NBER WORKING PAPER SERIES

OPTIMAL INCOME, EDUCATION, AND BEQUEST TAXES IN AN INTERGENERATIONAL MODEL

Stefanie Stantcheva

Working Paper 21177
http://www.nber.org/papers/w21177

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2015, Revised September 2015

I want to thank James Poterba, Ivan Werning and Emmanuel Saez for comments and feedback. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Income, Education, and Bequest Taxes in an Intergenerational Model
Stefanie Stantcheva
NBER Working Paper No. 21177
May 2015, Revised September 2015
JEL No. H21,H24,H31,I24

ABSTRACT

This paper considers dynamic optimal income, education, and bequest taxes in a Barro-Becker dynastic setup. Parents can transfer resources to their children in two ways: First, through education investments, which have heterogeneous and stochastic returns for children, and, second, through financial bequests, which yield a safe, uniform return. Each generation's productivity and preferences are subject to idiosyncratic shocks. I derive optimal linear formulas for each tax, as functions of estimable sufficient statistics, robust to underlying heterogeneities in preferences, and at any given level of all other taxes. It is in general not optimal to make education expenses fully tax deductible and the optimal education subsidy, income tax and bequest tax can, but need not, move together at the optimum. I also show how to derive optimal formulas using "reform-specific elasticities" that can be targeted to empirical estimates from existing reforms. I extend the model to an OLG model with altruism to study the effects of credit constraints on optimal policies. Finally, I solve for the fully unrestricted policies and show that, if education is highly complementary to children's ability, it is optimal to distort parents' trade-off between education and bequests and to tax education investments relative to bequests.

Stefanie Stantcheva
Department of Economics
Littauer Center 232
Harvard University
Cambridge, MA 02138
and NBER
sstantcheva@fas.harvard.edu

1 Introduction

Investing in the education of their children is a key concern for many parents. From primary school to college, education expenses can be a major financial burden on families.[1] However, education is only one way through which parents can transfer resources to their children: financial bequests are another channel to invest in the future generation. Education investments and bequests can have different distributional incidences and efficiency implications.[2]

Parental decisions regarding education and bequests are naturally *jointly* affected by income and bequest taxes or education subsidies.[3] Bequest taxes affect the choice between transferring resources through education purchases or through financial bequests. Income taxes confiscate part of the next generation's returns to education, but also redistribute resources towards low income parents, which facilitates their education investments. Education in turn directly impacts earnings and, hence, the income and bequest tax bases. Finally, bequests affect the incentives to work and, hence, the revenues from income taxes. Despite these interactions, most of the optimal tax literature treats parental education choices as delinked from bequest decisions.[4]

This paper jointly studies the optimal income tax, bequest tax, and education subsidy in a dynamic intergenerational model *à la* Barro-Becker. In the model, presented in Section 2, each generation lives for one period and cares about the expected discounted utility of the next generation. Parents can transfer resources to their children in two ways: by purchasing education for them and by leaving them bequests. Bequests yield a safe and uniform return. On the contrary, parental investments in education yield a risky return for their children and are subject to an idiosyncratic and persistent shock, called "ability." More precisely, the wage of each individual is a function of his parents' endogenous education investment and the stochastic ability.[5] I also allow for idiosyncratic preference shocks for each generation that affect its taste for work relative to consumption. The government's objective is to maximize the expected utility of all dynasties from the point of view of the current generation.

The first goal of this paper is to derive a simple optimal formula for the education subsidy in terms of estimable statistics that are robust to heterogeneity in preferences and primitives, and to do so for any (not necessarily optimally set) labor and bequest taxes. Accordingly, in Sections 2–5, taxes

[1]Indeed, parents typically bear the full burden of the bill for primary and secondary schooling. Beyond this, parents covered around 40% of total college expenses in 2012 – down from 50% before the financial crisis (Sallie Mae, 2012). Around 60% of students receive some help from their parents for college (Hader and McGarry, 2012).

[2]Nordblom and Ohlsson (2011) document the correlation between bequests and education transfers in Sweden.

[3]In their study of education choices in the kibbutzim, Abramitzky and Lavy (2014) show that investments in schooling by parents are quite responsive to redistributive policies.

[4]Some papers do endogenize either the education investments or the bequest transfers and are reviewed in detail in Section 1.1 below.

[5]As in the standard Mirrlees (1971) income taxation model, ability is a comprehensive measure of the exogenous component of productivity and can capture, for instance, labor market or health shocks.

and subsidies are restricted to be linear and history-independent. It has been shown that in models without human capital, income tax formulas can typically be reformulated in terms of behavioral elasticities that capture the efficiency costs of taxation and distributional parameters that capture the redistributive value of taxation (Piketty and Saez, 2013a,b) and the same applies here to the dynamic intergenerational education subsidy. To build the intuition, I first briefly consider the optimal policies in a simplified one generation model in Section 3, before moving to the optimal policies in the full-fledged intergenerational model in Section 4.

The dynamic intergenerational formula retains the spirit of usual static optimal tax formulas, but with the behavioral elasticities appropriately redefined to capture the long-term expected discounted effects on each tax base. The long-run behavioral elasticities and cross-elasticities are functions of cumulatve substitution and income effects for both parents and children. However, to evaluate optimal taxes and reforms, it is not necessary to decompose those elasticities into primitives. In a similar spirit, I also derive formulas for the optimal income and bequest taxes. Because I obtain the optimal formulas for each tax for any given level of the other taxes, these formulas can be used in particular to evaluate reforms around the (potentially suboptimal) status quo.

Crucial determinants of education subsidies and bequest taxes are their distributional incidences, i.e., how concentrated education expenses and bequests are among high marginal utility agents. Since redistributive concerns are isolated in the redistributive factors, different social objectives regarding education and bequest policies can easily be incorporated without having to rederive optimal formulas.[6]

The second goal is to determine how the tax system should account for education investments and bequests. Should parents' education expenses for their children be tax deductible? A strong result by Bovenberg and Jacobs (2005) states that income taxes and education subsidies are "Siamese Twins" i.e., that they should be set equal to each other, which is equivalent to making education expenses fully tax deductible. Even in the one-generation version of the model, this result does not always hold and it is not optimal to make education expenses fully tax deductible, unless the relative efficiency cost and the relative distributional effect of education subsidies and income taxes are equal.

In even starker contrast to the "Siamese Twins" result, it need not even be the case that the education subsidy and the income tax move together at the optimum. Indeed, given that each optimal tax or subsidy is derived without assuming that the other taxes are optimally set, the formulas contain "fiscal externality" terms that account for the effect of one tax on all other tax bases. In the full-fledged intergenerational model, bequest taxes can also move positively or negatively with education subsidies

[6] In particular, since the optimal formulas are expressed in terms of social marginal welfare weights, those standard weights can be replaced by the generalized welfare weights of Saez and Stantcheva (2015) to capture alternative social criteria and fairness and justice principles. I illustrate this for several different social criteria.

depending on whether bequests and education transfers are overall substitutes or complements for parents.

In Section 5, I consider how credit constraints affect optimal education subsidies. To do so, I augment the model to an overlapping generations model, in which each generation lives for three periods: young, adult, and old. Agents have to invest in the education of their children during their adult period, but face credit constraints. Old agents leave bequests. Stronger credit constraints will tend to increase optimal education subsidies, decrease optimal income taxes, and leave bequest taxes unaffected.

The third purpose of the paper is methodological. Once one steps into the realm of dynamic intergenerational models, the empirical burden of estimating the relevant elasticities to tax policy increases. Are there alternative reformulations of the optimum that can specifically be targeted to estimates that are easier to obtain? I show how to derive optimal formulas in terms of "reform-specific elasticities" that capture the same trade-off between the efficiency and distributional impacts of each instrument, but that are different for each reform under consideration.[7] In this sense, the theory can adapt to the data we might already have (or could easily make) available.

In Section 6, I then contrast the restricted linear policies to a fully unrestricted tax system using a dynamic mechanism design approach as in Farhi and Werning (2013b) and Stantcheva (2012). This approach shows us the best allocations that a government could hope to achieve and the distortions that would remain despite having the most sophisticated tax instruments available. I show that at the optimum parental choices between education and bequests are typically distorted for redistributive and insurance purposes. This is true unless the wage of children as a function of education inputs and stochastic ability has a multiplicatively separable form. Indeed, while bequests yield the same risk-free return for all types of children, education is risky and its returns interact with children's unobserved ability, which in turn affects incentive constraints. If education is highly complementary to ability, in the sense that high ability children benefit more in proportional terms from their parents' education investments, then the return to education investments will be reduced below that on bequests. Put differently, education investments by parents will be taxed relative to bequests.

1.1 Related literature

This paper contributes to the optimal income taxation literature since Mirrlees (1971) and to its linear counterpart (Sheshinski, 1972). Most closely related are the papers by Saez (2001) and Piketty and Saez (2013a,b) that explore the expression of tax formulas in terms of sufficient statistics and

[7] This is related to the "policy elasticities" in Hendren (2014).

estimable elasticities. Findeisen and Sachs (2015) study income and capital taxation with restricted tax tools. The most complete analysis to date of income and capital taxes in a dynamic framework is the new contribution by Golosov, Tsyvinski, and Werquin (2014), who highlight the different welfare and revenue effects of different tax instruments. Werquin (2014) studies optimal taxation when there are labor supply frictions.

In addition, this paper adds to the long-standing literature on human capital formation as developed since Heckman (1976a) and Heckman (1976b) and that has been pushed forward in recent years by Heckman, Lochner, and Todd (2006), Cunha, Heckman, and Navarro (2005), Cunha and Heckman (2007), Cunha and Heckman (2008), Cunha, Heckman, Lochner, and Masterov (2006), and Heckman, Lochner, and Taber (1998).

In their closely related paper, Piketty and Saez (2013b) focus on bequest taxation in an intergenerational model, and highlight that with uncertainty and distributional concerns, the optimal bequest tax is generically different from zero. Farhi and Werning (2010) study nonlinear bequest and income taxation in a dynamic Mirrleesian framework, and find that a progressive bequest subsidy is optimal. Instead, with more general altruistic preferences, either taxes or subsidies could be optimal (Farhi and Werning, 2013a). In this paper, I consider an additional way of transferring resources to the next generation, namely education investments by parents.

