - y(n)/w - n \cdot \alpha(n)] + v(n) \\ + u[e(n)] + \Phi[\alpha(n)] - R \end{bmatrix} \right] \cdot f$$
$$+ \mu \cdot \big[h'[1 - y(n)/w - n \cdot \alpha(n)] \cdot y(n)/w^2\big],$$

where $\mu(w)$ denotes the co-state multiplier and λ is the multiplier associated with the government revenue constraint.

Formulating the necessary first-order conditions, employing the *transversality* conditions, the household first order conditions and following some re-arrangements (see appendix F for details) yields the following expression for the optimal marginal child subsidy (where some of the arguments of the functions are omitted to abbreviate notation):

$$(19) \quad b_n = (1 - a_y) \cdot y' + \left[1 - \frac{D(w)}{D(\underline{w})}\right] \cdot \frac{1 - F(w)}{f(w) \cdot w} \cdot \big[h'' \cdot (y'/w + \alpha + n \cdot \alpha') \cdot y/w - a_y \cdot y'\big].$$

Notice, that as in the general (means-tested) case, one cannot directly derive the marginal child benefit, b_n. Instead, we define, as before, the marginal child benefit as the marginal rate of substitution of net income (*c*+*ne*) for children (*n*). Formally, using the individual first-order condition in (4), yields:

(15') $b_n = -v' + \alpha \cdot h' + e.$

Fully differentiating the first-order conditions in (5') and (7') with respect to *n*, applying *Cramer's Rule* and re-arranging, yields:

(20) $$y'(n) = \frac{-h''/w \cdot (\alpha \cdot \Phi'' + \Phi')}{(h''/w^2 + a_{yy}) \cdot (n^2 \cdot h'' + \Phi'') - (n \cdot h''/w)^2},$$

where the denominator of the term on right-hand side of (20) is positive by the second order condition for the household optimization.

Thus, by virtue of the concavity of *h* and our parametric assumption ($\alpha \cdot \Phi'' + \Phi' < 0$), it follows that $y'(n) < 0$. That is, an increase in the number of children (say in response to offering a marginal subsidy) results in a reduction in the labor supply, and consequently the gross level of income, as expected.

Fully differentiating the first-order condition in (5') with respect to *n* yields:

(21) $a_{yy} \cdot y' + h''/w \cdot (y'/w + \alpha + n \cdot \alpha') = 0.$

Substituting from (21) into equation (19) and re-arranging yields:

(22) $$b_n = \left[(1 - a_y) + \left[1 - \frac{D(w)}{D(\underline{w})} \right] \cdot \frac{1 - F(w)}{f(w) \cdot w} \cdot \left[-a_{yy} \cdot y - a_y \right] \right] \cdot y'.$$

Assuming that the second-order conditions for the government program are satisfied, a necessary and sufficient condition for the desirability of subsidizing children at the margin, $b_n > 0$, is:

(23) $$\left. \frac{\partial H}{\partial n} \right|_{b_n = 0} > 0 \Leftrightarrow \left[(1 - a_y) + \left[1 - \frac{D(w)}{D(\underline{w})} \right] \cdot \frac{1 - F(w)}{f(w) \cdot w} \cdot \left[-a_{yy} \cdot y - a_y \right] \right] \cdot y' > 0.$$

Namely, starting from a system where the marginal child subsidy is set to zero, social welfare will rise by introducing a small marginal child subsidy (thereby increasing the number of children).

By virtue of the fact that y'<0, the condition in (23) holds if-and-only-if:

$$(24) \qquad \left[1-\frac{D(w)}{D(\underline{w})}\right]\cdot\left[\frac{1-F(w)}{f(w)\cdot w}\right]\cdot\left[a_{yy}\cdot y+a_{y}\right]>(1-a_{y}).$$

The two first terms in brackets on the left-hand-side of equation (24) are positive [recall that $D(w)$ is decreasing with respect to w]; hence, they work in the direction of providing a marginal child subsidy. The sign of the third term in brackets on the left-hand-side of (24) is however ambiguous. Therefore, the sign of the whole expression on left-hand side of condition (24) is ambiguous too. One can show (see appendix G) that when the marginal child subsidy is set to zero (b_n =0), the third term in brackets [hence, the left-hand side of condition (24)] has the opposite sign of $n'(w)$, which reflects the correlation between earning ability and family size. The term on the right-hand-side is the marginal income tax rate, which is exogenously given in our formulation. It is plausibly assumed that this term is positive, as our model focuses on the intensive margin of individual labor supply choice; hence, it works in the direction of levying a marginal tax on children. Thus, one cannot a-priori determine the sign of b_n. Naturally, and as is also evident from condition (24), determining whether providing a marginal child subsidy would be socially desirable or not depends on the properties of the income tax schedule.

To gain some intuition, we consider several special cases. Consider first the simple case in which the marginal tax rate is zero for all levels of income (that is, either there is no tax in place, or, a lump-sum tax is being levied). In such a case, $1-a_y=0$ and $a_{yy}=0$. It follows then that $n'(w)<0$ and the term on the left-hand side of condition (24) is positive. Because the term on the right-hand side of condition (24) vanishes, it follows that providing a marginal child subsidy would be unambiguously socially desirable. The rationale for the clear-cut result obtained for this special case is as follows. In the absence of taxes, low-skill households will have a comparative advantage in raising children, and will hence choose to raise more children than high-skill ones [namely, $n'(w) < 0$]. The negative correlation between earning ability and family size in this case can be employed by the government for re-distributive purposes. Subsidizing children at the margin allows the government to target benefits to low-ability (poor) households, thereby to enhance re-distribution.

We turn next to the case where a flat income tax is in place; namely, $1 - a_y > 0$ and $a_{yy} = 0$. As can be observed from condition (24), both the left-hand side term and the right hand side term are unambiguously positive. Thus, one cannot determine a-priori whether a marginal child subsidy would be desirable. Similar to the case where no tax is in place, the positive sign of the term on the left-hand side derives from the fact that with a flat tax in place, low-skill families still choose to 'specialize' in quantity (namely, $n'(w) < 0$); hence, the government can still employ the ensuing negative correlation between ability and family size for re-distributive purposes by subsidizing children at the margin. However, unlike the case where the marginal income tax rate is zero, the desirability of a marginal subsidy is not forgone conclusion, as the sign of the term on the right-hand side is also positive. This term, which is equal to the marginal income tax rate, reflects the cost associated with a fiscal crowding out effect due to the interaction between the income tax and the child benefit instruments. A child subsidy will induce households to give birth to more children and hence to spend less hours in the labor market. This will reduce the government revenues collected from the income tax system and hence, indirectly, the level of re-distribution.

