No. 2011-003

Family Size and Optimal Taxation in the Extensive Model

Yoshihiro TAKAMATSU*†

Graduate School of Commerce, Waseda University

August 11, 2011

Abstract

This paper studies optimal taxation with extensive labor supply margin and variable family size. Households are differentiated by market and non-market (child-care) ability, and the number of children. When the government only employs non-linear income tax, in-work credit can be optimal within every family size class. The joint distribution of ability pattern affects the size of in-work credit. The relationship between income tax liabilities and family size depends on the government's redistributive taste, extensive labor response and lump-sum transfer. When the government also imposes a tax/subsidy on child-specific commodities, the desirability of in-work credit appears to be ambiguous. The role of differential commodity tax is that it reveals her child-care ability whenever the demand for child-specific commodity is related to the child-care ability. Furthermore, we confirm some conditions that Atkinson-Stiglitz Theorem remains valid in this context.

JEL classification: H21, H23, J13, D13

Keywords: Optimal taxation; Family size; Extensive margin

* E-mail: yoshihiro.takamatsu@gmail.com

† I would like to thank Professor Nobutake Yokota for providing invaluable guidance. I would also like to thank Professor Takao Kataoka and Professor Koichi Takase for their insightful comments. Any errors are the responsibility of the author.

1. Introduction

In the past two decades, most developed countries have introduced or extended in-work support through tax credits and work-conditioned transfers as a means of providing cash assistance to low income families with children. These programs intend to alleviate poverty without creating adverse incentives for participation in the labor market. The precise structure of these programs, however, differs substantially from country to country. In general, these programs are part of the income tax system and depend not only on the level of income (or the hours of labor) but also on family size. Therefore, we try to design an optimal (direct and indirect) tax system that maximizes a utilitarian social welfare function and takes into account variable family size.

The optimal income tax literature has developed models to analyze the design of income tax-transfer programs, following the seminal theoretical contribution of Mirrlees (1971). In that framework, which focuses exclusively on the case in which households choose hours of work or intensity of work (i.e., on the intensive margin), it can be shown that negative marginal tax rates can never be optimal, ruling out in-work credits. Moreover, numerical simulations have shown that, in this model, optimal marginal tax rate at the bottom are very high (see, e.g., Tuomala (1984) and Saez (2001)). Redistribution thus takes a form of traditional Negative Income Tax type program with a substantial guaranteed income support and a large phase-out tax rate. Diamond (1980), on the other hand, extended the model of optimal income taxation by focusing only on the extensive margin. In the extensive margin model, people of type (or productivity) n choose only whether to work at a job of type n or not to work. In this setting, there is no need to concern about a household of type n choosing to work at a job of type $n-1$ and earning a lower income, while in the intensive margin model, there is a possibility that a high productivity household may mimic a low productivity household. Saez (2002) has also demonstrated that the incorporation of extensive labor margin has important implications for the theory of optimal income taxation. Following Saez (2002), the optimal income transfer program is similar to U.S. the Earned Income Tax Credit (EITC) with negative marginal tax rates at low income levels and a small guaranteed income.

In this context, there is a question whether family sizes should be taxed or subsidized. A large number of studies have been made on this subject. Cremer, Dellis, and Pestieau (2003), assuming exogenous fertility, argue that families with a large number of children should receive a more favorable income tax treatment, the rationale for this being that large households face a higher total cost of children and should therefore be compensated by the tax system. Balestrino, Cigno, and Pettini (2002) , on the other hand, take fertility to be endogenous and assume that the tax system comprises non-linear income and children taxes, as well as linear commodity taxes. They find that it may not be optimal to design the tax system so that an additional child would lighten the net tax burden on his or her parents. These results, however, have been obtained in the intensive labor model. Little attention has been given to the point of the extensive margin.

In this paper, we explore the implications of extensive labor margin for the size of family and child care. There are several reasons for our interest. One is that the empirical literature has shown that labor supply effects on the extensive margin tend to be more important. Meghir and

Phillips (2010), for example, show that the decision whether or not to take paid work at all is quite sensitive to taxation and benefits for women and mothers in particular while hours of work do not respond particularly strongly to the financial incentives created by tax changes for men and a little more responsive for married women and lone mothers. Another is that households with a large number of children, especially lone mothers, are eligible for generous transfer programs. These programs introduce distortions that might lead to substantial disincentives in labor market. Thus, extensive margin and family size are two topics that offer the key to the structure of second-best redistributive tax system.

Throughout this paper, we assume that the number of children varies across families but that it is exogenous. Endogenous fertility could have some important implications, but here we limit the discussion to fertility as exogenous.[1] Furthermore, we assume that children welfare depends on their parents decisions. Thus, we consider an economy where households are heterogeneously endowed with two unobserved characteristics and a observed one: first two are their skill level in work force and in home, especially child care, and the other is the number of children. According to the size of disposal income, households may participate in work or concentrate on child care (not work).

The structure of the paper is the following. In Section 2, we present a general model of extensive labor supply with variable family size. In Section 3, we examine optimal income tax problem and focus on conditions whether family size is tax asset or not. Section 4 reconsider optimal tax problem when the government have the instruments to tax the child-specific commodities. Finally, Section 5 concludes.

2. Model

2.1. The Household

Consider a society where each household consists of a parent and dependent children. The dependent child can not supply her labor service. A parent is described by a set of exogenous characteristics, denoted by $\boldsymbol{a} = (w, s, n)$. The first coordinate of $\boldsymbol{a}$, $w \in [0, \bar{w}]$, denotes her ability to earn money, represented by the wage rate. The second coordinate of $\boldsymbol{a}$, $s \in [s_0, \bar{s}]$, denotes her ability to raise children, represented by a domestic productivity parameter. These characteristics are private information and not observable by the government. As described below, however, a parent's labor supply behavior in our model is only extensive margin so that when she works her wage is revealed to and observed by the government. The third coordinate of $\boldsymbol{a}$, $n \in [0, \bar{n}]$, denotes a number of household's children. n is assumed to be observable by the government. We assume that the total population is normalized to one and $\boldsymbol{a}$ have the joint density $f(\boldsymbol{a})$ and corresponding c.d.f. $F(\boldsymbol{a})$.

The labor-leisure choice of the parent in our model is whether to participate, or not, in the work force (extensive margin). The participation status of parent is described with a function

[1] Balestrino, Cigno, and Pettini (2002) pointed out that the number of children conveys a great deal of information about that person's characteristics and it helps to relax the self-selection constraints on the design of policy.

$L(\boldsymbol{a})$, where $L(\boldsymbol{a})$ is equal to 0 (no work) or 1 (work). As is standard practice in optimal taxation models, we assume that (industrial) production uses labor as the only input and takes place in perfectly competitive industries endowed with linear technologies. Therefore, producer prices of commodities are fixed and can be normalized to one. When parent $\boldsymbol{a}$ participates in the work force, she produces w unit of commodity and thus w is also the parent's earnings (pre-tax income), while she does not produce any marketable good when she does not participate. In our model, a parent whose wage is w can not mimic her wage to the other one's.[2]

Household preferences are described by a concave utility function,

$$U^a = U(x, Q; \boldsymbol{a}),$$

where x is parent's consumption, Q is an index of the children's quality of life ("quality" for short). We assume same Q is assigned to all children born in the same household. We may think of Q as of a composite consumption good, specific to children of that particular household, domestically produced by the child's parent with inputs of own time and child-specific commodities bought from the market (Becker(1991)). Alternatively, we may think of Q as of (the parental perception of) a child's lifetime utility, conditional on how much time and money the parent has spent on that child. If we favor the second interpretation, U becomes a kind of household-level social welfare function (Cigno (1991)). Either way, Q will depend on the quantity of child-specific commodities per child, z, and parental time ("attention") per child, h, provided to each child, and the domestic ability parameter, s, $Q = Q(z, h; s)$. Thus children welfare only depends on their parental decisions in our model.