A series of papers have considered optimal taxation jointly with education subsidies and using simpler policy tools. The benchmark result in a static model, which will be one of the focal points of this paper, is by Bovenberg and Jacobs (2005) who find that education subsidies and income taxes are "Siamese Twins," and should be set equal to each other, which is equivalent to making education expenses fully tax deductible. Bovenberg and Jacobs (2011) also study education subsidies and taxation with a more general earnings function. Jacobs (2007) considers in addition the general equilibrium effects from human capital, which I abstract from in this paper. Benabou (2002) jointly analyses taxes and education in a dynastic Ramsey model. Findeisen and Sachs (2012) consider the role of income-contingent loans and income taxes.

Several papers explore dynamic intergenerational models with parental investments in human capital from a computational and quantitative perspective and complement the theoretical analysis in this paper. Krueger and Ludwig (2013, 2014) study an overlapping generations general equilibrium model, in which education investments occur before agents enter the labor market, there are borrowing constraints, and parents transmit both bequests and ability to their children.[8] The formulas I obtain for optimal policies could fruitfully be calibrated from the quantitative estimates in those papers.[9]

[8]The paper is also related to Kindermann and Krueger (2014), but endogeneizes education investments.

[9]For instance, we could obtain the long-run elasticities needed in the formulas. Note that I do not need to impose

Finally, the role of credit constraints for educational investments has also been explored in the literature (Jacobs and Yang, 2014; Jacobs, Schindler, and Yang, 2011; Lochner and Monge-Naranjo, 2011, 2012). In Section 5, I derive optimal formulas in the presence of credit constraints in the dynamic intergenerational model.

2 An Intergenerational Model of Human Capital Investment and Bequests

2.1 Preferences and Dynastic Utility

This section starts with the setup of an intergenerational model of human capital investment. The economy consists of agents who live for one period. Agents are born, have one single child each, and then die. Total population is hence constant and normalized to 1, so that average per capita and aggregate variables are the same. Denote the agent from dynasty i at generation t by ti.

Parents are the ones who purchase education for their children. Parents in generation t can buy an education amount s_{t+1} for their child of generation $t+1$ at a linear cost s_{t+1}. In turn, generation t also receives the human capital that their own parents from generation $t-1$ purchased for them. Human capital completely depreciates between generations. The first generation of dynasty i at time 1 has an exogenously given distribution of human capital s_{1i}. This setup mirrors the fact that most investments in human capital occur before and during college and that parents account for a large share of these expenses.

Agents receive their human capital from their parents before they start working and consuming. The wage rate w_{ti} of any agent is determined by his stock of human capital and his stochastic ability θ_{ti}:

$$w_{ti}(s) \equiv w(s, \theta_{ti})$$

w is strictly increasing in education and ability and concave in education. Ability θ_{ti} is drawn from a stationary, ergodic distribution that allows for correlation between generations. Ability to earn income can be stochastic for several reasons, among which are health shocks, individual labor market idiosyncrasies or luck. Parents know the process for θ which means that they can form expectations about the ability of their children. If the process for ability is highly persistent across generations, parents have a very good advance information about their children's abilities, but, unless there is perfect persistence, still face some uncertainty regarding their children's ability realizations at the time when they are making education investment decisions.

restrictions on the heterogeneity in the formulas obtained and they are, in that sense, not model specific.

If an agent works $l_{ti} \geq 0$ hours at a wage rate w_{ti}, he earns gross income $y_{ti} = w_{ti} l_{ti}$. His utility is given by:

$$u_{ti}(c, y, s) \equiv u\left(c, \frac{y}{w(s, \theta_{ti})}; \eta_{ti}\right)$$

where c is consumption and η_{ti} is an idiosyncratic preference shock. u is increasing in consumption c and decreasing in labor effort y/w.

In addition to financing their education, parents can also leave financial bequests to their children. Bequests left by generation t are denoted by b_{t+1i} and earn a generational gross rate of interest R. Thus, generation t inherits a pre-tax bequest of Rb_{ti} from their parents. The initial generation 1 has an exogenously given distribution of bequests b_{1i}.

The expected utility of a dynasty i as of generation 1, U_{1i}, discounted by a generational discount factor β, is given by:

$$U_{1i} = E\left(\sum_{t=1}^{\infty} \beta^{t-1} u_{ti}(c_{ti}, y_{ti}, s_{ti})\right) \tag{1}$$

where the expectation operator represents the cross-sectional expectation over all dynasties, i.e., for any variable x: $E(x_{ti}) \equiv \int_i x_{ti} di$. This notation is a shortcut for the expectation over all realizations of θ and η for all i. Rewritten recursively:

$$U_{ti} = u_{ti} + \beta U_{t+1i}$$

2.2 Taxes and Budget Constraints

Government policies consist of a linear labor income tax τ_{Lt}, a linear human capital subsidy τ_{St}, a linear tax on the capitalized bequests τ_{Bt}, and a lump-sum demogrant G_t. Hence, the budget constraint of person ti is:

$$c_{ti} + b_{t+1i} + (1 - \tau_{St}) s_{t+1i} = Rb_{ti}(1 - \tau_{Bt}) + w_{ti}(s_{ti}) l_{ti}(1 - \tau_{Lt}) + G_t$$

Note here that each generation only lives for one period, so that parental bequests and education investments deliver their income at the same moment in life of the beneficiaries (their children). In a more realistic overlapping generations model with multiple periods of life, this would not be the case; bequests will in general be received later in the beneficiary's life than education. This would not matter if there were no credit constraints or if there was no progressive realization of the uncertainty about children's outcomes. Parents have consistent preferences with their children and would make the optimal trade-off between education and bequests without delay. When there is uncertainty in the child's ability and earning potential, being able to decide on bequests later in life, when the earning potential has been revealed, might make a difference. Indeed, parents may then try to target bequests to children who have received low ability and low income shocks so as to smooth consumption and provide insurance across generations. Section 5 considers an overlapping generations model with three periods life for each generation and credit constraints.

2.3 Equilibrium

Aggregate consumption, human capital investments, bequests received, and output are denoted respectively by c_t, s_{t+1}, b_t and y_t for generation t. I assume that the stochastic processes for θ and η are ergodic, so that, at constant policies (i.e., constant linear tax rates and demogrant), there is a unique ergodic steady-state equilibrium which is independent of the initial distribution of bequests and human capital (see also Piketty and Saez, 2013b). Hence, if tax policies $(\tau_{Lt}, \tau_{St}, \tau_{Bt}, G_t)$ converge to constant levels $(\tau_L, \tau_S, \tau_B, G)$ in the long run, then human capital s_{t+1}, output y_t, and bequests b_t also converge to their steady state levels and depend on the steady state tax policies.

It is assumed that the government sets the policies so as to satisfy a period-by-period budget constraint:[10]

$$G_t = \tau_{Lt} y_t + \tau_B R b_t - \tau_{St} s_{t+1} \tag{2}$$

3 Optimal Static Taxes and Subsidies

To build the intuition, let us start from a simple one-period model, and abstract from any intergenerational considerations. Each agent lives for one period, invests in his own human capital, and the distribution of shocks η and θ is iid over time. This leads to a sequence of identical static problems. In this case, preferences are simply given by $U_i = u_i(c_i, y_i, s_i)$ and the budget constraint is:

$$c_i + (1 - \tau_S) s_i = w_i(s_i) l_i (1 - \tau_L) + G$$

Social welfare is a weighted sum of individual utilities, with ω_i the Pareto weight on individual i:

$$SWF = \int_i \omega_i u_i(c_i, y_i, s_i) \, di$$

The government needs to satisfy his single period budget constraint:

$$G = \tau_L y - \tau_S s$$

where y and s are aggregate output and aggregate human capital.

The optimal labor tax in the presence of human capital: Suppose that the government sets the optimal income tax rate τ_L taking the education subsidy as given. Denote the individual elasticity

[10] Piketty and Saez (2013b) show that, if debt were allowed and the government was optimizing the economy-wide capital accumulation, then a modified golden rule would hold, with $\beta R = 1$ and the optimal formulas would be unaffected. Despite the fact that their model does not contain human capital, their result carries over unchanged.

of output to the net of tax rate $1-\tau_L$ by $\varepsilon_{yi} \equiv d\log(y_i)/d\log(1-\tau_L)$, the individual elasticity of education to the net of tax rate by $\varepsilon^y_{si} \equiv d\log(s_i)/d\log(1-\tau_L)$ and the aggregate weighted elasticities of, respectively, output and human capital to the net of tax rate by:

$$\varepsilon_Y \equiv \frac{d\log y}{d\log(1-\tau_L)} = \int_i \varepsilon_{yi}\frac{y_i}{y}di$$

$$\varepsilon^Y_S \equiv \frac{d\log s}{d\log(1-\tau_L)} = \int_i \varepsilon^y_{si}\frac{s_i}{s}di$$

Using the individuals' optimization and the envelope condition, a straightforward maximization implies that the change in welfare from a change in the linear tax rate $d\tau_L$ is given by:

$$\left(\int_i \omega_i u_{c,i}di\right)\left(y + \frac{\tau_L}{1-\tau_L}\int_i \varepsilon_{yi}y_i - \frac{\tau_S}{1-\tau_L}\int_i \varepsilon^y_{si}s_i\right)d\tau_L - \int_i \omega_i u_{c,i}y_i di d\tau_L = 0$$

Define the distributional characteristic of income to be:

$$\bar{y} \equiv \frac{\int_i \omega_i u_{c,i}y_i di}{y\int_i \omega_i u_{c,i}di}$$

$\bar{y} < 1$ since y_i is typically lower for those with high marginal social welfare weights (i.e., high marginal utilities of consumption). The following proposition gives the optimal static income tax for any given human capital subsidy.