Obviously, when the marginal income tax rate is zero, that is $1 - a_y = 0$, this term disappears (there is no crowding out effect). In general, this term will work in the direction of levying a tax on children. Thus, although the negative correlation between ability and family size is maintained under a flat (linear) income tax system, one cannot determine a-priori whether a marginal subsidy is desirable or not.

In the two cases examined above the marginal income tax rate is constant across different levels of income. Hence, the term on the left-hand side of (24), which captures the welfare gain from 'tagging', was unambiguously positive. Clearly, this need not be the case with a non-linear income tax system in place. To see this, consider the case where the marginal income tax rate rises with respect to income, that is $a_{yy} < 0$. When the marginal income tax rate rises sufficiently rapidly (that is, a_{yy} is sufficiently negative), then the third term in brackets on the left-hand-side of condition (24), and with it the entire expression on the left-hand-side of this condition, become negative. In such a case, the expression on the left-hand side of (24) will work, all-in-all, in the direction of levying a marginal tax on children. The rationale for this result is as follows. In general, we expect high-ability households to choose a higher level of gross labor income than that chosen by low-ability households. Thus, high-ability households face a higher marginal income tax rate than that faced by low-ability households. When the marginal tax rate will rise sufficiently rapidly, the net-of-tax wage rate of high-ability households may fall below that of low-ability households. When the net-of-tax wage rate of high-ability households will be sufficiently smaller than that of low-ability ones, the patterns of comparative advantage of child-rearing will reverse, and high-ability households will choose to raise more children than low-ability ones (namely, $n'(w) > 0$). The ensuing positive correlation between ability and family size implies that a marginal child tax (rather than a subsidy) would be desirable.

Naturally, in the case where the marginal income tax rate diminishes with respect to income (namely, $a_{yy} > 0$), the net-of-tax wage rate of high-ability households is higher than that of low-ability households, with the difference becoming even larger than in the flat-tax case. Hence, the negative correlation between earning ability and family size becomes yet stronger. The term on the left-hand side of condition (24) is definitely positive, and hence calls for subsidizing children at the margin as a 'tagging' device. If this effect is stronger than the crowding out effect reflected by the positive term on the right-hand side of condition (24), then a marginal child subsidy is desirable.[23]

To sum up, we have demonstrated that the desirability of subsidizing children at the margin under a universal child allowance system is far from being forgone conclusion and is highly sensitive to the properties of the income tax schedule. Notably, a necessary condition for the desirability of subsidizing children (at the margin) is that the tax-and-transfer system does not exhibit increasing (effective) marginal tax rates (marginal-rate progressivity).

We turn next to examine whether the condition in (24) for the desirability of a marginal child subsidy can hold under reasonable parametric assumptions. Saez (2002) approximates the US tax system by a flat tax schedule with a constant marginal tax rate of 40 percent. That is, we set $a_y = 0.6$ and $a_{yy} = 0$. Following Diamond (1998), we assume a single peaked density of skills (a property satisfied by commonly used distributions like the log-normal distribution), which is approximated by a Pareto

[23] One may argue that, in practice, most income tax systems exhibit marginal-tax rate progressivity; namely, the statutory marginal tax rate is rising with (gross) income. However, when the bulk of welfare (transfer) programs are means-tested, the effective marginal tax rate at low levels of gross income is relatively high. Therefore, the integrated tax-transfer system exhibits marginal-tax regressivity at the lower end of the income distribution; namely, the effective marginal tax rate is decreasing with (gross) income.

distribution above the modal skill level. Thus, the term $\frac{1-F(w)}{f(w)\cdot w}$ initially decreases up to the modal skill level and is then constant. It follows that the term $\frac{1-F(w)}{f(w)\cdot w}$ is bounded from below, where the lower bound is given by one over the coefficient of the Pareto distribution. Following Finberg and Poterba (1993), we assume a Pareto coefficient in the range 0.5 to 1.5, which implies that that the lower-bound of the term $\frac{1-F(w)}{f(w)\cdot w}$ varies in the range of 2/3 to 2. Assuming a *Rawlsian* social welfare function implies that $D(w)$=0. Substituting the parametric values into the condition in (24) implies that a marginal subsidy is desirable over the entire range of productivities (wage rates).

Obviously, with a higher marginal tax rate than the one used above, the desirability of providing a marginal subsidy across-the-board (or even providing a subsidy at all) may fail to hold. However, with a flat tax in place and a *Rawlsian* government, our parametric assumptions about the skill distribution imply that there would be a cutoff level of productivity, below which a marginal child subsidy would be provided and above which a marginal child tax would be imposed [this follows from the fact that the term $\frac{1-F(w)}{f(w)\cdot w}$ on the left-hand side of (24) is (weakly) decreasing with respect w]. Thus, while in general the desirability of providing a marginal subsidy is ambiguous, it is more likely that such a subsidy would be given to low-skill households.

So far we have examined the desirability of subsidizing/taxing children at the margin, taking the income tax system as given. We turn now to address the same question, while assuming, instead, that an optimal income tax system is in place.

Assuming that the marginal child subsidy is set to zero ($b_n = 0$), one can show (see appendix H for details) that the formula for the optimal income tax rule is given by:

$$(25) \qquad \frac{1-a_y}{a_y} = \left[1 - \frac{D(w)}{D(\underline{w})}\right] \cdot \frac{1-F(w)}{f(w)\cdot w} \cdot \left[1 + \frac{1}{\varepsilon_L}\right],$$

where ε_L denotes the labor supply elasticity, given by:

$$\varepsilon_L = -\frac{w \cdot a_y}{h'' \cdot y} \cdot \frac{1}{(1/w + n' \cdot \alpha + n \cdot \alpha')}.$$

The formula given in (25) is similar to the standard optimal tax formulae available in the literature [see, e.g., Salanie (2003)]. However, there is a subtle difference. In our setting the household's shift in labor supply in response to a change in the after-tax wage rate is accompanied by adjustments in family size and the allocation of time to nurturing activities (as opposed to the single labor-leisure margin which appears in the standard setting). The elasticity given in the expression in (25) takes these adjustments into account. Most estimates of the intensive margin elasticity (hours of work conditional on participating in the labor market) are fairly small [see, e.g., the survey by Blundell and MaCurdy (1999)]. However, the notion of elasticity used in our setting captures also traditional 'extensive margin' considerations related to the family composition. Parents will respond to higher anticipated net wage rates by choosing to give birth to a smaller number of children (focusing on their career, instead) and replacing their own (parental-attention) time with paid day-care services. These, in turn, enable the parents to put yet more efforts in their jobs. Thus, the commonly measured intensive margin elasticity underestimates the notion of elasticity used in our setting.