Normalizing the time endowment to one, we can write the time constraint as

$$L + H = 1,$$

where $H = nh$, is the total amount of time allocated to the care of children and L the labor supply. This means that there is no "pure leisure".

The government has two sets of instruments at its disposal; first, possibly non-linear income tax-transfer schedule, $T(w, n)$ to be conditioned on the number of children, n; second, child-specific commodities tax/subsidy, t (and the rate of tax on adult-specific commodities is normalized to zero). The child-specific commodity tax is indirect taxation so that the tax rate is linear. We denote the consumer price of child-specific commodities by p, $p = 1+t$. As mentioned above, for each household the government only observes the number of children, n and the wage w if she works. Since the government does not make any distinction among the unemployed households, the benefit for unemployed conditioned on the number of children, $T(0, n)$ are set uniformly. The government knows the joint distribution of $\boldsymbol{a}$ in the overall population, and therefore the conditional distributions, given the observables, as well.

[2] This means that labor inputs are imperfect substitutes. We can interpret this as an occupation (job) choice model where wages are different in each occupation and households choose their occupation among a continuum.

The household budget constraint is, if her wage is not 0,[3]

$$x + pZ = \begin{cases} w - T(w,n) = B(w,n), & \text{(if } L(a) = 1\text{)}, \\ -T(0,n) = B(0,n), & \text{(otherwise)}, \end{cases}$$

where $Z = nz$ is the household's total demand for child-specific commodities and B is a disposal income.

We can describe the household's choice as a two-step decision procedure.[4] At the first, a fixed amount of disposal income B is optimally allocated over the parental consumption goods and child-specific commodities, taking parental labor supply and government's policy menus $(t, T(w,n), T(0,n))$ as given. This gives conditional indirect utility,

$$V(p, B, \boldsymbol{a}) = \max_{z} \{U(B - pZ,\ Q(z, h; s); \boldsymbol{a})\},$$

and conditional demand function, $z = z(p, B; \boldsymbol{a})$.

At the second stage, labor participation status $L(a)$ is chosen to maximize $V(p, B, \boldsymbol{a})$ subject to

$$\begin{cases} B(w,n) = w - T(w,n) & \text{if } L = 1, \\ B(0,n) = -T(0,n) & \text{otherwise } (L = 0), \end{cases}$$

and $L + H = 1$. Households who have same wage and same number of children have heterogeneous abilities to raise children and choose labor participation status according to the relative after tax income in work and no work and to the consumer price of child-specific commodities. The larger after tax income with no work and the higher ability to raise children, the larger number of parent choose no work.

In particular, consider an household who is indifferent between working with after tax income $B(w,n)$ and not working with disposal income $B(0,n)$. In such a case, there exists

$$s^*(w,n) = s(p, B(w,n), B(0,n); w, n),$$

such that

$$V(p, B(w,n); w, s^*(w,n), n) = V(p, B(0,n); w, s^*(w,n), n). \tag{1}$$

These s^* are switching points of labor status if $s \leq s^*$ then $L(\boldsymbol{a}) = 1$, and if $s > s^*(w,n)$ then $L(\boldsymbol{a}) = 0$. As a result, each parent's total demand for child-specific commodities $Z(p, B; \boldsymbol{a})$ is discontinuous at point $s^*(w,n)$ where the parent is indifferent between work and no-work. However, at these switching points, the parent is indifferent between work and no-work and thus gets the same utility in both status. As a result, the conditional indirect utility V is continuous in s.

[3] If her wage is 0, she does not participate in work force. Hence, her disposal income is necessarily $-T(0,n) = B(0,n)$.

[4] A similar procedure has been used by Christiansen (1984), Edwards, Keen, and Tuomala (1994), Balestrino, Cigno, and Pettini (2002, 2003).

We denote by $D_{w,n}$ the probability she works conditional on the parent wage, w and on the number of children, n,

$$D_{w,n}(s^*(w,n)) = \Pr\left(s \leq s^*(w,n)|w,n\right) = \int_{s_0}^{s^*(w,n)} f(s|w,n)ds. \tag{2}$$

Presumably, $s^*(w,n)$ is increasing in $B(w,n)$ because if disposal income in work increases while prices and disposal income in no work remain constant, labor participation become more attractive and some no-work parents may switch to work. Similarly, $s^*(w,n)$ is presumably decreasing in $B(0,n)$. Hence, $D_{w,n}$ is also increasing in $B(w,n)$ and decreasing in $B(0,n)$.

2.2. The Government Problem

We assume that the government adopts the utilitarian welfare criterion. The social welfare function can be simply characterized by a weighted sum of household utilities. These non-negative weights are denoted by $\beta(\boldsymbol{a})$. Then, the government sets income tax $T(w,n)$, $T(0,n)$ and child-specific commodity tax t so as to maximize social welfare function,

$$\int\limits_{\boldsymbol{a}} \left\{\beta(\boldsymbol{a})V(p,B(w,n);\boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)} + \beta(\boldsymbol{a})V(p,B(0,n);\boldsymbol{a})\mathbb{1}_{s>s^*(w,n)}\right\} dF(\boldsymbol{a}), \tag{3}$$

subject to the budget constraint,

$$\int\limits_{\boldsymbol{a}} \left\{[T(w,n)]\mathbb{1}_{s\leq s^*(w,n)} + T(0,n)\mathbb{1}_{s>s^*(w,n)} + tZ(p,B;\boldsymbol{a})\right\} dF(\boldsymbol{a}) = R, \tag{4}$$

where R is exogenous government's revenue requirement and $\mathbb{1}_{s\leq s^*(w,n)}$ is an indicator function.[5]

Using (2) and employing the relation $f(a) = f(s|w,n)\tilde{f}(w,n)$, where $\tilde{f}(w,n)$ is marginal density function of wage and number of children in the population at w,n (i.e., $\tilde{f}(w,n) = \int_S f(w,s,n)ds$), the government's budget constraint can be rewritten as

$$\int\limits_{W,N} [T(w,n) - T(0,n)]D_{w,n}(s^*)d\tilde{F}(w,n) + \int\limits_{\boldsymbol{a}} [T(0,n) + tZ(p,B;\boldsymbol{a})]dF(\boldsymbol{a}) = R,$$

where $\tilde{F}(w,n)$ is the distribution of wage and number of children in the population.

3. Optimal Income Taxation

In this section, we first focus on the situation where the government can only employ income tax-transfer so that $t = 0$ and $p = 1$ in the above model setting. The present model extends Saez (2002) extensive margin labor supply model to take into account variable family size and children-related expenses.

[5] Indicator function, $\mathbb{1}_{s\leq s^*(w,n)}$, means that the value is equal to one if domestic productivity parameter s is larger than s^* and zero otherwise.