Proposition 1 *The optimal income tax at any subsidy τ_S is:*

$$\tau^*_L = \frac{1-\bar{y}-\tau_S\frac{s}{y}\varepsilon^Y_S}{1-\bar{y}+\varepsilon_Y}$$

The expression for τ^*_L captures the typical trade-off between redistribution as measured by $\bar{y}$ and efficiency as measured by ε_Y. In addition, if education choices respond to the income tax, there is a type of fiscal externality to the education subsidy base from the income tax, which appears in the numerator term $\tau_S\frac{s}{y}\varepsilon^Y_S$. It is natural to assume that education choices respond negatively to the income tax ($\varepsilon^Y_S > 0$), because the income tax captures part of the return to human capital. If there is a pre-existing positive subsidy on education, $\tau_S > 0$, then, the income tax is naturally lower than if education choices were insensitive to income taxes and the more so the higher τ_S. This is because a higher education subsidy is a measure of how much incentives are provided for education and it is costly to counteract them with the distortive income tax. If the education subsidy were zero, there would be no such fiscal externality, and the income tax would be set according to a more standard formula (e.g. as in Piketty and Saez (2013a)).

The optimal education subsidy at any given income tax: We can symmetrically derive the optimal education subsidy. Denote the individual elasticity of human capital to the subsidy by $\varepsilon_{si} \equiv$

$\frac{ds_i}{d(\tau_S-1)}\frac{(\tau_S-1)}{s_i}$, the individual elasticity of income to the education subsidy by $\varepsilon^s_{yi} \equiv \frac{dy_i}{d(\tau_S-1)}\frac{(\tau_S-1)}{y_i}$ and the aggregate weighted elasticities by:

$$\begin{aligned} \varepsilon_S &\equiv \frac{ds}{d(\tau_S-1)}\frac{(\tau_S-1)}{s} = \int_i \varepsilon_{si}\frac{s_i}{s}di \\ \varepsilon^S_Y &\equiv \frac{dy}{d(\tau_S-1)}\frac{(\tau_S-1)}{y} = \int_i \varepsilon^s_{yi}\frac{y_i}{y}di \end{aligned}$$

Note that for a subsidy τ_S smaller than 1, we have $\varepsilon_{si} < 0$ since $ds_i/d\tau_s > 0$. Similarly to the definition of $\bar{y}$ above, define the distributional characteristic of education to be:[11]

$$\bar{s} \equiv \frac{\int_i \omega_i u_{c,i} s_i di}{s\int_i \omega_i u_{c,i} di}$$

The higher $\bar{s}$ and the more education is concentrated among high social welfare weight agents. In the standard utilitarian framework, $\omega_i u_{c,i} = u_{c,i}$ is just the marginal utility of income, so that higher consumption agents have lower social marginal welfare weights. Hence, if education is concentrated among high income agents, then $\bar{s}$ is small.[12] Again, the change in welfare from a change in the linear education subsidy $d\tau_S$ is given by:

$$\left(\int_i \omega_i u_{c,i} di\right)\left(-s - \frac{\tau_S}{\tau_S - 1}\int_i \varepsilon_{si}s_i + \tau_L \int_i \varepsilon^s_{yi} y_i\right)d\tau_S + \int_i \omega_i u_{c,i} s_i di d\tau_S = 0$$

This leads to the following proposition:

Proposition 2 *The static optimal human capital subsidy for a given labor tax τ_L is given by:*

$$\tau^*_S = \frac{1 - \bar{s} + \frac{y}{s}\varepsilon^S_Y \tau_L}{1 - \bar{s} + \varepsilon_S} \tag{3}$$

The optimal subsidy will in general not be zero because of the redistributive effect of education $1 - \bar{s}$ and the (finite) elasticity ε_S. The income tax appears in the numerator because of the fiscal spillover: if output responds positively to education subsidies, then a higher education subsidy has an additional positive effect on revenues raised, which is stronger the higher the income tax rate is.[13]

Full Optimum: optimizing jointly the income tax and education subsidy. At the full optimum, with both τ_L and τ_S optimally set, the labor and human capital taxes are given by:

$$\tau^*_S = \frac{(1 - \bar{s})(1 - \bar{y} + \varepsilon_Y) + \frac{y}{s}\varepsilon^S_Y (1 - \bar{y})}{\left[(1 - \bar{s} + \varepsilon_S)(1 - \bar{y} + \varepsilon_Y) + \varepsilon^S_Y \varepsilon^Y_S\right]}$$

[11] This distributional impact of education is very related to the distributional weights in Feldstein (1972).

[12] As Saez and Stantcheva (2015) show for the optimal income tax rate, the same formula derived below will hold if one replaces $\omega_i u_{c,i}$ by generalized social welfare weights. See the discussion in Section 4.

[13] Note that because τ_S is defined as a subsidy and because the elasticity is defined with respect to $\tau_S - 1$, which is negative for $\tau_S < 1$, the denominator $1 - \bar{s} + \varepsilon_S$ is typically negative. If output responds positively to education subsidies, then $\varepsilon^S_Y < 0$.

$$\tau_L^* = \frac{(1-\bar{y})(1-\bar{s}+\varepsilon_S) - \frac{s}{y}\varepsilon_S^Y(1-\bar{s})}{\left[(1-\bar{s}+\varepsilon_S)(1-\bar{y}+\varepsilon_Y)+\varepsilon_Y^S\varepsilon_S^Y\right]}$$

When both income taxes and education subsidies are optimized, the fiscal externalities are also perfectly internalized. The full optimum is discussed next in relation to the Bovenberg and Jacobs (2005) "Siamese Twins" result, stating that the linear income tax and the linear education subsidies should be set equal to each other.

3.1 The static "Siamese Twins" result revisited

A natural benchmark is the full deductibility of education expenses, i.e., $\tau_S^* = \tau_L^*$.[14] Bovenberg and Jacobs (2005) find that full deductibility of education expenses is optimal with a special form of the earnings function that guarantees that all agents benefit equally at the margin, in proportional terms, from human capital investments. In this generalized setup, where both the wage and the utility function are unrestricted, we can infer a similar result but based on the estimable elasticities and redistributive effects:

Corollary 1 *The education subsidy should optimally be set equal to the income tax rate, i.e., there should be full deductibility of education expenses, if and only if:*

$$\frac{\left(\frac{y}{s}\varepsilon_Y^S - \varepsilon_S\right)}{\left(\frac{s}{y}\varepsilon_S^Y + \varepsilon_Y\right)} = \frac{(1-\bar{s})}{(1-\bar{y})} \tag{4}$$

for $1-\bar{s} \neq 0$ *and* $1-\bar{y} \neq 0$.[15]

The left hand side in expression (4) is the ratio of the efficiency cost of the education subsidy and the efficiency cost of the income tax. The right hand side, is the ratio of their redistributive effects. The optimal education subsidy should be set equal to the optimal income tax if and only if the relative efficiency cost is equal to the relative redistributive effect. On the other hand, if the redistributive effect of education is disproportionately large relative to its efficiency cost, then it will be optimal to set $\tau_S^* > \tau_L^*$ and to subsidize education expenses beyond just making them tax deductible.

It is easy to check that the Bovenberg and Jacobs (2005) setting with a multiplicatively separable wage $w = \theta s$, isoelastic, separable, and quasilinear utility $u_i(c,y,s) = c - \frac{1}{\gamma}\left(\frac{y}{\theta_i s}\right)^\gamma$ implies that for any welfare weights, $\bar{y} = \bar{s}$, $\varepsilon_Y^S = \gamma$, $\varepsilon_Y = 1-\gamma$, $\varepsilon_S^Y = -\gamma$, $\varepsilon_S = \gamma - 1$, so that the equality (4) indeed holds.

Bovenberg and Jacobs (2011) show that the relative redistributive effects of education and output are determined, among others, by how complementary ability θ and human capital are in the wage

[14] Setting $\tau_S^* = \tau_L^*$ is equivalent to setting a subsidy of zero, but only taxing the agent based on his income minus education expenses, i.e., $wl - s$.

[15] If $1 = \bar{s}$ and $1 = \bar{y}$, then $\tau_S^* = \tau_L^* = 0$, a degenerate case.

function. If the marginal wage benefit of human capital is proportionately higher for higher ability agents, then $(1 - \bar{s})$ will be large relative to $(1 - \bar{y})$ and $\tau_S^* < \tau_L^*$, so that education expenses will be only partially tax deductible.[16]

More generally, not even a "relaxed" version of the Siamese Twins result must hold: it is not necessarily the case that the labor tax and the education subsidy should optimally move together. The effect of the education subsidy on output is the result of two effects. First, the subsidy increases the wage through an increase in education investments, which entails income and substitution effects on labor supply. Second, there is the income effect from the increased subsidy. Because the utility function is fully general, the net effect is not unambiguous. If the utility function is separable in consumption and labor, then all income effects on labor supply are absent and output is increasing in the human capital subsidy (so that $\varepsilon_Y^S < 0$, for any $\tau_S < 1$). From expression (3), hence, we see that the optimal subsidy would be increasing in the labor tax rate τ_L. Indeed, a higher subsidy stimulates output and allows to raise more revenues from the income tax base the higher τ_L is. However, if income effects are very strong, the opposite effect can occur.

3.2 Unobservable Education Investments

What if parental education expenses are not observable by the government or if they can be misreported? In this case, the government faces the constraint $\tau_S \equiv 0$. The optimal income tax will have to be adjusted differently, so as to compensate for the lack of a tool to directly control education investments. The optimal linear tax $\tau_L^{*,u}$ with unobservable education is given by:

$$\tau_L^{*,u} = \frac{1 - \bar{y}}{1 - \bar{y} + \varepsilon_Y}$$

The difference to the observable education case in formula (3) is the absence of the term $-\tau_S \frac{s}{y} \varepsilon_S^Y$ in the numerator. Suppose that the substitution effect dominates so that education responds negatively to income taxes ($\varepsilon_S^Y < 0$) and that with observable education, we would have $\tau_S^* > 0$. Then, the optimal income tax needs to be set lower when education is not observable. This is intuitive: with observable education, the government would like to subsidize education. Yet, if the subsidy on education is constrained to be zero, incentives for parental education investments can only be indirectly provided through a lower income tax.

Note that the elasticity of income ε_Y is a composite of the elasticity of labor supply and the elasticity of the wage (driven by the elasticity of education) to the income tax. Nevertheless, all that matters is the full elasticity of taxable income, as observed in the data.

[16] In addition, the redistributive effects could also be driven by the idiosyncratic preference shock η if we allows education s_{ti} to enter the utility function directly (as opposed to only indirectly through the wage). For instance, if higher ability agents tend to have a higher taste for education (so that θ and η would be positively correlated), $\bar{s}$ would be lower.