Dividing both sides of the condition given in (24) by a_y, then substituting for the term $(1-a_y)/a_y$ from (25) into (24) and re-arranging, yields the following necessary and sufficient condition for the desirability of subsidizing children at the margin:

$$(26) \qquad \frac{a_{yy} \cdot y}{a_y} > \frac{1}{\varepsilon_L}.$$

We turn next to examine whether the condition in (26) can hold under reasonable parametric assumptions. We assume, as before, a *Rawlsian* government. We further maintain the assumption about a single-peaked skill distribution approximated by a *Pareto* distribution above the modal skill level. Finally, as is common in the literature, we assume constant labor supply elasticity. By virtue of our parametric assumptions, it follows from condition (25) that the optimal marginal tax rate is constant for income levels chosen by households at the higher end of the skill distribution ($a_{yy} = 0$). Substituting into (26) implies that for these households, unambiguously, children should be taxed at the margin!

Turning next to households at the lower end of the distribution, it follows from (25), by virtue of the single-peaked skill-distribution, that the optimal marginal tax rate is declining with income ($a_{yy} > 0$). Thus, both the right-hand side and the left-hand side terms in (26) are positive. The ambiguous result is again due to the already alluded to trade-off between the 'tagging' component on the left-hand side of (26) and the fiscal crowding out effect (on the right-hand side). With rapidly declining marginal tax rates and relatively high labor supply elasticity, the 'tagging' component (on the left hand side) will prevail and will all in all, call for subsidizing children at the margin.[24]

[24] Notice, that over the income-range where the tax rate is optimally set to be flat, the marginal child subsidy is unambiguously negative (that is, a marginal tax). This is in contrast to the general case, where, with an arbitrary flat-rate system in place, we have shown the result to be ambiguous.

With more general welfare functions, the optimal marginal tax rate may well rise with income over some range. Indeed, the literature [see Diamond (1998) and Salanie (2003), amongst others] suggests that the optimal marginal tax schedule is likely to be U-shaped. It then follows from the conditions in (25) and (26) that over the range in which the marginal tax rate increases with income, children should be, unambiguously, taxed at the margin.

6. Conclusion

The economic literature, starting with the seminal contributions of Becker (1960) and Becker and Lewis (1973), viewed household family planning as an economic decision, where the household chooses the number of children to raise and the bundle of goods to consume, so as to maximize its utility. Thus, the size of the household is optimally determined by comparing the costs and benefits associated with raising children.

The literature has emphasized a fundamental trade-off between the quantity (of children) and their quality (e.g., parental investment in education and commodities consumed by the children). Under plausible assumptions (supported by empirical evidence), comparative advantage considerations would induce low-skill (poor) households to specialize in quantity, whereas high-skill (wealthy) households would choose to specialize in quality. Conventional wisdom therefore suggests that in a second-best setting, the (observed) family size could be used as a screening ('tagging') device for re-distributive purposes by an egalitarian government, and call for subsidizing children. Recent literature [see Cigno (2009) for an elaborate overview] has called this conventional wisdom into question.

In this paper we adopt a benchmark setting in which the quantity-quality paradigm holds; namely, there exists a negative correlation between family size and earning ability. We show, however, that contrary to conventional wisdom, the desirability of subsidizing children in this case is far from being forgone conclusion and crucially hinges on whether child benefits are provided on a universal or a means-tested basis. In fact, we demonstrate that when means-testing is allowed, it is optimal to tax children at the margin (namely, setting the total child benefits to decline with the number of children), rather than subsidizing them. Under a universal (non means-tested) child-allowance system, subsidizing children may be warranted when the tax-and-transfer system is not marginal-tax progressive (namely, when the effective-marginal tax of the integrated tax-and-transfer system does not increase with gross income).

Our results were derived under the plausible assumption that the quantity-quality paradigm holds, implying a negative correlation between observed family size and unobserved innate ability. Alternatively, assuming that the patterns of specialization are reversed; namely, that there is a positive correlation between family size and ability, will result in reversing our predictions, suggesting that under means-testing, children should be subsidized at the margin (and not taxed as one would predict in light of the positive correlation between family size and ability); whereas, under a universal system taxing children may be desirable. The surprising desirability of taxing children is, therefore, not a byproduct of our (calibrating) parametric assumptions, but is rather derived from the sharp difference between the means-tested system and the universal one, yielding opposite policy implications.

Appendix A: Derivations of Equations (8) and (9)

Substituting for $h'(\cdot)$ from (5) into (4) and (7) and setting $z_n = 0$ and $z_y = 1$ yields:

(A1) $v'(n) - \alpha \cdot w - e(n) = 0$,

(A2) $\Phi'(\alpha) - n \cdot w = 0$,

where $e(n)$ is implicitly defined by the household's first-order condition in (6):

(A3) $u'(e) - n = 0$.

The system of two equations [(A1) and (A2)] implicitly define the optimal solution for the number of children and the level of parental time invested per child, as a function of the wage rate [$n(w)$ and $\alpha(w)$]. Fully differentiating the two first-order conditions in (A1) and (A2) with respect to w yields:

(A4) $v''(n) \cdot n'(w) - [\alpha + w \cdot \alpha'(w)] - e'(n) \cdot n'(w) = 0$,

(A5) $\Phi''(\alpha) \cdot \alpha'(w) - [n + w \cdot n'(w)] = 0.$

Applying *Cramer's Rule*, one then obtains:

(A6) $$n'(w) = \frac{\alpha \cdot \Phi''(\alpha) + n \cdot w}{[v''(n) - e'(n)] \cdot \Phi''(\alpha) - w^2},$$

where $[v''(n) - e'(n)] \cdot \Phi''(\alpha) - w^2 > 0$, by the household's optimization second order conditions.

Substituting for the term $n \cdot w$ from (A2) into the numerator on the right-hand side of (A6) yields the condition in (8).

Fully differentiating the expression in (A3) with respect to w and re-arranging yield the condition in (9).

Appendix B: Reformulation of the Government Optimization Problem

In this appendix we derive the expressions for the revenue constraint [equation (12)] and the incentive constraint [equation (13)] that appear in the re-formulated government problem in the main body of the text.

Following the standard approach in the optimal tax literature, the government is essentially offering the individuals with a tax schedule given by the triplet of functions, $z(w), y(w)$ and $n(w)$, that is, bundles comprised of the net-income, gross income and the number of children, from which each household is self-selecting the optimal bundle. The total tax revenues collected given the optimal choices taken by the individuals satisfy the government (pre-determined) revenue needs.