The Lagrangian for the government's maximization problem can be written as

$$\mathcal{L} = \int\limits_{\boldsymbol{a}} \left[\beta(\boldsymbol{a})V(B(w,n);\boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)} + \beta(\boldsymbol{a})V(B(0,n);\boldsymbol{a})\mathbb{1}_{s>s^*(w,n)}\right] dF(\boldsymbol{a})$$
$$+\lambda\left\{\int\limits_{W,N}[T(w,n)-T(0,n)]D_{w,n}(s^*)d\tilde{F}(w,n) + \int\limits_{\boldsymbol{a}} T(0,n)dF(\boldsymbol{a}) - R\right\}, \quad (5)$$

and the first-order conditions with respect to $T(w,n)$ and $T(0,n)$ are, using $B(w,n) = w - T(w,n)$ and $B(0,n) = -T(0,n)$,

$$\int\limits_{S}\left[1-\frac{\beta(\boldsymbol{a})}{\lambda}V_B(B(w,n);\boldsymbol{a})\right]\mathbb{1}_{s\leq s^*(\cdot)}dF(s|w,n) = \hat{T}(\cdot)D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(w,n)}, \quad (6)$$

and

$$\int\limits_{W,S}\left[1-\frac{\beta(\boldsymbol{a})}{\lambda}V_B(B(0,n);\boldsymbol{a})\right]\mathbb{1}_{s>s^*(\cdot)}dF(w,s|n) = \int\limits_{W,S}\hat{T}(\cdot)D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(0,n)}dF(w,s|n), \quad (7)$$

where $\hat{T}(\cdot) = T(w,n) - T(0,n)$, is the tax liability excluding the lump-sum transfer, $f(s|w,n)$ is p.d.f of the domestic ability parameters conditional on the parent wage, w and on the number of children, n, and $f(w,s|n)$ is joint p.d.f of wage and the domestic ability parameters conditional on the number of children, and $F(s|w,n)$ and $F(w,s|n)$ are corresponding c.d.f respectively.

When obtaining (6) and (7), it is important to note that, because of the envelope theorem, the effect of an infinitesimal change in $B(w,n)$ has no first-order effect on welfare for households moving in or out of labor force, and therefore there is no need to take into account, in the first term of (6), the effect of a change of $B(w,n)$ on the distribution.

Similar to Saez (2002), we define the marginal social welfare weight for household working at wage w and with number of children, n, as

$$g(w,n) = \frac{1}{\lambda D_{w,n}(s^*)}\int\limits_{S}\beta(\boldsymbol{a})V_B(B(w,n);\boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)}dF(s|w,n). \quad (8)$$

This weight represents the money equivalent value for the government of distributing an extra money uniformly to households working at wage w and with dependent children, n. Using definition of (8), the first-order condition (6) can be rewritten as

$$[1-g(w,n)]D_{w,n}(s^*) = \hat{T}(w,n)D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(w,n)}. \quad (9)$$

The size of the behavioral responses is captured by the elasticity of participation with respect to the after-tax income. Formally, we define for all w, n, as

$$\eta(w,n) = \frac{B(w,n)}{D_{w,n}}\frac{\partial D_{w,n}}{\partial B(w,n)}. \quad (10)$$

The elasticity, $\eta(w,n)$, is positive and measures the percentage number of employed workers at wage w and with dependent children n who decide to leave the labor force when the disposable incomes in employment decreases by 1 percent. Then, we get the following result.

Proposition 1.

For all household working at wage, w, and with dependent children, n, the optimal income tax/transfer satisfies

$$\frac{\hat{T}(w,n)}{w-T(w,n)} = \frac{1-g(w,n)}{\eta(w,n)}. \tag{11}$$

Proof: Substituting (10) in (9) and rearranging it. ∎

This formula is a simple inverse elasticity tax rule that is similar to Saez (2002) and Choné and Laroque (2005) under a Rawlsian criterion. We have shown that inverse elasticity tax rule obtained in the extensive labor model continues to hold here within every family size class. As noted by Choné and Laroque (2005), generally, the first-order condition does not characterize the solution. However, when the problem is well behaved, given $B(0,n)$, (11) implicitly gives the optimal income tax schedule.

We first note that $w - T(w,n)$ cannot be negative because nobody would choose to work at income level w. Hence, (11) represents that difference of after tax income in labor status $T(w,n) - T(0,n)$ depends negatively on the term $\eta(w,n)$ so that if the elasticity of participation is relatively elastic, difference of after tax income became small. The difference of after tax income also depends on the size of $g(w,n)$. Turning to the structure of g, it is useful to define the marginal social welfare weight for household not working, as similar to $g(w,n)$,

$$g(0,n) = \frac{1}{\lambda\left[1-\int_{W,S} dF(w,s^*|n)\right]} \int_{W,S} \beta(\boldsymbol{a}) V_B(B(0,n);\boldsymbol{a}) \mathbb{1}_{s>s^*(w,n)} dF(w,s|n). \tag{12}$$

Using (12), we can get the following result.

Lemma 1.

When the government only set income tax, the average level of $g(\cdot,n)$ is, given n,

$$\int_{W,S} g(w,n)dF(w,s|n) = 1 - \int_{W,S} \hat{T}(w,n)D'_{w,n}(s^*)\left[\frac{\partial s^*(w,n)}{\partial B(w,n)} + \frac{\partial s^*(w,n)}{\partial B(0,n)}\right] dF(w,s|n). \tag{13}$$

Furthermore,

(i) if labor supply do not have income effect, the average level of $g(\cdot,n)$ is equal to one,

(ii) if parental time to child is a normal (inferior) good, the average level of $g(\cdot,n)$ is larger (smaller) than one.

Proof: See Appendix A.2. ∎

We may recall that $g(\cdot,n)$ represents government's preferences among various redistributive form. We first assume that the government has an redistributive tastes such that for each number of children $g(\cdot,n)$ decrease with w. From lemma 1, with no income effect, the average of marginal social welfare weight for household for each number of children is one, and if parental time is a normal good, it is larger than one.[6] In these cases, there is an wage w^* such that $g(w,n) \geq 1$ for $w \leq w^*$ and $g(w,n) < 1$ for $w > w^*$.

Hence, (11) and lemma 1 imply that $T(w,n) - T(0,n) > 0$ for $w > w^*$ and that $T(w,n) - T(0,n) \leq 0$ for $w \leq w^*$. When $w^* > 0$, the government provides a higher transfer to low skilled workers $-T(w,n)$ than to the unemployed $-T(0,n)$ even though the social marginal utility of consumption is highest for the unemployed. It means that the government should subsidize the wages of the poorest household in work force within every family size class in order to increase the size of transfers as income increases. In this sense, in-work credits can be optimal.

If household were differentiated by labor market ability only, as in conventional optimal taxation models, redistribution would come about for equity reasons, and go from high wage to low wage households. Although it is reasonable to assume that the government give higher social marginal weight for whom they have lower abilities, if household are differentiated also by skill on the children's quality or domestic production as in our present model, there can be an efficiency as well as an equity motive for redistributing. The key reason why there is an efficiency gain from redistribution is simply because complete specialization may achieve allocative efficiency. At the laissez-faire equilibrium, the budget constraint may prevent households who are comparatively better at raising children from specializing completely in child care. In the second best environment, the government can redistribute towards households who have a comparative advantage in child care so as to raise social welfare quite independently of equity considerations.

Hence, we need to reconsider how joint distribution pattern affects the optimal income tax schedule. We first consider the case in which domestic ability parameter and wage rate are negatively correlated. In this case, high w households tend to have low s ability (and vice versa). It is reasonable to assume that switching points s^* is increasing in w so that high w households may want to participate in labor force and high s households may want to specialize completely in domestic activities. Then the government intervention may improve efficiency. Since few households have both lower abilities, the government do not have incentives to stress equity consideration and is implied that the variance of $g(\cdot,n)$ may be smaller. It follows that the size of in-work credits becomes small.

We next consider the case in which domestic ability parameter and wage rate are positively correlated so that high w households tend to have high s ability. In this cases, there exists possibilities that households who have a relative high ability do not participate in labor force. Thus among the unemployed there are mixed high and low ability household. It means that the size of $g(0,n)$ is smaller. On the other hand, relative low wage households in labor force are given to higher weight. Therefore, the variance of $g(\cdot,n)$ is large. It follows that the size of wage subsidy is large.