4 Optimal Linear Policies in a Dynastic Model

Turning back to the intergenerational model from Section 2, the government maximizes the expected welfare of the current generation, which takes into account all future generations' welfare through the altruistic preferences.[17]

$$SWF_0 = \max E \sum_{t=1}^{\infty} \beta^{t-1} \left[u_{ti} \left(\left(1-\tau_{Lt}\right) y_{ti} - s_{t+1i} \left(1-\tau_{St}\right) + R \left(1-\tau_{Bt}\right) b_{ti} - b_{t+1i} + G_t, y_{ti}, s_{ti} \right) \right]$$

subject to

$$G_t = \tau_{Lt} y_t + \tau_{Bt} R b_t - \tau_{St} s_{t+1}$$

I consider in turn two approaches to solve for the optimal policies. The first consists in changing one policy instrument at a time, taking into account the effect of that change on tax revenue. The second consists in simultaneously adjusting other instruments so as to keep revenue unaffected. Both of these approaches lead to different formulas, which can be useful under different circumstances, discussed in this section.

4.1 A Dynamic Reform Approach

The perturbation used to derive optimal policies is as follows. Consider a small reform $d\tau_{St} = d\tau_S$ that affects subsidy rates for all generations after some T. There is perfect foresight about the time and magnitude of the reform, and, hence, the dynasty will have anticipatory effects and start adjusting its choices even before time T. At the optimal τ_S, the change in social welfare from the reform must be zero.

From the envelope theorem linked to the agents' first-order conditions, the change in social welfare, $dSWF_0$ is:[18]

$$\begin{aligned} dSWF_0 &= \sum_{t \geq T} \beta^{t-1} E \left(s_{t+1i} u_{c,ti} \right) d\tau_S - \sum_{t \geq T} \beta^{t-1} E \left(u_{c,ti} \right) s_{t+1} d\tau_S \\ &\quad + \sum_{t<T} \beta^{t-1} E \left(u_{c,ti} \right) \left(-\tau_{St} ds_{t+1} + \tau_{Lt} dy_t + \tau_{Bt} R db_t \right) \\ &\quad + \sum_{t \geq T} \beta^{t-1} E \left(u_{c,ti} \right) \left(-\tau_{St} ds_{t+1} + \tau_{Lt} dy_t + \tau_{Bt} R db_t \right) \end{aligned}$$

The first term is the direct welfare effect of the reform. By the envelope theorem, it is equal to the weighted reduction in consumption from the subsidy change. This is true even if parents make the

[17] The current generation is normalized to be generation 1 and the maximization takes place at time $t = 0$, i.e., before that generation's uncertainty is realized.

[18] Recall that the envelope theorem states that we only need to take into account the direct effect of the tax change (i.e., the reduction in consumption from the tax change) on an agent's utility and can ignore to a first-order the indirect effects which act through changes in the agents' actions, as those are second-order at the agent's optimum.

choices instead of their children, as social preferences perfectly respect the dynastic preferences. If parents did not value their children's education the same way that the social planner did, then there would be an additional social welfare effect on the children from the parents' change in investment decisions. This would also occur if children's education (rather than children's utility) directly entered parental utility: $u_{ti}\left(c_{ti}, y_{ti}, s_{ti,} s_{t+1i}\right)$.

The second term is the mechanical revenue effect driven by the loss in tax revenue from the higher subsidy, at constant individual choices. These two effects only take place after the reform. The last two terms are the behavioral responses, before and after the reform respectively. The behavioral responses before the reform ($t < T$) are anticipatory effects.

Define the total elasticities of aggregate human capital s_{t+1}, output y_t, and bequests b_t to a small change in the education subsidy $d\tau_S$ for all $t > T$ to be:[19]

$$\varepsilon_{St+1} = \frac{ds_{t+1}}{d\left(\tau_S - 1\right)} \frac{\left(\tau_{St} - 1\right)}{s_{t+1}}, \qquad \varepsilon^S_{Yt} = \frac{dy_t}{d\left(\tau_S - 1\right)} \frac{\left(\tau_{St} - 1\right)}{y_t}, \qquad \varepsilon^S_{Bt} = \frac{db_t}{d\left(\tau_S - 1\right)} \frac{\left(\tau_{St} - 1\right)}{b_t}$$

In an ergodic stationary steady state in which all policies (τ_L, τ_S, and τ_B) are constant, for $t > T$ after the reform, each of these elasticities converges to the corresponding long-run elasticity when $t \to \infty$. Before the reform, the elasticities correspond to anticipatory elasticities, since τ_S has not been changed. Even though generations can in principle start reacting a long time before the reform, the responses are attenuated the further away in the future the reform is. It is convenient, hence, to a first approximation, to assume that the anticipatory effects of the reform only start once the steady state paths of all variables have been reached (i.e., T is large enough, so that aggregate variables have converged even before then).

On the steady state path, human capital, output, and bequests are constant. Hence, we can divide through by $E\left(u_{c,ti}\right) s_{t+1}$, constant in the steady state, where the expectation is taken over all dynasties i:

$$dSWF_0 = \sum_{t \geq 1}^{\infty} \beta^{t-1} \frac{1}{s_{t+1}} \left(-\tau_{St} \varepsilon_{St+1} \frac{s_{t+1}}{\left(\tau_S - 1\right)} + \tau_{Lt} \varepsilon^S_{Yt} \frac{y_t}{\left(\tau_S - 1\right)} + \tau_{Bt} \varepsilon^S_{Bt} \frac{Rb_t}{\left(\tau_S - 1\right)} \right) + \sum_{t \geq T} \beta^{t-1} \frac{E\left(s_{t+1i} u_{c,ti}\right)}{E\left(u_{c,ti}\right) s_{t+1}} - \sum_{t \geq T} \beta^{t-1}$$

Let the elasticities ε'_S, $\varepsilon^{S\prime}_Y$ and $\varepsilon^{S\prime}_B$ be the long-run elasticities of the present discounted value of each corresponding tax base, i.e.:

$$\varepsilon'_S \equiv (1-\beta) \sum_{t \geq 1} \beta^{t-1-T} \varepsilon_{St+1}, \qquad \varepsilon^{S\prime}_Y \equiv (1-\beta) \sum_{t \geq 1} \beta^{t-1-T} \varepsilon^S_{Yt}, \qquad \varepsilon^{S\prime}_B \equiv (1-\beta) \sum_{t \geq 1} \beta^{t-1-T} \varepsilon^S_{Bt} \tag{5}$$

[19]It is important the elasticities are defined as the reactions to the full policy, i.e., to the change in τ_S for all $t > T$, and not just to a one-period change.

For instance, ε'_S is the elasticity of the present discounted value of the education subsidy base with respect to a distant subsidy change. It is the sum of the discounted average of the standard post-reform elasticities and of the discounted average of the anticipatory elasticities. The same is true for the cross-elasticities of output and bequests to a distant subsidy change, $\varepsilon_Y^{S\prime}$ and $\varepsilon_B^{S\prime}$.

As in the static setup, define the redistributive incidences of output, education and bequests to be:

$$\bar{y} \equiv \frac{E\left(u_{c,ti} y_{ti}\right)}{E\left(u_{c,ti}\right) y_t}, \qquad \bar{s} \equiv \frac{E\left(u_{c,ti} s_{t+1i}\right)}{E\left(u_{c,ti}\right) s_{t+1}}, \qquad \bar{b} \equiv \frac{E\left(u_{c,ti} b_{ti}\right)}{E\left(u_{c,ti}\right) b_t} \tag{6}$$

To reiterate, each of these factors measures the strength of the covariance between the corresponding variable and marginal utility. The larger $\bar{y}$, $\bar{s}$, or $\bar{b}$ are, the more output, education, and bequests are concentrated among those with high marginal utilities of consumption.

At the optimal τ_S the change in welfare $dSWF_0$ must be zero, for any given level of τ_L and τ_B. Then, we can rearrange the expression to obtain the optimal education subsidy τ_S^*.

Proposition 3 *The optimal education subsidy, for any τ_L and τ_B, is given by:*

$$\tau_S^* = \frac{1 - \bar{s} + \varepsilon_Y^{S\prime} \tau_L \frac{y}{s} + \varepsilon_B^{S\prime} \tau_B R \frac{b}{e}}{1 - \bar{s} + \varepsilon'_S} \tag{7}$$

with $\bar{s}$ the distributional characteristic of education as defined in (6), and ε'_S, $\varepsilon_Y^{S\prime}$, and $\varepsilon_B^{S\prime}$, respectively, the long-run elasticities of the discounted education, income, and bequest tax bases as defined in (5).

Several features of optimal formula (7) are worth noting. First, the typical inverse elasticity effect is apparent: the subsidy is smaller when ε'_S (which is negative for any $\tau_S < 1$) is larger. It is also possible that the optimal subsidy is actually a tax. This is most likely to occur if the distributional value of education $\bar{s}$ is small, i.e., mostly high consumption agents acquire education.

Tax deductibility of education expenses: In the dynamic model, full tax deductibility of expenses such that $\tau_S = \tau_L$ is in general not optimal. A weaker result would be that education subsidies, income taxes, and bequest taxes move tpgether at the optimum.

$\varepsilon_Y^{S\prime}$ is the long-run elasticity of the discounted income tax base to a change in the education subsidy. As a result, it mixes both parents' and children's behavioral responses. A change in the education subsidy has three effects on income. First, through the substitution effect, a higher education subsidy induces parents to buy more education for their children. Second, however, there is an income effect on parents which reduces the need to work. Third, there are both substitution and income effects on children through their higher wages. On balance, depending on the utility functions the total effect could go either way. If there were no labor supply effects, then unambigously, $\varepsilon_Y^{S\prime} < 0$ (for a regular subsidy $\tau_S < 1$). In this case, τ_S and τ_L optimally move together.[20] However, this need not be the case with strong income effects.

Similarly when bequest taxes are higher, the human capital subsidy should be reduced if a lower human capital subsidy encourages more bequests through a strong substitution effect ($\varepsilon_S^{B\prime} < 0$). However, a higher education subsidy also has an income effect that might increase bequests, which are a normal good. Depending on which effect dominates, both the income tax and the education subsidy on the one hand, and the bequest tax and the education subsidy on the other hand could be positively related or not.