We first turn to derive the incentive constraint given by equation (13). We let $J[w,z,y,n]$ denote the maximal level of utility derived by a household with ability level *w*, net income level z, gross income level *y* and number of children *n*. Formally,

(B1) $$J[w,z,y,n] = \max_{e,\alpha}\left[z - n\cdot e + v(n) + u(e) + h(1 - y/w - n\cdot\alpha) + \Phi(\alpha)\right].$$

By definition of the indirect utility function in (3) it follows that:

(B2) $$U(w) = \max_{z,y,n} J[w,z,y,n)].$$

Denoting by $z(w)$, $y(w)$ and $n(w)$, the optimal choices of the *w*-household, it follows that:

(B3) $$U(w) \equiv J[w,z(w),y(w),n(w)].$$

Fully differentiating the identity in (B3) with respect to *w* yields:

(B4) $$\frac{dU(w)}{dw} = \frac{\partial J}{\partial w} + \frac{\partial J}{\partial z}\cdot z'(w) + \frac{\partial J}{\partial y}\cdot y'(w) + \frac{\partial J}{\partial n}\cdot n'(w).$$

The *w*-household's incentive constraint is defined by the following condition:

(B5) $$w = \arg\max_{w'} J[w,z(w'),y(w'),n(w')].$$

In words, the *w*-household has no incentives to mimic other types; hence it will choose to reveal its true type (rather than to pretend to be some other type, *w'*). Re-formulating the condition in (B5) as a first-order condition yields:

(B6) $$\frac{\partial J}{\partial z}\cdot z'(w)+\frac{\partial J}{\partial y}\cdot y'(w)+\frac{\partial J}{\partial n}\cdot n'(w)=0\,.$$

Substituting from (B6) into (B4) then yields:

(B7) $$U'(w)=\frac{\partial J\left[w,z(w),y(w),n(w)\right]}{\partial w}\,.$$

Thus, the condition in (B6) holds if-and-only-if the condition in (B7) holds. Employing (B1) yields then the following expression, which is identical to equation (13):

(B8) $$U'(w)=h'\left[1-y(w)/w-n(w)\cdot\alpha[n(w),y(w)]\right]\cdot y(w)/w^2.$$

We turn next to the government revenue constraint, given by equation (12). For convenience, we modify the tax schedule offered by the government [given by the triplet of functions, $z(w), y(w)$ and $n(w)$] by replacing the function $z(w)$, the net income level, with the function $U(w)$, the indirect utility level in the individual optimization.[25] Recalling that the tax function is implicitly defined by the difference between the gross income and the net income, $t(y,n)\equiv y-z(y,n)$, the revenue constraint in (11) can be re-written as follows:

(B9) $$\int_{\underline{w}}^{\overline{w}}\left[y(w)-z[y(w),n(w)]\right]dF(w)=R\,.$$

[25] Notice, that, given the optimal choice of the level of education as a function of the number of children [determined by the first-order-condition in (6)] and the time dedicated by parents to nurturing activities as a function of the gross level of income and the number of children [determined by the first-order condition in (7)], the triplet $< z(w), y(w), n(w) >$ uniquely determines the level of utility, $U(w)$, by virtue of condition (3). Thus, our transformation is with no loss in generality.

Recalling that $z(w)$, $y(w)$, $n(w)$, $\alpha(w)$ and $e(w)$, denote the optimal choices of the w-household; substitution for $z[y(w),n(w)]$ from condition (3) into (B9) and re-arrangement yields the following expression, which is identical to equation (12):

(B10)

$$\int_{\underline{w}}^{\bar{w}} \begin{bmatrix} y(w) - U(w) - n(w) \cdot e[n(w)] + h\big[1 - y(w)/w - \alpha[n(w), y(w)] \cdot n(w)\big] + v[n(w)] \\ + u\big[e[n(w)]\big] + \Phi\big[\alpha[n(w), y(w)]\big] \end{bmatrix} dF(w) = R.$$

This completes the derivation.

Appendix C: Derivation of Equation (18)

In this appendix we derive the formula for the optimal marginal child benefit given in equation (18). Formulating the necessary first-order conditions for the *Hamiltonian* in (17) yields:

(C1)
$$\frac{\partial H}{\partial n} = \lambda \cdot f \cdot \begin{bmatrix} -e - n \cdot e'(n) - h'(1 - y/w - n \cdot \alpha) \cdot \left[\alpha + n \cdot \frac{\partial \alpha}{\partial n} \right] \\ + v'(n) + u'(e) \cdot e'(n) + \Phi'(\alpha) \cdot \frac{\partial \alpha}{\partial n} \end{bmatrix}$$
$$- \mu \cdot h''(1 - y/w - n \cdot \alpha) \cdot \frac{y}{w^2} \cdot \left[\alpha + n \cdot \frac{\partial \alpha}{\partial n} \right] = 0,$$

(C2)

$$\frac{\partial H}{\partial y} = \lambda \cdot f \cdot \left[1 - h'(1 - y/w - n \cdot \alpha) \cdot \left[1/w + n \cdot \frac{\partial \alpha}{\partial y} \right] + \Phi'(\alpha) \cdot \frac{\partial \alpha}{\partial y} \right]$$
$$+ \mu \cdot \left[-h''(1 - y/w - n \cdot \alpha) \cdot \frac{y}{w^2} \cdot \left[1/w + n \cdot \frac{\partial \alpha}{\partial y} \right] + h'(1 - y/w - n \cdot \alpha) \cdot \frac{1}{w^2} \right] = 0,$$

(C3) $$\frac{\partial H}{\partial U} = G'(U) \cdot f - \lambda \cdot f = -\mu'.$$

The *transversality* conditions are given by:

(C4) $\mu(\underline{w}) = \mu(\overline{w}) = 0$, ($\lim_{\overline{w} \to \infty} \mu(\overline{w}) = 0$, when the distribution of skills is unbounded).

Integrating condition (C3), employing the *transversality* condition, $\mu(\overline{w}) = 0$, yields:

(C5) $$\mu(w) = \int_{w}^{\overline{w}} \left[G'[U(t)] - \lambda \right] dF(t).$$

Employing the second *transversality* condition, $\mu(\underline{w}) = 0$, yields:

(C6) $$\lambda = \int_{\underline{w}}^{\overline{w}} G'[U(t)] dF(t).$$

Now define the function D by:

(C7) $$D(w) = \frac{1}{1-F(w)} \int_{w}^{\bar{w}} G'[U(t)]dF(t) .$$

In words, the function *D* measures the average social marginal utility of income over the interval $[w, \bar{w}]$. By virtue of the concavity of *G*, it follows that *D*(*w*) is decreasing. Moreover, employing (C5) and (C6) yields:

(C8) $$\mu(w) = [1 - F(w)] \cdot [D(w) - D(\underline{w})] ,$$

(C9) $$\lambda = D(\underline{w}) .$$

Substituting from (C8) and (C9) into (C1), employing the first-order condition in (4), yields, after some re-arrangements, the following expression [which is identical to equation (18)]:

(C10) $$z_n = \left[1 - \frac{D(w)}{D(\underline{w})}\right] \cdot \frac{1-F(w)}{f(w)} \cdot h''[1 - y/w - n \cdot \alpha] \cdot \frac{y}{w^2}\left[\alpha + n \cdot \frac{\partial \alpha}{\partial n}\right].$$