Finally, we would now like to examine some systematic relationships between family size and income tax liability or wage subsidies (negative marginal tax rate). We mainly focus on the logic

[6] Saez (2002) showed that with no income effect the average of marginal social welfare weight for household is one.

of evidence that the wage subsidies increase with family size like U.S. EITC.

Rearranging (11), we get

$$T(w,n) = \frac{\eta(w,n)T(0,n) + [1-g(w,n)]w}{1-g(w,n)+\eta(w,n)}. \tag{14}$$

Differentiating (14) with respect to n, we obtain

$$\frac{\partial T(w,n)}{\partial n} = \frac{\{[1-g(\cdot)]\eta_n(\cdot) + g_n(\cdot)\eta(\cdot)\}\,[T(0,n)-w] + \eta(\cdot)[1-g(\cdot)+\eta(\cdot)]T_n(0,n)}{[1-g(\cdot)+\eta(\cdot)]^2}, \tag{15}$$

where the subscripts indicate derivatives. Equation (15) gives us information on the relationship between family size and income tax liabilities. If the R.H.S of (15) is smaller (larger) than 0, a child should be a tax asset (liability) for that household. It depends on the size of the behavioral response, η, the government's redistributive taste, g, lump-sum transfer, $T(0,n)$, and these derivatives respectively.

We first consider a situation that the government does not change lump-sum transfer with respect to the number of children (i.e., $T_n(0,n) = 0$). Since $T(0,n)$ is negative, $T(0,n) - w$ is also negative. Then the sign of the R.H.S of (15) is determined by the following relationship:

$$\frac{\partial T(w,n)}{\partial n} \gtreqless 0 \Leftrightarrow [1-g(w,n)]\eta_n(w,n) + \eta(w,n)g_n(w,n) \lesseqgtr 0. \tag{16}$$

Equation (16) implies that a child should be a tax asset when the government's redistributive taste is increasing with n, and the behavioral response is increasing with n unless $1-g$ is smaller than 0. We further assume that the government set in-work credits at the first. Then, $1-g(w,n)$ is smaller than 0. We can see whenever the government's redistributive taste are stronger (i.e., $g_n > -[1-g]\eta_n/\eta$), the wage subsidies increase with family size.

We next consider the case in which the lump-sum transfer is increasing with the number of children. Recall that the lump-sum transfer is $-T(0,n)$ so that it means $T_n(0,n) < 0$. In a real world, most developed countries have adopted that feature to achieve some horizontal equity to compensate families for children-related expenses. Then we can see the existence of this feature accelerate children as a tax asset. Indeed, we can show that $1-g+\eta > 0$.[7] In this setting, it seems reasonable to conclude that family size should be subsidized.

4. Optimal Mixed Taxation

In this section, we present a general model of public policy for child care. The government can not only employ income tax-transfer but also a tax on the child-specific commodities. We now examine the possible advantage of introducing indirect tax.

[7] We assume that the government redistributes the uniform guaranteed income level (lump-sum transfer) for all household. Then, $-T(0,n)$ and $w - T(0,n)$ are positive. Therefore, as $w > T(0,n)$, we have $w - T(w,n) > T(0,n) - T(w,n)$. This implies that $\hat{T}(w,n)/[w - T(w,n)] > -1$. Combining (11), it means that $[1 - g(w,n)]/\eta(w,n) > -1$. Therefore, we can get $1 - g(w,n) + \eta(w,n) > 0$.

4.1. Effective Tax Rate

The Lagrangian associated with this optimization problem is

$$\mathcal{L} = \int_{\boldsymbol{a}} \left[\beta(\boldsymbol{a})V(p, B(w,n); \boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)} + \beta(\boldsymbol{a})V(p, B(0,n); \boldsymbol{a})\mathbb{1}_{s>s^*(w,n)}\right] dF(\boldsymbol{a})$$

$$+ \lambda \left\{ \int_{W,N} [T(w,n) - T(0,n)]D_{w,n}(s^*)d\tilde{F}(w,n) + \int_{\boldsymbol{a}} T(0,n)dF(\boldsymbol{a}) \right.$$

$$\left. + t\int_{\boldsymbol{a}} \left[Z(p, B(w,n); \boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)} + Z(p, B(0,n); \boldsymbol{a})\mathbb{1}_{s>s^*(w,n)}\right] dF(\boldsymbol{a}) - R \right\}. \tag{17}$$

The first-order conditions with respect to $T(w,n)$ and $T(0,n)$ are, using $B(w,n) = w - T(w,n)$ and $B(0,n) = -T(0,n)$,

$$\left[\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))\right] D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(w,n)} + t\int_S \frac{\partial Z(\cdot)}{\partial B(w,n)}\mathbb{1}_{s\leq s^*(\cdot)}dF(s|w,n)$$

$$= \int_S \left[1 - \frac{\beta(\boldsymbol{a})}{\lambda}V_B(p, B(w,n); \boldsymbol{a})\right] \mathbb{1}_{s\leq s^*(w,n)}dF(s|w,n), \tag{18}$$

and

$$\int_{W,S} \left[\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))\right] D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(0,n)}dF(w,s|n) + t\int_{W,S} \frac{\partial Z(\cdot)}{\partial B(0,n)}\mathbb{1}_{s>s^*(\cdot)}dF(w,s|n)$$

$$= \int_{W,S} \left[1 - \frac{\beta(\boldsymbol{a})}{\lambda}V_B(p, B(0,n); \boldsymbol{a})\right] \mathbb{1}_{s>s^*(w,n)}dF(w,s|n), \tag{19}$$

where $\hat{T}(\cdot) = T(w,n) - T(0,n)$ and $\hat{Z}(s^*(\cdot)) = Z(p, B(w,n); w, n, s^*) - Z(p, B(w,0); w, n, s^*)$.[8] Similar to section 3, we define the marginal social welfare weight for household working at wage w and with number of children, n, as

$$g(w,n) = \frac{1}{\lambda D_{w,n}(s^*)}\int_S \beta(\boldsymbol{a})V_B(p, B(w,n); \boldsymbol{a})\mathbb{1}_{s\leq s^*(w,n)}dF(s|w,n). \tag{20}$$

This weight represents the money equivalent value for the government of distributing an extra money uniformly to households working at wage w and with dependent children, n.

Using definition of (20) and the elasticity (10), we can get following result.

Proposition 2.
For all household working at wage, w, and with dependent children, n, the optimal tax mix satisfies

$$\frac{\hat{T}(w,n) + t\left[\hat{Z}(s^*(w,n)) + \int_S \frac{\partial Z(\cdot)}{\partial D_{w,n}(s^*)}\mathbb{1}_{s\leq s^*(w,n)}dF(s|w,n)\right]}{w - T(w,n)} = \frac{1 - g(w,n)}{\eta(w,n)}. \tag{21}$$

[8] The same procedure in the previous section applies to this section. Therefore, when obtaining (18), there is no change in welfare due to the behavioral response.

Proof: See Appendix A.3. ∎

We characterize the optimal effective tax and transfer schedule from (21). Equation (21) shows that the inverse elasticity tax rule also holds in the mixed taxation model. Comparing (11) and (21), we can see that tax rules in the mixed taxation model are added on the difference of commodity tax liability between work and no-work and total change of the demand of child-specific commodities. In this sense, we can interpret (21) as optimal effective tax rule.[9] Then, in-work credit may be sub-optimal since the first term and second term in the L.H.S of (21) tend to offset whenever child-specific commodity tax rate is negative.