Evaluating education subsidy reforms: The aggregate elasticities and the distributional parameters are of course endogenous to the taxes and subsidies, and the formula is, as usual in the optimal tax literature, merely an implicit formula. However, it is especially useful for evaluating reforms around the current status quo. Indeed, the formula holds for any bequest and income taxes τ_B and τ_L, so that the right-hand side could be evaluated at the current tax and subsidy levels. If the implied τ_S^* is above the current τ_S, a reform that decreases τ_S would improve social welfare, and vice versa.

The distributional characteristic of education matters: The optimal subsidy is higher when those with high social welfare weights, that is, those with low consumption, also have high education expenses. The redistributive value of education is closely linked to how complementary education and ability are in the wage function, as explained in subsection 3.1 and explored in detail in Bovenberg and Jacobs (2011) and Stantcheva (2012). In addition, it can also be driven by idiosyncratic preferences for human capital and work, as captured by η.

However, it will not only depend on the technological and preference primitives, but also on the institutional setup of the education system and on who acquires the *subsidized* education. Indeed, suppose that the government provides basic education for all children through a public school system and that the subsidy under consideration actually only applies to additional education expenses, done at parental discretion, such as private school fees or tutoring lessons. If mostly high income agents incur such additional expenses, the approximation $\bar{s} \approx 0$ might be reasonable. On the other hand, if the subsidy applies to *all* education expenses, even low income agents might be beneficiaires of it.

The Optimal Income Tax: The optimal linear income tax τ_L^* can be similarly derived for given education subsidy and bequest tax:

$$\tau_L^* = \frac{1 - \bar{y} + \varepsilon_S^{Y\prime} \frac{s}{y} \tau_S - \varepsilon_B^{Y\prime} \tau_B R \frac{b}{y}}{1 - \bar{y} + \varepsilon_Y^{\prime}} \tag{8}$$

[20]Recall that the denominator in (7) is negative.

with $\varepsilon_S^{Y\prime}$, $\varepsilon_B^{Y\prime}$, and ε_Y' the long-run elasticities of the discounted present value of, respectively s, b, and y to a change in $(1-\tau_L)$ taking place for all generations after T. These elasticities are again the composites of the anticipatory elasticities and the post-reform elasticities. The typical trade-off between the redistributive and insurance benefit of taxation (in $\bar{y}$) and the efficiency cost of taxation (in ε_Y') is present. The fiscal spillovers to the education subsidy base and the bequest tax base enter as well, in a symmetric way and with the same explanation as for the optimal education subsidy previously explained.

The Optimal Bequest Tax: Finally, the optimal bequest tax is, at given τ_S and τ_L:

$$\tau_B^* = \frac{1 - \bar{b} + \varepsilon_S^{B\prime}\frac{s}{b}\tau_S - \varepsilon_Y^{B\prime}\tau_L\frac{y}{b}}{1 - \bar{b} + \varepsilon_B'} \tag{9}$$

The bequest tax is generically not zero. There are two different sets of reasons for this. First, if education subsidies and income taxes are already set, and the government needs to maximize the bequest tax holding the former fixed, there are fiscal spillovers to the education subsidy and the income tax bases. Bequest taxes indirectly influence education purchase and work decisions. Second, even in the absence of income taxes and education subsidies, the bequest tax is not zero as long as $\bar{b} \neq 1$. We would have $\bar{b} = 1$ if either utility were linear, or if everyone left the same bequest amount due to homogeneous, quasilinear preferences.[21]

4.2 Generalized Social Welfare Weights

Saez and Stantcheva (2015) show that, in standard optimal income tax formulas, it is possible to replace the standard social welfare weights, equal to $u_{c,ti}/E\left(u_{c,ti}\right)$, by generalized social welfare weights $g_{ti}/E(g_{ti})$ that directly place a social marginal value on an additional dollar transferred to person i. The same can be done here for education subsidies (in (7)), income taxes (in (8)) and bequest taxes (in (9)), which are all derived as functions of their distributional characteristics $\bar{s}$, $\bar{y}$, and $\bar{b}$. In particular, we can more generally define $\bar{s}$, $\bar{y}$, and $\bar{b}$ as, respectively, the education, income, and bequests weighted by generalized social marginal welfare weights:

$$\bar{s} = \frac{E(g_{ti}s_{ti})}{E(g_{ti})s_t} \quad \bar{y} = \frac{E(g_{ti}y_{ti})}{E(g_{ti})y_t} \quad \bar{b} = \frac{E(g_{ti}b_{ti})}{E(g_{ti})b_t}$$

This can allow us to incorporate different social objectives and equity considerations, without having to rederive new formulas: all these social criteria will translate into different values for $\bar{s}$, $\bar{y}$ and $\bar{b}$. I next illustrate the implications of different social welfare weights for the optimal education subsidy.

[21] This result is entirely consistent with Piketty and Saez (2013b), who do not consider education as an alternative way for parents to transfer resources to their children. Farhi and Werning (2013a) find that a wide range of positive or negative bequest taxes can be optimal depending on the social objective. A more direct link to Farhi and Werning (2013a) is drawn in Section 6 which considers optimal nonlinear taxation.

For instance, suppose that society thinks that children who receive no education from their parents should be compensated. In this limit case, $\bar{s} = 0$ and τ_S is set to the "Rawlsian" subsidy rate:

$$\tau_S^{\text{Rawls}} = \frac{1 + \varepsilon_Y^{S\prime} \tau_L \frac{y}{e} + \varepsilon_B^{S\prime} \tau_B R \frac{b}{e}}{1 + \varepsilon_S'}$$

On the other hand, it could be that society places a lot of value on parents who invest in their children's human capital. This would lead to a very large subsidy if $\bar{s} >> 1$. Social preferences may of course depend on the institutional setting and on who exactly acquires the subsidized education. For instance, if society is not concerned by redistribution, then $\bar{s} = 1$ and the optimal subsidy, denoted by $\tau_S^{\text{Efficiency}}$, is only driven by efficiency considerations, and takes into account the fiscal spillovers to the income tax and bequest tax bases:[22]

$$\tau_S^{\text{Efficiency}} = \frac{\varepsilon_Y^{S\prime} \tau_L \frac{y}{s} + \varepsilon_S^{B\prime} \tau_B R \frac{b}{e}}{\varepsilon_S'}$$

Finally, suppose that society only cares about people from a poor background. Then we could have simple binary weights such that $g_i = 1$ if an agent comes from a poor background and $g_i = 0$ otherwise. Then,

$$\bar{s} = \frac{E(s_{ti} \text{ for poor background kids})}{Prob(\text{poor background}) s_t},$$

so that $\bar{g}$ measures the relative education that agents from poor backgrounds have relative to the average education level across all backgrounds. Under this normative point of view, optimal education subsidies are likely to be small if, as is to be expected, agents from poor backgrounds only invest little in education.

4.3 Unobservable Education Investments

As in subsection 3.2, if human capital is unobservable, the optimal income tax will adjust, so as to compensate for the lack of a tool to directly control human capital investments. Supposing that bequests remain observable, the optimal linear tax will be:

$$\tau_L^{*,unobs} = \frac{1 - \bar{y} - \frac{b}{y} \varepsilon_B^{Y\prime} \tau_B}{1 - \bar{y} + \varepsilon_Y'}$$

Exactly as in the static case, the difference to the observable human capital case in formula (7) is the lack of term $-\tau_S \frac{s}{y} \varepsilon_S^{Y\prime}$ in the numerator, with $\varepsilon_S^{Y\prime}$ the long-run elasticity of the discounted education base to the retention rate. Given that the present discounted value of the education base should respond negatively to income taxes ($\varepsilon_S^{Y\prime} < 0$), then if $\tau_S^* > 0$ would have been optimal if education were observable, the optimal income tax will now need to be set lower with unobservable human capital (and, vice-versa if $\tau_S^* < 0$ would have been optimal).

[22] If in addition there are no revenue requirements, then $\tau_S = \tau_B = \tau_L = 0$.

Similarly, the optimal bequest tax, for any given labor tax, will have to be adjusted to compensate for the lack of an education subsidy. The optimal linear bequest tax with unobservable education will be:

$$\tau_B^{*,unobs} = \frac{1-\bar{b}-\varepsilon_Y^{B\prime}\tau_L\frac{y}{b}}{1-\bar{b}+\varepsilon_B^{\prime}}$$

The term $\varepsilon_S^{B\prime}\frac{s}{b}\tau_S$ is missing form the numerator. If education and bequests are substitutes overall, $\varepsilon_S^{B\prime} < 0$ so that if a positive τ_S had been optimal, the bequest tax will now be increased so as to indirectly encourage parents to channel resources into their children's education. In both cases, the formula looks like the formula if there was no human capital at all. But the value of the elasticities will of course depend on the presence of education.

4.4 Reform elasticities

There is another way to determine optimal tax and subsidy rates, which might prove to be more convenient in some situations. Indeed, the shortcoming of (7) is that it relies on the (endogenous) cross-elasticities of output or wealth to education subsidies, $\varepsilon_Y^{S\prime}$ and $\varepsilon_B^{S\prime}$, at a given τ_L and τ_B. However, often, all that is observed in the data is the full response of some set of variables to a reform. A full reform is often a combination of changes in several tax tools, with or without revenue-neutrality.

Thus, in any country, there might have been specific reforms already implemented, which can serve as natural experiments to estimate elasticities. For each reform, one can derive the implied optimal education subsidy as a function of the "reform elasticities," i.e., the *full* responses observed during that particular reform. It is then not crucial to know what the underlying cross-elasticities are. The analysis can be performed for different types of reforms, and is illustrated below for a change in education subsidies financed by an increase in income taxes. It bears repeating, however, that the same type of formula can be derived for any reform considered and, hence, adapted to the empirical evidence available.