Appendix D: Proof of Proposition 1

We reproduce, for convenience, the condition for the optimal marginal child benefit given in (18):

(D1) $$z_n = \left[1 - \frac{D(w)}{D(\underline{w})}\right] \cdot \frac{1-F(w)}{f(w)} \cdot h''\left[1 - y/w - n\cdot\alpha\right] \cdot \frac{y}{w^2}\left[\alpha + n\cdot\frac{\partial\alpha}{\partial n}\right],$$

where $D(w) = \frac{1}{1-F(w)}\int_{w}^{\bar{w}} G'[U(t)]dF(t).$

By the concavity of G, it follows that $D(w)$ is decreasing. Thus, by virtue of the concavity of h, it follows that:

(D2) $$z_n < 0 \Leftrightarrow \alpha + n\cdot\frac{\partial\alpha}{\partial n} > 0 .$$

Suppose by negation that $z_n \geq 0$, hence, $\alpha + n\cdot\frac{\partial\alpha}{\partial n} \leq 0$.

Differentiating the first-order condition in (7) with respect to n yields:

(D3) $$-n\cdot h''(1-y/w-n\cdot\alpha)\cdot\left[\alpha + n\cdot\frac{\partial\alpha}{\partial n}\right] + h'(1-y/w-n\cdot\alpha) = \Phi''(\alpha)\cdot\frac{\partial\alpha}{\partial n}.$$

By our presumption, it follows from (D3) that:

(D4) $$\alpha + n\cdot\frac{\partial\alpha}{\partial n} = \left[\frac{\Phi''(\alpha)\cdot\frac{\partial\alpha}{\partial n} - h'(1-y/w-n\cdot\alpha)}{-n\cdot h''(1-y/w-n\cdot\alpha)}\right] \leq 0 .$$

Thus, by virtue of the concavity of h it follows that:

(D5) $$\Phi''(\alpha)\cdot\frac{\partial\alpha}{\partial n} - h'(1-y/w-n\cdot\alpha) \leq 0 .$$

Substituting for h' from the first-order condition in (7) into (D5) and re-arranging yields:

(D7) $$\Phi''(\alpha)\cdot n\cdot\frac{\partial\alpha}{\partial n} - \Phi'(\alpha) \leq 0 .$$

By virtue of our presumption it follows:

(D8) $\alpha \leq -n \cdot \frac{\partial \alpha}{\partial n}$.

Hence it follows from (D7) that:

(D9) $-\Phi''(\alpha) \cdot \alpha - \Phi'(\alpha) \leq 0 \Leftrightarrow \frac{-\Phi''(\alpha) \cdot \alpha}{\Phi'(\alpha)} \leq 1$

This violates our parametric assumption, given by condition (A) in the main text, that:

$\frac{-\Phi''(\alpha) \cdot \alpha}{\Phi'(\alpha)} > 1$.

We thus obtain the desired contradiction.

Appendix E: A Means-Tested System Strictly Dominates the Universal one

In this appendix we demonstrate the dominance of a means-tested child benefit system over a universal one. We restrict attention to a setting with a discrete distribution of skill levels and restrict attention to smooth (continuously differentiable) tax-and-transfer functions.

Consider an economy with a finite number of skill levels denoted by: $w_i; i = 1,2....N$, where $0 < w_1 < w_2 < ......w_{N-1} < w_N$. We normalize the number of individuals of each type to unity with no loss in generality. We let $J[w_i, z(y,n), y, n]$ denote the maximal level of utility derived by a household with ability level w_i, gross income level y, net income level $z(y,n)$ and number of children n. Formally,

(E1)

$$J[w_i, z(y,n), y, n] = \max_{e,\alpha}\left[z(y,n) - n \cdot e + v(n) + u(e) + h(1 - y/w_i - n \cdot \alpha) + \Phi(\alpha)\right].$$

We assume that the government is implementing a universal system; namely:

(E2) $z(y,n) = a(y) + b(n)$.

We turn to show that any universal system of the form given in condition (E2) can be replaced by a means-tested system that maintains the government revenue requirement and attains a higher level of welfare.

Denote by (y_i, n_i) the optimal choice of an individual with skill-level w_i, faced with the universal system given in condition (E2). Consider the following (means-tested) alternative tax-transfer schedule:

(E3) $$\hat{z}(y,n) = \begin{cases} 0 & y = y_1, n \neq n_1 \\ z(y,n) & \text{otherwise} \end{cases}$$

Notice, that indeed the schedule given in (E3) is means-tested. To see this consider some arbitrary number of children, $n \neq n_1$. Then, the marginal child subsidy is given by *b'*(*n*) for levels of income $y \neq y_1$ and zero when the income level is y_1.

It is straightforward to verify that the optimal choices of the individuals under the modified system given by (E3) will coincide with choices taken under the system in (E2).

By revealed preference considerations, it follows that:

(E4) $$J[w_i, z(y_i, n_i), y_i, n_i] \geq \max_n J[w_i, z(y_1, n), y_1, n] \geq J[w_i, z(y_1, n_1), y_1, n_1],$$

for all *i*>1.

In fact, the last inequality in condition (E4) is strict. To see this, notice that by the individual first-order conditions:

(E5) $$v'(n_1) + b'(n_1) - \alpha_1 \cdot h'(1 - y_1 / w_1 - n_1 \cdot \alpha_1) - e_1 = 0,$$

(E6) $$u'(e_1) - n_1 = 0,$$

(E7) $$\Phi'(\alpha_1) - n_1 \cdot h'(1 - y_1 / w_1 - n_1 \cdot \alpha_1) = 0$$

Fixing n_1 and y_1, and differentiating the first-order condition in (7) with respect to *w*, yields:

(E8) $$\Phi''(\alpha) \cdot \frac{\partial \alpha}{\partial w} = n_1 \cdot h''(1 - y_1 / w - n_1 \cdot \alpha) \cdot \left[y_1 / w^2 - n_1 \cdot \frac{\partial \alpha}{\partial w} \right].$$

Re-arranging yields:

(E9) $$\frac{\partial \alpha}{\partial w} = \left[\frac{n_1 \cdot h''(1 - y_1 / w - n_1 \cdot \alpha) \cdot y_1 / w^2}{\Phi''(\alpha) + n_1^{\;2} \cdot h''(1 - y_1 / w - n_1 \cdot \alpha)} \right] > 0,$$

where the sign of inequality follows by the concavity of *h* and Φ.