It must be noted that the commodity tax also affects the level of $g(w,n)$. Therefore, before turning to the possible usefulness and the sign of commodity tax, we shall consider the mean level of $g(w,n)$. Similar to (12), we can define the marginal social welfare weight for household not working, $g(0,n)$. Using the expression $g(0,n)$, the first-order condition (19) can be rewritten as

$$\begin{aligned}[1-g(0,n)]\left[1-\int\limits_{W,S} dF(w,s^*|n)\right] &= \int\limits_{W,S}[\hat{T}(w,n)+t\hat{Z}]D'_{w,n}(s^*)\frac{\partial s^*(w,n)}{\partial B(0,n)}dF(w,s|n)\\ &\quad + t\int\limits_{W,S}\frac{\partial Z(\cdot)}{\partial B(0,n)}\mathbb{1}_{s>s^*(w,n)}dF(w,s|n).\end{aligned} \tag{22}$$

Lemma 2.
In the tax mix model, mean g given n,

$$\begin{aligned}\int\limits_{W,S} g(w,n)dF(w,s|n) =& 1-\int\limits_{W,S}\hat{T}(w,n)D'_{w,n}(s^*)\left[\frac{\partial s^*(w,n)}{\partial B(w,n)}+\frac{\partial s^*(w,n)}{\partial B(0,n)}\right]dF(w,s|n)\\ & - t\int\limits_{W,S}\left\{\hat{Z}(s^*(w,n))D'_{w,n}(s^*)\left[\frac{\partial s^*(w,n)}{\partial B(w,n)}+\frac{\partial s^*(w,n)}{\partial B(0,n)}\right]\right.\\ &\left. + \frac{\partial Z(\cdot)}{\partial B(w,n)}\mathbb{1}_{s\le s^*(\cdot)}+\frac{\partial Z(\cdot)}{\partial B(0,n)}\mathbb{1}_{s>s^*(\cdot)}\right\}dF(w,s|n).\end{aligned} \tag{23}$$

Proof: Similar procedures in derivation of Lemma 1 apply. ∎

The difference between (13) and (23) can be illustrated as follows. Equation (23) takes into account of the commodity tax revenue change that results from disposal income changes. The commodity tax revenue change is composed of two parts. One is a change with labor status change (the second line of the R.H.S of (23)). Second is a income effect of child-specific commodities (the last line of the R.H.S of (23)). If commodity tax rate is negative and total tax revenue change is also negative, the R.H.S of (23) is larger than 1.

---

[9] Edwards, Keen, and Tuomala (1994) use the term 'effective' as the total tax paid at an income level in the form of both income and commodity taxation.

4.2. The Role of Commodity Taxation

Turning to the structure of second-best linear commodity tax, the first-order condition with respect to t is,

$$-\int_{\boldsymbol{a}} \frac{\beta(\boldsymbol{a})}{\lambda} V_p(\cdot) dF(\boldsymbol{a}) = \int_{\boldsymbol{a}} \left\{ [\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))] D'_{w,n}(s^*) \frac{\partial s^*(\cdot)}{\partial p} + \left[Z(\cdot) + t\frac{\partial Z(\cdot)}{\partial p} \right] \right\} dF(\boldsymbol{a}). \tag{24}$$

Then, we can get following result.

Proposition 3.

The structure of Pareto efficient linear commodity taxes satisfies,

$$t\int_{\boldsymbol{a}} \Psi(\boldsymbol{a}) dF(\boldsymbol{a}) = -\int_{\boldsymbol{a}} \hat{T}(w,n)\Gamma'(\boldsymbol{a}) dF(\boldsymbol{a}), \tag{25}$$

where

$$\Psi(\boldsymbol{a}) = \frac{\partial Z^c(p, B(w,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s \le s^*(\cdot)} + \frac{\partial Z^c(p, B(0,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s > s^*(\cdot)} + \hat{Z}(s^*(w,n))\Gamma'(\boldsymbol{a}), \tag{26}$$

$$\begin{aligned} \Gamma'(\boldsymbol{a}) = D'_{w,n}(s^*) \Bigg[& \frac{\partial s^*(w,n)}{\partial p} + \frac{\partial s^*(w,n)}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(w,n)} \\ & + \frac{\partial s^*(w,n)}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(w,n)} \Bigg], \end{aligned} \tag{27}$$

and $Z^c(\cdot)$ is compensated conditional demand function, as mentioned below.

Proof: See Appendix A.4. ∎

We first focus on the term $\Gamma'(\boldsymbol{a})$ (i.e., (27)). The first term in it represents the direct effect of child-specific commodity price for labor status choice. Turning to the second and third term in $\Gamma'(\boldsymbol{a})$, it is useful to define the compensated conditional demand function,

$$Z^c(p, L, u; \boldsymbol{a}) = \arg\min_{z} \{x + pZ | U(x, Q(z,h)) \ge u\},$$

and corresponding conditional expenditure function, $E(p, L, u; \boldsymbol{a})$. We note that

$$\frac{\partial E}{\partial p} = Z^c,$$

and

$$Z(p, E(p, L, u, w); \boldsymbol{a}) = Z^c(p, L, u; \boldsymbol{a}).$$

Therefore, second and third term in $\Gamma'(\boldsymbol{a})$ represent an indirect effect due to small commodity tax changes through the total expenditure change. As a result, the term $\Gamma'(\boldsymbol{a})$ (i.e., (27)) shows total effect for labor status choice results from commodity tax change.

Hence, Proposition 3 has a simple interpretation. It shows that second-best Pareto efficiency requires that the sum of the change in the aggregate compensated demand for child-specific commodities induced by the introduction of a small tax (holding her labor status and earnings constant) and change in aggregate demand for child-specific commodities according to labor status change which is induced by small tax change be the same amount by which income tax change induced by small tax change overall population.

This result enables a very direct account of the way in which differential linear commodity taxation may usefully supplement a non-linear income tax. The essential point in considering the optimal commodity tax structure is its participation effect. As mentioned above, parents have different unobservable ability to raise children. Although a parent whose wage is w can not mimic her wage to the other one's if she works, tax system affects her labor status. Income tax impose same tax liability for same earnings and cannot discriminate on the basis of ability to raise children. On the other hands, if the demand for child-specific commodity is related to her ability to raise children, commodity tax become an indirect tool to reveal it. In this sense, commodity tax has a role to improve efficiency.

Proposition 3 also gives us information on the sign of the child-specific commodity tax rate. Since the own-price Slutsky term is negative, the optimal tax rate on z, t, depends on the sign and relative size of Γ'. More precisely, if $\int_{\boldsymbol{a}} \hat{T}(w,n)\Gamma'(\boldsymbol{a})dF(\boldsymbol{a})$ and $\int_{\boldsymbol{a}} \Psi(\boldsymbol{a})dF(\boldsymbol{a})$ have same (opposite) sign, then the optimal commodity tax rate is negative (positive). The intuitive explanation is that, if an increase of child-specific commodity price decreases the proportion of parent in work force on average, a subsidy for that commodity helps to reveal her true ability and hence Pareto improving.