To illustrate reform-specific elasticities, consider a small revenue-neutral reform $d\tau_{St} = d\tau_S$ for $t > T$, but suppose now that there is a corresponding series of income tax reforms $d\tau_{Lt}$ to maintain budget balance, around constant τ_S and τ_L. The bequest tax τ_B is left unchanged. T is again large enough for all variables to have converged to their steady state paths. At an optimum, the change in social welfare from this reform must be zero. Using the envelope theorem from the agents' first-order conditions:

$$dSWF_0 = \sum_{t>T}\beta^{t-1}E\left(s_{t+1i}u_{c,ti}\right)d\tau_S - \sum_{t\geq 1}\beta^{t-1}E\left(y_{ti}u_{c,ti}\right)d\tau_{Lt} = 0$$

Define the long-run elasticities to the full reform at constant revenue as:

$$\varepsilon_{Bt} \equiv \frac{db_t}{d(\tau_S - 1)} \frac{\tau_{St} - 1}{b_t}|_G, \quad \varepsilon_{St+1} \equiv \frac{ds_{t+1}}{d(\tau_S - 1)} \frac{\tau_{St} - 1}{s_{t+1}}|_G, \quad \varepsilon_{Yt} \equiv \frac{dy_t}{d(1 - \tau_L)} \frac{1 - \tau_{Lt}}{y_t}|_G$$

where db_t, ds_{t+1}, and dy_t are the responses of, respectively, aggregate savings, aggregate human capital, and aggregate output to the *full* reform $(d\tau_S, d\tau_{Lt})$ at constant revenue G. Note that these elasticities capture the joint total effects of the simultaneous changes in τ_S and τ_L on b_t, s_{t+1} and y_t, with τ_B held constant. For each variable, they are composites of own-tax and cross-tax effects.[23] By contrast, the formula in (7) isolated the pure effect of the education subsidy on all other variables, holding other taxes fixed.

Exactly as above, due to the anticipatory effects of forward-looking life-cycle agents, the reaction of these variables to the reform may start even before the reform period T. We assume that T is large enough so that, to a first-order, the anticipatory reactions only start once all three choice variables have reached their steady state paths.

For the budget to remain balanced in all periods, the income tax needs to be adjusted such that, for $t > T$ and any $d\tau_S$:

$$\left(-\varepsilon_{St+1} \frac{\tau_S}{\tau_S - 1} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1} - 1\right) s_{t+1} d\tau_S = -\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right) y_t d\tau_{Lt}$$

and for $t \leq T$:

$$\left(-\varepsilon_{St+1} \frac{\tau_S}{\tau_S - 1} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1}\right) s_{t+1} d\tau_S = -\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right) y_t d\tau_{Lt}$$

Substituting for these tax changes in $dSWF_0$ and dividing by $s_{t+1} E(u_{c,ti})$ (constant in the steady state) yields:

$$dSWF_0 = \bar{s} \frac{\beta^{T-1}}{1 - \beta} + \bar{y} \left(\sum_{t<T} \beta^{t-1} \frac{\left(-\varepsilon_{St+1} \frac{\tau_S}{\tau_S - 1} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1}\right)}{\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right)} + \sum_{t \geq T} \beta^{t-1} \frac{\left(-\varepsilon_{St+1} \frac{\tau_S}{(\tau_S - 1)} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1} - 1\right)}{\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right)} \right) = 0$$

Let ε'_S and ε'_B be the total long-run response of the discounted human capital, bequest and output bases to the reform, as in subsection 4.1. Define ε'_Y as the composite elasticity of output that ensures that the following equality holds:

$$\frac{\left(-\varepsilon'_S \frac{\tau_S}{(\tau_S - 1)} + R \frac{b}{e} \varepsilon'_B \frac{\tau_B}{\tau_S - 1} - 1\right)}{\left(1 - \varepsilon'_Y \frac{\tau_L}{1 - \tau_L}\right)}$$

$$= (1 - \beta) \sum_{t<0} \beta^{t-T} \frac{\left(-\varepsilon_{St+1} \frac{\tau_S}{\tau_S - 1} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1}\right)}{\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right)} + (1 - \beta) \sum_{t \geq T} \beta^t \frac{\left(-\varepsilon_{St+1} \frac{\tau_S}{\tau_S - 1} + R \frac{b_t}{s_{t+1}} \varepsilon_{Bt} \frac{\tau_B}{\tau_S - 1} - 1\right)}{\left(1 - \varepsilon_{Yt} \frac{\tau_L}{1 - \tau_L}\right)}$$

[23] The normalization by $(\tau_S - 1)$ of ε_{Bt} is arbitrary. We could have normalized by $(1 - \tau_L)$ instead.

Setting $dSWF_0 = 0$ and rearranging the expression at constant τ_S and τ_L yields the optimal human capital subsidy, formulated according to the reform-specific elasticities.

Proposition 4 *The optimal human capital subsidy that maximizes the expected welfare of the dynasty, for any τ_L and τ_B, is given by:*

$$\tau_S^* = \frac{1 - \frac{\bar{s}}{\bar{y}}\left(1 - \varepsilon'_Y \frac{\tau_L}{1-\tau_L}\right) + R\frac{b}{s}\varepsilon'_B \tau_B}{1 - \frac{\bar{s}}{\bar{y}}\left(1 - \varepsilon'_Y \frac{\tau_L}{1-\tau_L}\right) + \varepsilon'_S} \tag{10}$$

where the long-run elasticities ε'_B, ε'_Y and ε'_S are the total elasticities to a revenue neutral reform that changes τ_S and adjusts τ_L to maintain budget balance, and $\bar{s}$ and $\bar{y}$ are the distributional factors of human capital and income, as defined in (6).

The elasticities in formula (10) are reform-specific, i.e., they measure the total impact on the aggregate variables b_t, s_{t+1}, and y_t of changing τ_S while adjusting τ_L to maintain revenue-neutrality and keeping τ_B constant. Hence, this formulation is most useful when there have been past reforms resembling exactly this one, so that the elasticities can be estimated in the data. Conversely, an analog of formula (10) can easily be derived again for other reforms, with differently defined elasticities. The advantage of formulation (10) is that, if such suitable reforms already exist, the elasticities are readily available without having to separately estimate all cross-tax effects. Again, it is not necessary to assume that either τ_L or τ_B are optimally set in the economy.

It is worth repeating that, with perfect estimation tools that would allow us to uncover all cross elasticities and all reform-specific elasticities, at any tax levels, formulas (7) and (10) would yield the same answer. Alternatively, if we knew or were willing to make assumptions on the primitives and could obtain the Slutsky matrices, the formulas would yield equal answers as well. They are merely two ways of approaching the same question, and one of these ways may be empirically easier.

5 Credit Constraints

Up to here, the analysis assumed away credit constraints that could prevent parents from investing in their children's education. However, as mentioned in the Introduction, a large literature documents the existence of credit constraints for parents.

To make the study of credit constraints meaningful, the model from Section 2 is now augmented to an overlapping generations model, in which each generation's lifetime lasts for three periods of time. Variable t indexes time periods. A generation is called generation t if it is born in period t. In the first period of life, agents of generation t are born and are "young." They receive an investment of human capital equal to s_t from their parents. In the second period of their life, $t+1$, agents of generation t become "adults" and have one child each. They work to earn income y_{t+1}, choose to save an amount k_{t+1} for their old age, and invest an amount s_{t+1} of human capital in their children. For simplicity, savings k_{t+1} yield a tax-exempt gross return of 1. Finally, in their third period of life, period $t+2$, agents from generation t become "old" and at the beginning of the period receive bequests b_{t+1} from their own parents. These bequests again yield a gross pre-tax return R. They also chose how much bequests b_{t+2} to leave to their own children at the end of the period before dying. This mirrors the fact that people typically receive bequests after education investments have taken place.

Hence, in each period, there is a mass of one of each type of agents: young, adult, and old and population size does not change. Given the simple setup, in every period, only adults work, invest in human capital and save, while only the old receive and leave bequests. Thus we denote by y_t the output produced in period t by the adults of that period (who were born in generation $t-1$), by s_t the human capital investment made by those same adults (which is immediately received by the young agents born in generation t), and by b_{t-1} the bequests received by the old in period t.

For simplicity, in this section only, labor supply is assumed to be inelastic.[24] The income of generation t, earned in period $t+1$, in dynasty i is given by:

$$y_{t+1i} = w_{t+1}(s_{ti}, \theta_{t+1i})$$

where θ_{t+1i} is as before the generation's stochastic productivity. Since each generation only consumes in old age, the utility of generation t of dynasty i is equal to:

$$u_{t+2i}(c_{t+2i}, \eta_{t+2i}),$$

were η is again a preference shock. The budget constraint of an adult agent born in generation t (and, hence, working and saving in period $t+1$) is:

$$(1-\tau_{Lt+1})w_{t+1}(s_{ti}, \theta_{t+1i}) = k_{t+1i} + s_{t+1i}(1-\tau_{St+1}) \tag{11}$$

The budget constraint of the same agent when he is old (in period $t+2$) is:

$$k_{t+1i} + Rb_{t+1i}(1-\tau_{Bt+2}) = c_{t+2i} + b_{t+2i} \tag{12}$$

Using (11) and (12), we can express consumption in old age as:

$$c_{t+2i} = Rb_{t+1i}(1-\tau_{Bt+2}) + (1-\tau_{Lt+1})w_{t+1}(s_{ti}, \theta_{t+1i}) - s_{t+1i}(1-\tau_{St+1}) - b_{t+2i} + G_{t+2} \tag{13}$$

[24]The results obtained below that credit constraints tend to increase the optimal education subsidy and decrease the optimal income tax still apply with elastic labor supply.

The government rebates the transfer to agents at the beginning of their old age (after the bequests received at the beginning of old age have been taxed). The transfer to generation born at $t-2$ and who is old at time t is:

$$G_t = \tau_{Lt-1} y_{t-1} + \tau_{Bt} R b_{t-1} - \tau_{St-1} s_{t-1} \tag{14}$$

Each generation values the utility (in old age) of the next generation, discounted by an intergenerational discount rate β. Thus, social welfare is as before:

$$SWF_0 = \max E \sum_{t=1}^{\infty} \beta^{t-1} \left[u_{ti} \left(\left(1 - \tau_{Lt-1}\right) y_{t-1i} - s_{t-1i} \left(1 - \tau_{St-1}\right) + R \left(1 - \tau_{Bt}\right) b_{t-1i} - b_{ti} + G_t \right) \right] \tag{15}$$

where, now, u_{ti} is the utility of the old agents at time t, so that u_{1i} is the old generation at time 1 born at time $t = -1$. The benchmark case without credit constraints imposes no further restrictions and the optimal formulas will be as in Section 4.1.