By re-arranging the first-order condition in (7) it follows that:

(E10) $$\alpha(w) \cdot h'\left[1 - y_1 / w - n_1 \cdot \alpha(w)\right] = \Phi'[\alpha(w)] \cdot \frac{\alpha(w)}{n_1}.$$

Hence,

(E11)
$$\frac{\partial}{\partial w}\left[\alpha(w)\cdot h'[1-y_1/w-n_1\cdot\alpha(w)]\right]$$
$$=\frac{\partial\alpha(w)}{\partial w}\cdot\frac{1}{n_1}\cdot\left[\Phi''[\alpha(w)]\cdot\alpha(w)+\Phi'[\alpha(w)]\right]<0,$$

where the inequality sign follows from our parametric assumption, given in condition (A) in the main text, and (E9).

Thus,

(E12) $$v'(n_1)+b'(n_1)-\alpha(w_i)\cdot h'[1-y_1/w_i-n_1\cdot\alpha(w_i)]-e_1>0,$$

for all $i>1$.

Thus, conditional on the level of income, y_1, the optimal number of children chosen by an individual of type $i>1$ differs from n_1.

Let $\delta_i \equiv \max_n J[w_i, z(y_1,n), y_1, n] - J[w_i, z(y_1,n_1), y_1, n_1]$. As we have just shown, $\delta_i > 0$ for all $i>1$. Further, let $\delta \equiv \min_{i>1}(\delta_i) > 0$.

Consider next the following tax-transfer schedule:

(E13) $$z^*(y,n)=\begin{cases}0 & y=y_1, n\neq n_1\\ z(y,n)+(N-1)\cdot\delta/N & y=y_1, n=n_1\\ z(y,n)-\delta/N & \text{otherwise}\end{cases}$$

The tax-transfer schedule in (E13) is a modification of the schedule given in (E3). This schedule is obtained by levying a lump sum tax by the amount δ on all income levels different than y_1, and then distribute the entailed extra tax revenues in a lump sum fashion. Notice that by construction of δ, the optimal choices of the individuals will still coincide with those obtained under the schedule given in (E3) and (E2). Moreover, the government revenue constraint will still be satisfied. However, as we redistribute resources from individuals of type $i>1$ towards the least well-off individual ($i=1$), we obtain a welfare gain, due the strictly concave welfare function [given in (10)].

The idea underlying the construction was that by forcing individuals of type $i>1$ who mimic the least well-off type ($i=1$), by choosing her level of income, also to choose the same number of children, we create a slack in the binding incentive constraints. We can then employ this slack to redistribute more towards the least well-off individual.

Appendix F: Derivation of Equation (19)

In this appendix we derive the formula for the optimal marginal child benefit given in equation (19). Formulating the necessary first-order conditions for the *Hamiltonian* in (17’) yields:

$$\frac{\partial H}{\partial n} = \lambda \cdot f \cdot \begin{bmatrix} y'(n) - e - n \cdot e'(n) - h'(1 - y/w - n \cdot \alpha) \cdot \left[y'(n)/w + \alpha + n \cdot \alpha'(n) \right] \\ + v'(n) + u'(e) \cdot e'(n) + \Phi'(\alpha) \cdot \alpha'(n) \end{bmatrix}$$

(F1)

$$+ \mu \cdot \begin{bmatrix} h'(1 - y/w - n \cdot \alpha) \cdot y'(n)/w^2 \\ - h''(1 - y/w - n \cdot \alpha) \cdot \frac{y}{w^2} \cdot \left[y'(n)/w + \alpha + n \cdot \alpha'(n) \right] \end{bmatrix} = 0,$$

(F2) $$\frac{\partial H}{\partial U} = G'(U) \cdot f - \lambda \cdot f = -\mu'.$$

The *transversality* conditions are given by:

(F3) $\mu(\underline{w}) = \mu(\overline{w}) = 0$, ($\lim_{\overline{w} \to \infty} \mu(\overline{w}) = 0$, when the distribution of skills is unbounded).

Integrating condition (F2), employing the *transversality* condition, $\mu(\overline{w}) = 0$, yields:

(F4) $$\mu(w) = \int_{w}^{\overline{w}} \left[G'[U(t)] - \lambda \right] dF(t).$$

Employing the second *transversality* condition, $\mu(\underline{w}) = 0$, yields:

(F5) $$\lambda = \int_{\underline{w}}^{\overline{w}} G'[U(t)] dF(t).$$

Now define the function D by:

(F6) $$D(w) = \frac{1}{1 - F(w)} \int_{w}^{\overline{w}} G'[U(t)] dF(t).$$

Employing (F4) and (F5) yields:

(F7) $$\mu(w) = [1 - F(w)] \cdot [D(w) - D(\underline{w})],$$

(F8) $$\lambda = D(\underline{w}).$$

Substituting from (F7) and (F8) into (F1), employing the household's first-order conditions in (4)-(7) yields, after some re-arrangements, the following expression [which is identical to equation (19)]:

(F9) $$b_n = (1-a_y)\cdot y' + \left[1-\frac{D(w)}{D(\underline{w})}\right]\cdot\frac{1-F(w)}{f(w)\cdot w}\cdot\left[h''\cdot(y'/w+\alpha+n\cdot\alpha')\cdot y/w - a_y\cdot y'\right].$$

Appendix G: The Correlation between Family Size and Earning Ability and the Properties of the Income Tax Schedule

In this appendix we state and prove the following claim:

Claim: $n'(w) >=< 0$ if and only if $a_{yy} \cdot y + a_y <=> 0$

Proof: We reproduce, for convenience, the *w*-household's first-order conditions, given in equations (4)-(7), assuming a universal system, namely, $z(y,n) = a(y) + b(n)$ and setting the marginal child subsidy to zero ($b_n = 0$):

(G1) $v'(n) - \alpha \cdot h'(1 - y/w - n \cdot \alpha) - e = 0$,

(G2) $a_y \cdot w - h'(1 - y/w - n \cdot \alpha) = 0$,

(G3) $u'(e) - n = 0$,

(G4) $\Phi'(\alpha) - n \cdot h'(1 - y/w - n \cdot \alpha) = 0.$

We let *e*(*n*) and $\alpha(w,n,y)$ denote the implicit solutions to (G3) and (G4), respectively. Substituting into (G1) and (G2) then yields:

(G1') $H(n,y,w) \equiv v'(n) - \alpha \cdot h'[1 - y/w - n \cdot \alpha(w,n,y)] - e(n) = 0,$

(G2') $K(n,y,w) \equiv a_y \cdot w - h'[1 - y/w - n \cdot \alpha(w,n,y)] = 0.$

The systems of two equations [(G1') and (G2')] provide an implicit solution for *n*(*w*) and *y*(*w*), the optimal choices of the *w*-household.