It is worthwhile examining a condition that $\int_{\boldsymbol{a}} \hat{T}(w,n)\Gamma'(\boldsymbol{a})dF(\boldsymbol{a})$ and $\int_{\boldsymbol{a}} \Psi(\boldsymbol{a})dF(\boldsymbol{a})$ have same sign. The R.H.S of (26) shows that the sign of $\Psi(\boldsymbol{a})$ depends on term $\Gamma'(\boldsymbol{a})$ since first and second term of it are negative. Thus if $\Gamma'(\boldsymbol{a})$ is negative on average, $\int_{\boldsymbol{a}} \hat{T}(w,n)\Gamma'(\boldsymbol{a})dF(\boldsymbol{a})$ and $\int_{\boldsymbol{a}} \Psi(\boldsymbol{a})dF(\boldsymbol{a})$ have same negative sign. As $\Gamma'(\boldsymbol{a})$ represent total effect for labor status choice results from commodity tax change, $\Gamma'(\boldsymbol{a}) < 0$ implies that parental time and child-specific commodity are substitutes in child's quality or domestic production function. That is not necessarily true if parental time and child-specific commodity are complements, in which case it may happen that $\int_{\boldsymbol{a}} \hat{T}(w,n)\Gamma'(\boldsymbol{a})dF(\boldsymbol{a}) > 0$ but $\int_{\boldsymbol{a}} \Psi(\boldsymbol{a})dF(\boldsymbol{a}) < 0$. Then commodity tax rate should be positive. In this sense, our result are consistent with Christiansen (1984) in intensive labor model.[10]

4.3. Separable Utility and relation to Atkinson-Stiglitz Theorem

From Atkinson and Stiglitz (1976) we know that in the context of the Mirrlees (1971) intensive labor model of income taxation with many consumption goods and in the presence of an optimal non-linear income tax, commodity taxation is useless when utility is weakly separable between leisure and consumption goods. Saez (2004) also showed that in the context of extensive labor model, the Atkinson-Stiglitz Theorem remains valid with imperfect substitution in labor types. More precisely, assumptions of weak separability and common sub-utility of consumption goods to all households imply that the tax on labor income is enough and that there is no need to tax commodities at the optimum. On the other hand, Balestrino, Cigno, and Pettini (2003) showed

[10] Christiansen (1984) showed that goods that are complementary with leisure should be taxed.

that in the context of intensive labor and household production model, when households differ in domestic as well as market ability, it is optimal to use indirect taxation alongside an optimally designed income tax even if the utility function is separable in commodities and non-labor time, as long as an input in domestic production and domestic skill parameter are uncorrelated.

The key point of obtaining different conclusion is whether or not the assumption of homogeneity preference for consumption goods is adopted in addition to weakly separability condition.[11] Let us see whether this argument is still true in our variable family size setting. If we assume that the utility function is weakly separable in x and Q, and the child's life time quality in h and z, parental time is weakly separable from commodities. Furthermore, we assume that child-specific commodity and parental ability to raise child as well as the number of children and each child quality are not correlated. Then, we can specify utility function as $U = U(\phi(x), nQ(\psi_1(z), \psi_2(h, s)))$. In this setting, we can show that Atkinson-Stiglitz Theorem is still valid. To see this, as in Christiansen (1984) and Saez (2004), starting from no commodity tax and optimal income tax, we consider the case in which the introduction of a small tax dt on child-specific commodity accompanied by reductions $dT(w, n) = -Z(p, B)dt$ in the income tax liabilities. Then, we can show the effects on tax revenue and welfare of this reform can be fully canceled out.

First note that this tax reform is well defined because the demand for child-specific commodity is the same for given disposal income and for any family size. This implied that uncorrelations between child-specific commodity and parental ability to raise child, and between family size and each child quality are key assumptions. Second, from Roy's identity, $dV = -V_B[Zdt + dT]$, this tax reform has no effect on household utility and hence on welfare. Third, any households who switch labor status because of one of the tax changes also switch labor status because of the other one. Therefore the behavioral responses to the tax reform are identical. Thus there is no effect on tax revenue due to behavioral responses. Last, the mechanical change in tax revenue also have no effect because of assumption $dT(w, n) = -Z(p, B)dt$. Therefore, the small commodity tax is fully equivalent to the small income tax change. We summarize this result as following:

Proposition 4.
In extensive labor model with heterogeneous family size, the Atkinson-Stiglitz Theorem remains valid with weakly separable preference and uncorrelations between child-specific commodity and parental abilities, and between family size and each child quality.

5. Conclusion

The purpose of this paper was to study optimal tax system with extensive margin and variable sizes of family. We derived the following results. First, assuming that the government is restricted to taxing income, the inverse elasticity tax rule obtained in the extensive labor model continues to hold within every family size class. Furthermore, the government should subsidize the wages of the poorest household in work force within every family size class in order to increase the size of transfers as income increases. The joint distribution of wage and skill on children's quality

[11] Notice that substituting the domestic production function into the utility function gives non-labor time a role analogous to that of leisure in the standard labor-leisure choice model.

affects the optimal income tax schedule. The negative correlation will lead to relative small size of in-work credit and vice versa. We also examine relationship between family size and income tax-transfer. It implies that in-work credit should become larger with the number of children, provided that government's redistributive taste, extensive labor response, and lump-sum transfer are increasing with the number of children.

Second, in case the government can use child-specific commodity tax, the inverse elasticity tax rule holds within every family size class, but effective tax term. Hence, in-work credit is not necessarily optimal depending on the size of tax/subsidy on the child-specific commodity. The differential commodity taxation may usefully supplement a non-linear income tax. The main reason is that if the demand for child-specific commodity is related to her ability to raise children, commodity tax become an indirect tool to reveal it. This argument is analogy to the role of commodity taxation with non-linear income tax in intensive labor model, in which commodity tax can play a role of relaxing the self-selection constraint which restricts ability to income tax.

Third, if we assume that weakly separable preference and uncorrelations between child-specific commodity and parental abilities and between family size and each child quality, we can say that the Atkinson-Stiglitz Theorem remains valid.

A. Appendix

A.1. First order conditions to the government's problem

The first-order condition with respect to $T(w,n)$ is, using $B(w,n) = w - T(w,n)$,

$$\begin{aligned}
&\int_S \frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(w,n)} dF(s|w,n) \tilde{f}(w,n) \\
&= \left[D_{w,n}(s^*) - \hat{T}(w,n) D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(w,n)} \right] \tilde{f}(w,n) \\
&\quad - t \left[\int_S \frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial B(w,n)} \mathbb{1}_{s \le s^*(\cdot)} dF(s|w,n) + \hat{Z}(s^*(w,n)) D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(w,n)} \right] \tilde{f}(w,n),
\end{aligned}$$

dividing both sides by $\tilde{f}(w,n)$ yields (18). Setting $t = 0$ (hence $p = 1$), we get (6).

The first-order condition with respect to $T(0,n)$ is,

$$\begin{aligned}
&\int_{W,S} \frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(w,n)} dF(w,s|n) \tilde{f}(n) \\
&\quad = \int_W \left[-D_{w,n}(s^*) - [\hat{T}(w,n) + t\hat{Z}(s^*(w,n))] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(0,n)} \right] d\tilde{F}(w,n) \\
&\qquad + \int_{W,S} d\tilde{F}(w,s|n) \tilde{f}(n) - t \int_{W,S} \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial B(0,n)} \mathbb{1}_{s > s^*(w,n)} dF(w,s|n) \tilde{f}(n),
\end{aligned}$$

where $f(w,s|n)$ is p.d.f. of wage and the domestic ability parameters conditional on the number of children, $F(w,s|n)$ is corresponding c.d.f., and $\tilde{f}(n)$ is marginal density function of the number of children in the population. Noting that $\tilde{f}(w,n) = \int_s f(w,s,n)$ and $f(w,s,n) = f(w,s|n)\tilde{f}(n)$,

$$\int\limits_{W,S} \frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s>s^*(w,n)} dF(w,s|n)\tilde{f}(n)$$
$$= \int\limits_{W,S} \left[-D_{w,n}(s^*) - [\hat{T}(w,n) + t\hat{Z}(s^*(w,n))] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(0,n)} \right] dF(w,s|n)\tilde{f}(n)$$
$$+ \int\limits_{W,S} d\tilde{F}(w,s|n)\tilde{f}(n) - t \int\limits_{W,S} \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial B(0,n)} \mathbb{1}_{s>s^*(w,n)} dF(w,s|n)\tilde{f}(n).$$