With credit constraints, savings during adulthood cannot be negative, so we need to impose:

$$k_{t-1i} = (1 - \tau_{Lt-1}) w_{t-1}(s_{t-2i}, \theta_{t-1i}) - s_{t-1i}(1 - \tau_{St-1}) \geq 0 \tag{16}$$

Let γ_{ti} be the multiplier on the credit constraint in period t on an adult agent of dynasty i. Changes in policies will now, in addition to the standard effects, involve effects on the credit constraints of parents. Note that parents already take into account the effects of the education investments on their children's credit constraints, so that, by the envelope theorem, the indirect effects of the subsidy on credit constraints through the education and income choices have a zero first-order welfare effect.

Define two incidence measures of credit constraints, respectively $\tilde{s}$ and $\tilde{y}$ by:

$$\tilde{s} \equiv \frac{E(\gamma_{ti} s_{t-1i})}{E(u_{c,ti}) s_{t-1}}, \qquad \tilde{y} \equiv \frac{E(\gamma_{ti} y_{t-1i})}{E(u_{c,ti}) y_{t-1}} \tag{17}$$

$\tilde{s}$ and $\tilde{y}$ measure how concentrated credit constraints are, respectively, on parents who invest a lot in their children's education and on parents with low incomes.

Then, similar calculations as in Section 4.1 yield the optimal human capital subsidy with credit constraints, denoted by $\tau_S^{*,cc}$:

Proposition 5 *The optimal human capital subsidy with credit constraints, for any τ_L and τ_B, is given by:*

$$\tau_S^{*,cc} = \frac{1 - (\bar{s} + \tilde{s}) + \varepsilon_Y^{S\prime} \tau_L \frac{y}{s} + \varepsilon_B^{S\prime} \tau_B R \frac{b}{s}}{1 - (\bar{s} + \tilde{s}) + \varepsilon_S^{\prime}} \tag{18}$$

with $\bar{s}$ the distributional characteristic of education as defined in (6),[25] *$\varepsilon_S^{\prime}$, $\varepsilon_Y^{S\prime}$, and $\varepsilon_B^{S\prime}$, respectively, the long-run elasticities of the discounted human capital, income, and bequest tax bases as defined in (5), and $\tilde{s}$ the credit constraint weighted education from (17).*

[25] With the only difference that s_{ti} in the definition is always replaced by s_{t-1i} and s_t by s_{t-1} due to the slightly different timing in the OLG model here.

Comparing the optimal subsidy $\tau_S^{*,cc}$ in the presence of credit constraints and the optimal subsidy without credit constraints in (10) there is only one additional term, namely $\tilde{s}$ that measures how concentrated credit constraints are among parents who invest a lot in their children's education. While it is not in general possible to compare the levels of the subsidies analytically, because all variables and elasticities will be endogenous to the presence of credit contraints, it is possible to discuss in which direction this term would tend to influence the optimal subsidy, all else equal. The incidence of credit constraints $\tilde{s}$ acts entirely symmetrically to the distributional incidence of the subsidy, $\bar{s}$. If credit constraints are concentrated among parents with high investments in their children's education, $\tilde{s}$ is higher and this acts to effectively increase the positive distributional impact of an education subsidy. Accordingly, $\tau_S^{*,cc}$ will tend to be higher.

The optimal labor tax will be similarly modified, compared to the formula without credit constraints in (8).

$$\tau_L^{*,cc} = \frac{1 - (\bar{y} + \tilde{y}) + \varepsilon_S^{Y\prime} \frac{s}{y} \tau_S - \varepsilon_B^{Y\prime} \tau_B R \frac{b}{y}}{1 - (\bar{y} + \tilde{y}) + \varepsilon_Y^{\prime}} \tag{19}$$

The more credit constraints are concentrated among low income agents ($\tilde{y}$ small), the higher the income tax is. Again, the incidence of credit constraints acts exactly like the distributional characteristic of output, $\bar{y}$.

On the other hand, the optimal bequest tax is unaffected and formula (9) still applies. There are however two important remarks related to this result. First, while the formula for optimal bequest taxes for any given level of τ_S and τ_L is unaffected, the actual level of the optimal bequest tax may be very different with credit constraints since agents' choices may change. Second, this result strongly hinges on the timing of bequests, and, in particular, on the assumption that bequests occur relatively late in life and do not relieve the credit constraints of (adult) children.

6 Nonlinear Dynamic Education and Bequest Taxation

I now turn to a full-fledged dynamic mechanism design approach to the previous problem. This approach is considered in detail in Stantcheva (2012) in a life cycle setting, without intergenerational considerations and bequests. It is adapted here to the intergenerational setting.

First, some additional structure for the stochastic process is imposed. Every generation has a stochastic unobserved ability θ_t, which follows a Markov processs $f^t(\theta_t|\theta_{t-1})$, hence alowing for persistence between generations. For instance, if the correlation between parental and child ability is positive, then more productive parents are more likely to have more productive children. This implies that parents have some advance information about the potential productivity of their children when making the human capital investment decision – and the more so if the Markov process exhibits more persistence– but do not know the exact ability realization of their child. We abstract from the preference shock η.

Second, to simplify the problem, I assume that utility is separable in consumption and labor with:

$$\tilde{u}_t\left(c_t, y_t, s_t; \theta_t\right) = u_t\left(c_t\right) - \phi_t\left(\frac{y_t}{w_t\left(\theta_t, s_t\right)}\right)$$

u_t is increasing, twice continuously differentiable, and concave. ϕ_t is increasing, convex and twice continuously differentiable.

Denote the partials of the wage with respect to ability and education respectively by w_θ and w_s, and by $w_{\theta s}$ the second order partial of the wage. Similarly, let ϕ_l and ϕ_{ll} denote, respectively, the first and second order partial of the disutility function. A crucial parameter for the optimal solution will be the Hicksian coefficient of complementarity between ability and education in the wage function at time t (Hicks, 1970; Samuelson, 1974), denoted by $\rho_{\theta s,t}$

$$\rho_{\theta s,t} \equiv \frac{w_{\theta s,t} w_t}{w_{s,t} w_{\theta,t}} \tag{20}$$

A positive Hicksian complementarity between education s and ability θ means that higher ability agents have a higher marginal benefit from human capital ($w_{\theta s} \geq 0$). A Hicksian complementarity greater than 1 means that higher ability agents have a higher *proportional* benefit from human capital, i.e., the wage elasticity with respect to ability is increasing in education, i.e., $\frac{\partial}{\partial s}\left(\frac{\partial w}{\partial \theta}\frac{\theta}{w}\right) \geq 0$.

6.1 Program

Denote by θ^t the history of ability shocks up to period t, by Θ^t the set of possible histories at t, and by $P\left(\theta^t\right)$ the probability of a history θ^t, $P\left(\theta^t\right) \equiv f^t\left(\theta_t|\theta_{t-1}\right)...f^2\left(\theta_2|\theta_1\right)f^1\left(\theta_1\right)$. We imagine that the government designs a direct revelation mechanism in which he assigns allocations as functions of the history of reports of an agent. An allocation $\left\{c\left(\theta^t\right), y\left(\theta^t\right), s\left(\theta^t\right)\right\}_{\Theta^t}$ specifies consumption, output, parental education investment (and, hence, bequests) for each period t, conditional on the history θ^t. Define the continuation utility of the dynasty after history θ^t recursively as:

$$\omega\left(\theta^t\right) = u_t\left(c\left(\theta^t\right)\right) - \phi_t\left(\frac{y\left(\theta^t\right)}{w_t\left(\theta_t, s\left(\theta^{t-1}\right)\right)}\right) + \beta\int \omega\left(\theta^{t+1}\right) f^{t+1}\left(\theta_{t+1}|\theta_t\right) d\theta_{t+1}$$

A first-order approach is followed, in which the incentive compatibility constraint of each generation t is replaced by the envelope condition. The envelope condition is:[26]

$$\dot{\omega}\left(\theta^t\right) := \frac{\partial \omega\left(\theta^t\right)}{\partial \theta_t} = \frac{w_{\theta,t}}{w_t} l\left(\theta^t\right)\phi_{l,t}\left(l\left(\theta^t\right)\right) + \beta\int \omega\left(\theta^{t+1}\right)\frac{\partial f^{t+1}\left(\theta_{t+1}|\theta_t\right)}{\partial \theta_t} d\theta_{t+1} \tag{21}$$

[26] The reader can refer to Stantcheva (2012) for all the technical steps.

To write the problem recursively, let the future marginal rent (the second term in the envelope condition) be denoted by:

$$\Delta\left(\theta^{t}\right) \equiv \int \omega\left(\theta^{t+1}\right) \frac{\partial f^{t+1}\left(\theta_{t+1}|\theta_{t}\right)}{\partial \theta_{t}} d\theta_{t+1} \tag{22}$$

The envelope condition can then be rewritten as:

$$\dot{\omega}\left(\theta^{t}\right) = \frac{w_{\theta,t}}{w_{t}} l\left(\theta^{t}\right) \phi_{l,t}\left(l\left(\theta^{t}\right)\right) + \beta\Delta\left(\theta^{t}\right) \tag{23}$$

Let $v\left(\theta^{t}\right)$ be the expected future continuation utility:

$$v\left(\theta^{t}\right) \equiv \int \omega\left(\theta^{t+1}\right) f^{t+1}\left(\theta_{t+1}|\theta_{t}\right) d\theta_{t+1} \tag{24}$$

Continuation utility $\omega\left(\theta^{t}\right)$ can hence be rewritten as:

$$\omega\left(\theta^{t}\right) = u_{t}\left(c\left(\theta^{t}\right)\right) - \phi_{t}\left(\frac{y\left(\theta^{t}\right)}{w_{t}\left(\theta_{t}, s\left(\theta^{t-1}\right)\right)}\right) + \beta v\left(\theta^{t}\right) \tag{25}$$

Define the continuation cost of the government for generation t, $K\left(v, \Delta, \theta_{-}, s_{-}, t\right)$, as a function of the promised utility, promised marginal utility, the previous' generations type and the education investments by parents. The program of the government is:

$$K\left(v, \Delta, \theta_{-}, s_{-}, t\right) = \min \int (c(\theta) + s(\theta) - w_{t}\left(\theta, s_{-}\right) l\left(\theta\right) + \frac{1}{R} K\left(v\left(\theta\right), \Delta\left(\theta\right), \theta, s\left(\theta\right), t+1\right)) f^{t}\left(\theta|\theta_{-}\right) d\theta$$

subject to:

$$\omega\left(\theta\right) = u_{t}\left(c\left(\theta\right)\right) - \phi_{t}\left(l\left(\theta\right)\right) + \beta v\left(\theta\right) \tag{26}$$

$$\dot{\omega}\left(\theta\right) = \frac{w_{\theta,t}}{w_{t}} l\left(\theta\right) \phi_{l,t}\left(l\left(\theta\right)\right) + \beta\Delta\left(\theta\right) \tag{27}$$

$$v = \int \omega\left(\theta\right) f^{t}\left(\theta|\theta_{-}\right) d\theta \tag{28}$$

$$\Delta = \int \omega\left(\theta\right) \frac{\partial f^{t}\left(\theta|\theta_{-}\right)}{\partial \theta_{-}} d\theta \tag{29}$$

where the maximization is over the functions $\left(c\left(\theta\right), l\left(\theta\right), s\left(\theta\right), \omega\left(\theta\right), v\left(\theta\right), \Delta\left(\theta\right)\right)$.