Fully differentiating the two conditions in (G1') and (G2') with respect to *w* yields:

(G5) $\partial H/\partial n \cdot n'(w) + \partial H/\partial y \cdot y'(w) + \partial H/\partial w = 0,$

(G6) $\partial K/\partial n \cdot n'(w) + \partial K/\partial y \cdot y'(w) + \partial K/\partial w = 0.$

Employing *Cramer's Rule* then yields:

(G7) $$n'(w) = \frac{-\partial H/\partial w \cdot \partial K/\partial y + \partial K/\partial w \cdot \partial H/\partial y}{\partial H/\partial n \cdot \partial K/\partial y - \partial K/\partial n \cdot \partial H/\partial y}$$

By the second-order conditions of the household's optimization it follows that $\partial H/\partial n \cdot \partial K/\partial y - \partial K/\partial n \cdot \partial H/\partial y$ >0. Thus, it follows:

(G8) $Sign[n'(w)] = Sign[-\partial H/\partial w \cdot \partial K/\partial y + \partial K/\partial w \cdot \partial H/\partial y]$.

Differentiating the household's first-order conditions in (G1') and (G2') yields:

(G9) $$\partial H/\partial w = -\frac{\partial \alpha}{\partial w} \cdot h'(1 - y/w - n\cdot\alpha) - \alpha \cdot h''(1 - y/w - n\cdot\alpha) \cdot \left[y/w^2 - n \cdot \frac{\partial \alpha}{\partial w} \right],$$

(G10) $$\partial H/\partial y = -\frac{\partial \alpha}{\partial y} \cdot h'(1 - y/w - n\cdot\alpha) - \alpha \cdot h''(1 - y/w - n\cdot\alpha) \cdot \left[-1/w - n \cdot \frac{\partial \alpha}{\partial y} \right],$$

(G11) $$\partial K/\partial y = a_{yy} \cdot w - h''(1 - y/w - n\cdot\alpha) \cdot \left[-1/w - n \cdot \frac{\partial \alpha}{\partial y} \right],$$

(G12) $$\partial K/\partial w = a_y - h''(1 - y/w - n\cdot\alpha) \cdot \left[y/w^2 - n \cdot \frac{\partial \alpha}{\partial w} \right].$$

Differentiating the first-order condition in (G4) with respect to *w* and with respect to *y*, yields:

(G13) $$\Phi''(\alpha) \cdot \frac{\partial \alpha}{\partial w} = n \cdot h''(1 - y/w - n\cdot\alpha) \cdot \left[y/w^2 - n \cdot \frac{\partial \alpha}{\partial w} \right],$$

(G14) $$\Phi''(\alpha) \cdot \frac{\partial \alpha}{\partial y} = n \cdot h''(1 - y/w - n\cdot\alpha) \cdot \left[-1/w - n \cdot \frac{\partial \alpha}{\partial y} \right].$$

Dividing (G13) by (G14) then yields, after re-arrangement:

(G15) $$-\frac{\partial \alpha}{\partial w} = \frac{\partial \alpha}{\partial y} \cdot \frac{y}{w}.$$

Substituting from (G9)-(G12) into (G7), then substituting for $\partial \alpha/\partial w$ from (G15) and re-arranging (omitting some of the argument to abbreviate notation) yields:

(G16) $$n'(w) = [a_{yy} \cdot y + a_y] \cdot \left[\alpha \cdot h''/w + \frac{\partial \alpha}{\partial y} \cdot [\alpha \cdot n \cdot h'' - h'] \right].$$

Substituting for $\partial \alpha/\partial y$ from (G14) into (G16) yields, after re-arrangement:

(G17) $$n'(w) = [a_{yy} \cdot y + a_y] \cdot \left[\frac{h'' \cdot (\alpha \cdot \Phi'' + n \cdot h')}{w \cdot (\Phi'' + n^2 \cdot h'')} \right].$$

By virtue of (G4) $n \cdot h' = \Phi'$. The result follows, then, by the concavity of *h* and Φ and our parametric assumption given in condition (A) in the main text ($\alpha \cdot \Phi'' + \Phi' < 0$).

Appendix H: Derivation of Equation (25)

In this appendix we derive the formula for the optimal marginal income tax given in equation (25). The *Hamiltonian* for the government program is given by:

(H1)
$$H = \left[G(U) + \lambda \cdot \begin{bmatrix} y - U - n(y) \cdot e[n(y)] + h[1 - y/w - n(y) \cdot \alpha(y)] + v[n(y)] \\ + u\big[e[n(y)]\big] + \Phi\big[\alpha[n(y)]\big] - R \end{bmatrix} \right] \cdot f$$
$$+ \mu \cdot \Big[h'\big[1 - y/w - n(y) \cdot \alpha[n(y)]\big] \cdot y/w^2\Big],$$

where $n(y), e[n(y)]$ and $\alpha(y)$ are implicitly given by the household's first-order conditions in (4), (6) and (7).

Formulating the necessary first-order conditions for the *Hamiltonian* in (H1), omitting the arguments to abbreviate notation, yields:

(H2)
$$\frac{\partial H}{\partial y} = \lambda \cdot f \cdot \big[1 - n' \cdot e - n \cdot e' \cdot n' + h' \cdot (-1/w - n' \cdot \alpha - n \cdot \alpha') + v' \cdot n' + u' \cdot e' \cdot n' + \Phi' \cdot \alpha'\big]$$
$$+ \mu \cdot \left[h'/w^2 + h'' \cdot \frac{y}{w^2} \cdot (-1/w - n' \cdot \alpha - n \cdot \alpha') \right] = 0,$$

(H3) $$\frac{\partial H}{\partial U} = G'(U) \cdot f - \lambda \cdot f = -\mu'.$$

The *transversality* conditions are given by:

$(H4)$ $\mu(\underline{w}) = \mu(\overline{w}) = 0$, ($\lim_{\overline{w} \to \infty} \mu(\overline{w}) = 0$, when the distribution of skills is unbounded).

Integrating condition (H3), employing the *transversality* condition, $\mu(\overline{w}) = 0$, yields:

$(H5)$ $$\mu(w) = \int_{w}^{\overline{w}} \big[G'[U(t)] - \lambda\big] dF(t).$$

Employing the second *transversality* condition, $\mu(\underline{w}) = 0$, yields:

$(H6)$ $$\lambda = \int_{\underline{w}}^{\overline{w}} G'[U(t)] dF(t).$$

Now define the function *D* by:

(H7) $$D(w)=\frac{1}{1-F(w)}\int_{w}^{\bar{w}}G'[U(t)]dF(t)\,.$$

Employing (H5) and (H6) yields:

(H8) $$\mu(w)=[1-F(w)]\cdot[D(w)-D(\underline{w})]\,,$$

(H9) $$\lambda=D(\underline{w})\,.$$

Substituting from (H8) and (H9) into (H2), employing the household's first-order conditions in (4)-(7) yields, after some re-arrangements, the following expression [which is identical to equation (25)]:

(H10) $$\frac{1-a_y}{a_y}=\left[1-\frac{D(w)}{D(\underline{w})}\right]\cdot\frac{1-F(w)}{f(w)\cdot w}\cdot\left[1-\frac{h''\cdot y}{w\cdot a_y}\cdot(1/w+n'\cdot\alpha+n\cdot\alpha')\right].$$

It remains to show that ε_L denotes the labor supply elasticity, given by:

(H11) $$\varepsilon_L=-\frac{w\cdot a_y}{h''\cdot y}\cdot\frac{1}{(1/w+n'\cdot\alpha+n\cdot\alpha')}.$$

To see this, let $w_n\equiv w\cdot a_y$ denote the after-tax wage rate. By virtue of the household first order condition in (5) it follows that:

(H12) $$w_n=h'(l)\,.$$

Hence,

(H13) $$\partial l/\partial w_n=1/h''(l)\,.$$

By definition,

(H14) $$\varepsilon_L=\frac{\partial(y/w)}{\partial w_n}\cdot\frac{w_n}{(y/w)}=\frac{\partial y}{\partial w_n}\cdot\frac{w_n}{y}\,.$$

By construction,

(H15) $$\frac{\partial y}{\partial w_n}=\frac{\partial l/\partial w_n}{\partial l/\partial y}$$

Moreover, as $l\equiv 1-y/w-\alpha\cdot n$, it follows that:

(H16) $\partial l / \partial y = -(1 / w + n' \cdot \alpha + n \cdot \alpha')$.

Substituting from (H13), H(15) and (H16) into (H14) yields the elasticity formula in (H11).

References

Akerlof, G. (1978). “The Economics of Tagging as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning,” American Economic Review, 68, 8-19.

Andreoni, J. (1990). "Impure Altruism and Donations to Public Goods: a Theory of Warm-Glow Giving," Economic Journal 38, 209–227.

Atkinson, A. and Stiglitz, J. (1976). “The Design of Tax Structure: Direct versus Indirect Taxation,” Journal of Public Economics, 6, 55-75.

Balcer, Y. and Sadka, E. (1982). “Horizontal Equity, Income Taxation and Self-selection with an Application to Income Tax Credits," Journal of Public Economics, 19, 291-309.

Balcer, Y. and Sadka, E. (1986). "Equivalence Scales, Horizontal Equity and Optimal Taxation under Utilitarianism," Journal of Public Economics, 29, 79-97

Barro, R.J. (1974). "Are Government Bonds Net Wealth?," Journal of Political Economy 82, 1095–1117.

Becker, G. (1960). “An Economic Analysis of Fertility,” in Demographic and Economic Change in Developed Countries, National Bureau of Economic Research Series, Number 11, Princeton, NJ: Princeton University Press, 209–231.

______ . (1991). “A Treatise on the Family” (2nd enlarged edition), Harvard University Press, Cambridge, Mass.

Becker, G. and Barro, R. (1988). “A Re-formulation of the Economic Theory of Fertility,” Quarterly Journal of Economics, 103, 1-25.

Becker, G. and Lewis, G. (1973) "On the Interaction between the Quantity and Quality of Children," Journal of Political Economy, 81, 279–288.

Blumkin, T. et al (2010) “Taxing Children: The Re-distributive Role of Child Benefits – Revisited”, CesIfo Working Paper #2970

Blundell, R. and MaCurdy, T. (1999) “Labor Supply: A Review of Alternative Approaches”, in O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, Volume IIIA, Amsterdam, North-Holland.

Bradshaw, J. and Finch, N. (2002). “A Comparison of Child Benefit Packages in 22 Countries,” Department for Work and Pensions Research Report No.174, Corporate Document Services: Leeds.

Cigno, A. (1986). “Fertility and the Tax-Benefit System: A Reconsideration of the Theory of Family Taxation,” Economic Journal, 96, 1035-1051

______ . (1993). “Intergenerational Transfers without Altruism: Family, Market and State,” European Journal of Political Economy

______ . (2001). “Comparative Advantage, Observability and the Optimal Tax Treatment of Families with Children,” International Tax and Public Finance, 8 455-470

______ . (2009). “Agency in Family Policy: A Survey,” CesIfo Working Paper # 2664.

Cigno, A. and Pettini A. (2002). “Taxing Family Size and Subsidizing Child-specific Commodities?” Journal of Public Economics, 87, 75-90.

Cohen, A., Dehejia, R. and Romanov, D. (2007). “Do Financial Incentives Affect Fertility?” NBER Working Paper 13700, Cambridge, Massachusetts, revised May 2009.

Cremer, H., Dellis, A. and Pestieau, P. (2003). "Family Size and Optimal Income Taxation," Journal of Population Economics 16, 37-54.

Cremer, H., Lozachmeur, J.M. and Pestieau, P. (2009). “Income Taxation of Couples and the Tax Unit Choice,” CORE DP 2007/13.

Diamond, P. (1998). “Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates,” American Economic Review, 88, 83-95.

Ebert, U. (1992). “A Reexamination of the Optimal Nonlinear Income Tax,” Journal of Public Economics 49, 47–73.

Kleven, H.J., Kreiner, C.T. and Saez, E. (2009). “The Optimal Income Taxation of Couples,” Econometrica, 77, 537-560.

Hanushek, E. (1992) "The Trade-Off between Child Quantity and Quality," Journal of Political Economy, 100, 84-117.

Hotz, V.J., Klerman, J.A. and Willis, R.J. (1997). "The Economics of Fertility in Developed Countries," in M.R. Rosenzweig and O. Stark (editors), Handbook of Population and Family Economics, Elsevier.

Laroque, G. and Salanie, B. (2008). “Does Fertility Respond to Financial Incentives?”, CesIfo Working Paper # 2339.

Mirrlees, J. (1971) “An Exploration in the Theory of Optimum Income Taxation,” Review of Economic Studies, 38, 175-208.

Mirrlees, J. (1976) "Optimal Tax Theory: A Synthesis," Journal of Public Economics, 6, 327-358.

Moav, O. (2005) "Cheap Children and the Persistence of Poverty," Economic Journal, 115, 88-110.

Phelps, E. (1973) "Taxation of Wage Income for Economic Justice," Quarterly Journal of Economics, 87, 331-354.

Sadka, E. (1976) “On Income Distribution, Incentive Effects and Optimal Income Taxation,” Review of Economic Studies, 43, 261-268.

Saez, E. (2002) "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses", Quarterly Journal of Economics, 117, 1039-1073.

Salanie, B. (2003). “The Economics of Taxation”, The MIT Press, Cambridge, Mass.

Seade, J. (1977). On the Shape of Optimal Tax Schedules. Journal of Public Economics 7: 203–235.

Stiglitz, J. (1982). "Self Selection and Pareto Efficient Taxation," Journal of Public Economics, 17, 213-240