Dividing both sides by $\tilde{f}(n)$ yields

$$\int\limits_{W,S} \frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s>s^*(w,n)} dF(w,s|n)$$
$$= \int\limits_{W,S} \left[-D_{w,n}(s^*) - [\hat{T}(w,n) + t\hat{Z}(s^*(w,n))] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(0,n)} \right] dF(w,s|n) + 1$$
$$- t \int\limits_{W,S} \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial B(0,n)} \mathbb{1}_{s>s^*(w,n)} dF(w,s|n).$$

We next note that $\int_{W,S} D_{w,n}(s^*) dF(w,s|n)$ satisfy the following relationship;

$$\int\limits_{W,S} D_{w,n}(s^*) dF(w,s|n) = \iint\limits_{W,S} \left[\int_{s_0}^{s^*(w,n)} f(s|w,n) ds \right] \frac{f(w,s,n)ds}{\int f(w,s,n) dw ds} dw$$
$$= \int\limits_{W} \left[\frac{\int_{s_0}^{s^*(w,n)} f(w,s,n) ds}{\int f(w,s,n) ds} \right] \frac{\int f(w,s,n)ds}{\int f(w,s,n) dw ds} dw$$
$$= \frac{\int_W \int_{s_0}^{s^*(w,n)} f(w,s,n) dw ds}{\int_{W,S} f(w,s,n) dw ds} = \frac{\int_W \int_{s_0}^{s^*(w,n)} f(w,s,n) dw ds}{\tilde{f}(n)}$$
$$= \int\limits_{W,S} dF(w, s^*|n). \tag{A-1}$$

Using (A-1) yields (19). Setting $t = 0$ (hence $p = 1$), we get (7).

A.2. Proof of Lemma 1

Using definition of (12) and employing (A-1), (7) can be rewritten as

$$[1 - g(0,n)] \left[1 - \int\limits_{W,S} dF(w, s^*|n) \right] = \int\limits_{W,S} \hat{T}(w,n) D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(0,n)} dF(w,s|n). \tag{A-2}$$

Multiplying both sides of (9) by $f(w,s|n)$, and integrating with respect to w, s, we get

$$\int\limits_{W,S} [1 - g(w,n)] D_{w,n}(s^*) dF(w,s|n) = \int\limits_{W,S} \hat{T}(w,n) D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial B(w,n)} dF(w,s|n). \tag{A-3}$$

Combining (A-3) and (A-2) and using (A-1),

$$\int_{W,S} [1-g(w,n)]D_{w,n}(s^*)dF(w,s|n) + [1-g(0,n)]\left[1-\int_{W,S} dF(w,s^*|n)\right]$$
$$= \int_{W,S} \hat{T}(w,n)D'_{w,n}(s^*)\left[\frac{\partial s^*(w,n)}{\partial B(w,n)}+\frac{\partial s^*(w,n)}{\partial B(0,n)}\right]dF(w,s|n).$$

Hence, rearranging gives us

$$\int_{W,S} g(w,n)dF(w,s^*|n) + g(0,n)\left[1-\int_{W,S} dF(w,s^*|n)\right]$$
$$= 1-\int_{W} \hat{T}(w,n)D'_{w,n}(s^*)\left[\frac{\partial s^*(w,n)}{\partial B(w,n)}+\frac{\partial s^*(w,n)}{\partial B(0,n)}\right]dF(w,s|n). \qquad \text{(A-4)}$$

Note that $\left[1-\int_{W,S} dF(w,s^*|n)\right]$ represents the number of household not working when their children is n. The L.H.S of (A-4) is, therefore, population average of marginal social welfare weight given their children n.

We first assume that there are no income effects in labor supply. In this case, increasing both after tax income in work and no-work by a constant amount does not change the household labor participation decisions. Then we can write labor status switching point by $s^*(w,n) = s(B(w,n)-B(0,n))$. This implies that $\partial s^*/\partial B(w,n) + \partial s^*/\partial B(0,n) = 0$.

We next assume that parental time to child is a normal good. In this case, increasing both after tax income in work and no-work by a constant amount needs to decrease the proportion of household in work. Noting that $D_{w,n}$ is number of household choosing in work, this means $D'_{w,n}(s^*)[\partial s^*/\partial B(w,n) + \partial s^*/\partial B(0,n)] < 0$. In the inferior good case, it is vice versa.

A.3. Proof of Proposition 2

Using definition (20), the first-order condition (18) can be rewritten as

$$[1-g(\cdot)]D_{\cdot}(s^*) = \left[\hat{T}(\cdot)+t\hat{Z}(s^*(\cdot))\right]D'_{\cdot}(s^*)\frac{\partial s^*(\cdot)}{\partial B(\cdot)} + t\int_{S}\frac{\partial Z(p,B(\cdot);\boldsymbol{a})}{\partial B(\cdot)}\mathbb{1}_{s\le s^*(\cdot)}dF(s|\cdot). \qquad \text{(A-5)}$$

Multiplying both sides of (A-5) by $B(w,n)/D_{w,n}(s^*)$ and substituting $\eta(w,n)$ in it,

$$[1-g(\cdot)]B(\cdot) = \hat{T}(\cdot)\eta(\cdot) + t\left[\hat{Z}(s^*(\cdot))\eta(\cdot) + \int_{S}\frac{B(\cdot)}{D_{w,n}(s^*)}\frac{\partial Z(\cdot)}{\partial B(\cdot)}\mathbb{1}_{s\le s^*(\cdot)}dF(s|\cdot)\right],$$

and rearranging it,

$$\frac{1-g(\cdot)}{\eta(\cdot)} = \frac{1}{w-T(\cdot)}\left\{\hat{T}(\cdot)+t\left[\hat{Z}(s^*(\cdot)) + \int_{S}\frac{B(\cdot)}{\eta(\cdot)D_{w,n}(s^*)}\frac{\partial Z(\cdot)}{\partial B(\cdot)}\mathbb{1}_{s\le s^*(\cdot)}dF(s|\cdot)\right]\right\}.$$

Using the definition $\eta(w,n)$, we can get (21).

A.4. Proof of Proposition 3

Rearranging (24),

$$
\begin{aligned}
&-\int_{\boldsymbol{a}} \left[\frac{\beta(\boldsymbol{a})}{\lambda} V_p(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} + \frac{\beta(\boldsymbol{a})}{\lambda} V_p(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)}\right] dF(\boldsymbol{a}) \\
&\quad = \int_{\boldsymbol{a}} \Bigg\{ [\hat{T}(w,n) + t\hat{Z}(s^*(w,n))] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial p} \\
&\qquad + Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} + Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)} \\
&\qquad + t \left[\frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s \le s^*(\cdot)} + \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s > s^*(\cdot)}\right] \Bigg\} dF(\boldsymbol{a}). \qquad \text{(A-6)}
\end{aligned}
$$

Using Roy's identities, $Z = -V_p/V_B$, (A-6) is rewritten as,

$$
\begin{aligned}
&\int_{\boldsymbol{a}} \left\{ \left[\frac{\beta(\boldsymbol{a})}{\lambda} V_B - 1\right] Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} + \left[\frac{\beta(\boldsymbol{a})}{\lambda} V_B - 1\right] Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)} \right\} dF(\boldsymbol{a}) \\
&\quad = t \int_{\boldsymbol{a}} \left[\frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s \le s^*(w,n)} + \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s > s^*(w,n)}\right] dF(\boldsymbol{a}) \\
&\qquad + \int_{\boldsymbol{a}} \left[\hat{T}(w,n) + t\hat{Z}(s^*(w,n))\right] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial p} dF(\boldsymbol{a}).
\end{aligned}
$$

Rearranging it,

$$
\begin{aligned}
&t \int_{\boldsymbol{a}} \left[\frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s \le s^*(w,n)} + \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s > s^*(w,n)}\right] dF(\boldsymbol{a}) \\
&\quad = \int_{\boldsymbol{a}} \left[\frac{\beta(\boldsymbol{a})}{\lambda} V_B Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(w,n)} + \frac{\beta(\boldsymbol{a})}{\lambda} V_B Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(w,n)}\right] dF(\boldsymbol{a}) \\
&\qquad - \bar{Z} - \int_{\boldsymbol{a}} \left[\hat{T}(w,n) + t\hat{Z}(s^*(w,n))\right] D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial p} dF(\boldsymbol{a}), \qquad \text{(A-7)}
\end{aligned}
$$

where $\bar{Z}$ is the average level of Z in the overall population.

Multiplying both sides of (A-5) by $\int_s Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} dF(s|w,n)$ and both sides of (22) by $\int_{w,s} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)} dF(w,s|n)$, and integrating with respect to w, n,

$$
\begin{aligned}
&\int_{\boldsymbol{a}} Z(w,n) g(w,n) dF(\boldsymbol{a}) \\
&= \bar{Z} - \int_{\boldsymbol{a}} \left[\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))\right] A(\boldsymbol{a}) dF(\boldsymbol{a}) \\
&\quad - t \int_{\boldsymbol{a}} \left[\frac{\partial Z}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} + \frac{\partial Z}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)}\right] dF(\boldsymbol{a}),
\end{aligned}
$$

where

$$
A(\boldsymbol{a}) = D'_{w,n}(s^*) \left[\frac{\partial s^*(w,n)}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \le s^*(\cdot)} + \frac{\partial s^*(w,n)}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)}\right].
$$

Using definitions $g(\cdot)$,

$$\int_{\boldsymbol{a}} \left[\frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(w,n); \boldsymbol{a}) Z(w,n) \mathbb{1}_{s \leq s^*(\cdot)} + \frac{\beta(\boldsymbol{a})}{\lambda} V_B(p, B(0,n); \boldsymbol{a}) Z(w,n) \mathbb{1}_{s > s^*(\cdot)}\right] dF(\boldsymbol{a})$$
$$= \bar{Z} - \int_{\boldsymbol{a}} \left[\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))\right] A(\boldsymbol{a}) dF(\boldsymbol{a})$$
$$- t \int_{\boldsymbol{a}} \left[\frac{\partial Z}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \leq s^*(\cdot)} + \frac{\partial Z}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)}\right] dF(\boldsymbol{a}). \tag{A-8}$$

Substituting (A-8) in (A-7) and rearranging yields

$$t \int_{\boldsymbol{a}} \left[\frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s \leq s^*(w,n)} + \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial p} \mathbb{1}_{s > s^*(w,n)}\right] dF(\boldsymbol{a})$$
$$= -t \int_{\boldsymbol{a}} \left[\frac{\partial Z}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \leq s^*(\cdot)} + \frac{\partial Z}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)}\right] dF(\boldsymbol{a})$$
$$- \int_{\boldsymbol{a}} \left[\hat{T}(\cdot) + t\hat{Z}(s^*(\cdot))\right] \left[D'_{w,n}(s^*) \frac{\partial s^*(w,n)}{\partial p} + A(\boldsymbol{a})\right] dF(\boldsymbol{a}).$$

Further rearranging yields

$$t \int_{\boldsymbol{a}} \left\{ \left[\frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} + Z(p, B(w,n); \boldsymbol{a}) \frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial B(w,n)}\right] \mathbb{1}_{s \leq s^*(\cdot)} \right.$$
$$\left. + \left[\frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial p} + Z(p, B(0,n); \boldsymbol{a}) \frac{\partial Z(p, B(0,n); \boldsymbol{a})}{\partial B(0,n)}\right] \mathbb{1}_{s > s^*(\cdot)} + \hat{Z}(s^*(\cdot)) \Gamma'(\boldsymbol{a}) \right\} dF(\boldsymbol{a})$$
$$= - \int_{\boldsymbol{a}} \hat{T}(w,n) \Gamma'(\boldsymbol{a}) dF(\boldsymbol{a}),$$

where

$$\Gamma'(\boldsymbol{a}) = D'_{w,n}(s^*) \left[\frac{\partial s^*(w,n)}{\partial p} + \frac{\partial s^*(w,n)}{\partial B(w,n)} Z(p, B(w,n); \boldsymbol{a}) \mathbb{1}_{s \leq s^*(\cdot)} \right.$$
$$\left. + \frac{\partial s^*(w,n)}{\partial B(0,n)} Z(p, B(0,n); \boldsymbol{a}) \mathbb{1}_{s > s^*(\cdot)} \right].$$

Using the Slutsky equation,

$$\frac{\partial Z^c(p, B(w,n); \boldsymbol{a})}{\partial p} = \frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial p} + Z(p, B(w,n); \boldsymbol{a}) \frac{\partial Z(p, B(w,n); \boldsymbol{a})}{\partial B(w,n)},$$

where Z^c is compensated demand function, gives (25).

References

[1] Atkinson, A. B. and J. E. Stiglitz (1976) "The Design of Tax Structure: Direct versus Indirect Taxation", *Journal of Public Economics*, 6, 55-75.

[2] Balestrino, A., Cigno, A. and A. Pettini (2002) "Endogenous Fertility and the Design of Family Taxation", *International Tax and Public Finance*, 9, 175-193.

[3] Balestrino, A., Cigno, A. and A. Pettini (2003) “Doing Wonders with an Egg: Optimal Re-Distribution When Households Differ in Market and Non-Market Abilities”, *Journal of Public Economic Theory*, 5(3), 479-498.

[4] Becker, G. (1991) *A Treatise on the Family*, 2nd enlarged ed., Harvard University Press.

[5] Choné, P. and G. Laroque (2005) “Optimal Incentives for Labor Force Participation”, *Journal of Public Economics*, 89, 395-425.

[6] Christiansen, V. (1984) “Which Commodity Taxes Should Supplement the Income Tax”, *Journal of Public Economics*, 24, 195-220.

[7] Cigno, A. (1991) *Economics of the Family*, Clarendon Press and Oxford University Press.

[8] Cremer, H., Dellis, A. and P. Pestieau (2003) “Family Size and Optimal Income Taxation”, *Journal of Population Economics*, 16, 37-54.

[9] Diamond, P. (1980) “Income Taxation with Fixed Hours of Work”, *Journal of Public Economics*, 13, 101-110.

[10] Edwards, J., Keen, M. and M. Tuomala (1994) “Income Tax, Commodity Taxes and Public Good Provision: A Brief Guide”, *FinanzArchiv*, 51, 472-487.

[11] Laroque, G. (2005) “Income Maintenance and Labor Force Participation”, *Econometrica*, 73(2), 341-376.

[12] Meghir, C. and D. Phillips (2010) “Labour supply and taxes”, *Dimensions of Tax Design: The Mirrlees Review*, Ch.3, Oxford University Press.

[13] Mirrlees, J. A. (1971) “An Exploration in the Theory of Optimal Income Taxation”, *Review of Economic Studies*, 38, 175-208.

[14] Saez, E. (2001) “Using Elasticities to Derive Optimal Income Tax Rates”, *Review of Economic Studies*, 68(1), 205-229.

[15] Saez, E. (2002) “Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses”, *Quarterly Journal of Economics*, 117(3), 1039-1073.

[16] Saez, E. (2004) “Direct or indirect tax instruments for redistribution: short-run versus long-run” , *Journal of Public Economics*, 88, 503-518.

[17] Tuomala, M. (1984) “On the Optimal Income Taxation: Some Further Numerical Results”, *Journal of Public Economics*, 23, 351-366.