6.2 Optimal policies

In the second best, marginal distortions relative to a *laissez-faire* economy are described using "wedges." (For a detailed explanation of the wedges see Golosov, Tsyvinski, and Werning (2006) or Stantcheva (2012)). As their definitions reflect, these wedges are similar to locally linear subsidies and taxes. For any allocation, define the intratemporal wedge on labor $\tau_{L}\left(\theta^{t}\right)$

$$\tau_{L}\left(\theta^{t}\right) \equiv 1 - \frac{\phi_{l,t}(l_{t})}{w_{t} u_{t}'\left(c_{t}\right)} \tag{30}$$

as the gap between the marginal rate of substitution and the marginal rate of transformation between consumption and labor. A positive labor wedge means that labor is distorted downwards relative to the laissez-faire. Similarly, define the intertemporal wedge on bequests, τ_B as the gap between the marginal rate of substitution between consumption of today's and tomorrow's generations and the return on bequests:

$$\tau_B\left(\theta^t\right) \quad \equiv \quad 1-\frac{1}{R\beta}\frac{u_t'\left(c_t\right)}{E_t\left(u_t'\left(c_{t+1}\right)\right)} \tag{31}$$

The first result, which is standard in the dynamic taxation literature and which continues to hold in this intergenerational model with human capital, is that at the optimum, the Inverse Euler Equation holds as in the model of (Rogerson, 1985). This, combined with Jensen's inequality, implies that there is a positive wedge on bequests, $\tau_B > 0$.

More interesting is the relation between bequests and human capital at the optimum, which is very simple and characterized in the following proposition.

Proposition 6 *At the optimum, the following relation needs to be satisfied:*

$$R = E\left(w_{s,t+1}l_{t+1}(1+\tau_{Lt+1}\frac{\varepsilon_{t+1}^c}{1+\varepsilon_{t+1}^u}\left(1-\rho_{\theta s,t+1}\right))\right)$$

where ε_t^u and ε_t^c are, respectively, the uncompensated and compensated labor supply elasticities to the net wage at fixed bequests.[27] The left-hand side is simply the (social) return on bequests.[28] The right-hand side is the social return to education. The first part of the social return to education is just the wage increase of the next generation from education. The second part captures the incentive implications of education for the next generation.

Education has two effects on the incentives of children. First, it encourages their work effort, which relaxes their incentive constraints. This is the so-called "labor supply effect." Second, depending on the sign of the complementarity between human capital and ability, education may increase or decrease pre-tax inequality. If $\rho_{\theta s} > 0$, education increases pre-tax inequality and benefits mostly able kids. This tends to reduce the effective incentive-adjusted benefit of education and is called the "inequality effect." The net effect on children's incentives depends on the sign of $(1-\rho_{\theta s})$, called the redistributive and insurance effect of human capital (Stantcheva, 2012). The redistributive and insurance effect of

[27] ε^c and ε^u are defined as in the static framework (Saez, 2001), at constant bequests:

$$\varepsilon^u = \frac{\phi_l(l)/l + \frac{\phi_l(l)^2}{u'(c)^2}u''\left(c\right)}{\phi_{ll}(l) - \frac{\phi_l(l)^2}{u'(c)^2}u''\left(c\right)} \qquad \varepsilon^c = \frac{\phi_l(l)/l}{\phi_{ll}(l) - \frac{\phi_l(l)^2}{u'(c)^2}u''\left(c\right)}$$

With per-period utility separable in consumption and labor, $\frac{\varepsilon_t^c}{1+\varepsilon_t^u-\varepsilon_t^c}$ is the Frisch elasticity of labor.

[28] The private return is $R(1-\tau_B)$ where τ_B is the bequest wedge.

education is scaled by $\tau_L \frac{\varepsilon^c_{t+1}}{1+\varepsilon^u_{t+1}}$, which captures the efficiency cost of taxation, i.e., the value of relaxing children's incentive constraints.

At the optimum, the return on bequests is not equated to the return on human capital investments: instead, it needs to be equated to the expected, incentive-adjusted return on education that takes into account the direct increase in earnings and the labor supply effect and the inequality effect on the incentive constraint. While bequests benefit all types uniformly in marginal terms, human capital investments have redistributive incentive effects.[29]

If education is highly complementary to ability with $\rho_{\theta s} > 1$, which is equivalent to high ability children benefitting more in proportional terms from their parents' education investments, then the return to education investments will be reduced below that on bequests. Put differently, education investments by parents will be taxed relative to bequests. The opposite happens when education is not too complementary to children's ability ($\rho_{\theta s} < 1$), in which case parental education investments should be subsidized relative to bequests.

Note that the typically used wage function in the human capital literature (Bovenberg and Jacobs, 2005) with $\rho_{\theta s} = 1$ would imply that relative to one another, parental education investments and bequest choices should not be distorted, i.e.,

$$R = E(w_{s,t+1} l_{t+1}).$$

Note that Farhi and Werning (2010) also find that it is optimal to not distort the trade-off between bequests and human capital purchases by parents, despite a very general wage function. Their result, however, is driven by the fact that children in the second period of their two-period model do not work, so that there is no incentive compatibility constraint for them.

7 Dynamic Complementarities in Education and Early Childhood Investments

A large body of important work has emphasized how crucial early childhood investments are relative to investments early in life. A few papers among the many others written by James Heckman on this topic are for instance Cunha and Heckman (2007), Cunha and Heckman (2008), Cunha, Heckman, Lochner, and Masterov (2006). Heckman (2011) and Heckman, R., and P. (2013) go into more detail into the formation of cognitive and non-cognitive skills in early childhood years. What is emphasized by these papers is two important features which require special attention when it comes to policy design. The first, treated in this section, pertains to dynamic complementarities between investments

[29] Bequests would have income effects that would interact with agents' types if utility were not separable in consumption and labor.

in earlier periods and those in later periods. The second, treated in Section 8 is that there are several *types* of skills, such as cognitive and non-cognitive skills, each with a different production function and different impacts on the wage.

To look into the issues of dynamic complementarities, consider a specialization of the model in Section 6 which includes three periods. In the first period, children start with a level of human capital of s_0 and an investment e_1 is made by parents. This leads to the production of a skill level

$$s_1 = f_1(s_0, e_1) \tag{32}$$

. In the second period, a further investment e_2 is done and produces an updated skill level s_2 according to the production function:

$$s_2 = f_2(s_1, e_2) \tag{33}$$

Following Cunha and Heckman (2007) and Cunha and Heckman (2008), we call investments in periods 1 and 2 complementary if:

$$\frac{\partial f_2^2(s_1, e_2)}{\partial e_2 \partial s_1} > 0 \tag{34}$$

We call investments *critical* if investing in skills has no marginal product unless it is done in period 1, i.e.:

$$\frac{\partial f_2(s_1, e_2)}{\partial e_2} = 0 \tag{35}$$

Finally, the authors call a period sensitive if investment in that period is more productive than investment in the other period. For instance, period 1 is sensitive if:

$$\frac{\partial f_2(s_1, e_2)}{\partial e_2}|_{s_1=s, e_2=e} < \frac{\partial f_1(s_0, e_1)}{\partial e_1}|_{s_0=s, e_1=e} \tag{36}$$

Note that in this case, we can rewrite the monetary cost of human capital from aboce as a function of both s_{t+1} and s_t, i.e., $M_t(e_t) = M_t(f_f^{-1}(s_{t-1}, s_t))$.

Finally, in period 3, the child enters the labor market and earns a wage equal to, as before, a function of the stochastic shock (which, recall, proxies for labor market risk and earnings uncertainty) θ and his skill stock acquired throughout childhood, s_2. Hence the wage of the child born in generation t is given by:

$$w_t(s_2, \theta_t) \tag{37}$$

If we set up the same problem as in Section 6 now with this augmented production technology, we can solve for the optimal subsidies on parental investments in each period.

Proposition 7 *For period 1 investment, at the optimum, the following relation needs to hold:*

$$R + E\left(\frac{\partial M_2}{\partial e_2}\frac{\partial f_2^{-1,e}}{\partial s_1}\right) = E\left(w_{s,t+1} l_{t+1}(1 + \tau_{Lt+1}\frac{\varepsilon^c_{t+1}}{1+\varepsilon^u_{t+1}}(1-\rho_{\theta s,t+1}))\right)$$

where $f_2^{-1,e}$ *denotes the partial inverse of function* f_2 *with respect to its second argument* e*.*

Hence, period 1 investment, which is easily interpretable as early childhood investment, is very special. The new term on the left-hand side (relative to the optimum expression in proposition 6) captures how and by how much period 1 investments will influence the cost of period 2 investments. As long as investments done in early childhood reduce the cost of later investments, there is an argument for increasing the subsidy on them beyond what was optimal in the case without dynamic complementarities. I.e., this holds true as long as period 1 and 2 investments are complementary, in the language of Cunha, Heckman, Lochner, and Masterov (2006).

Note that this argument does not rely at all on early childhood investments having a stronger marginal product, but purely on the dynamic complementarities in investments. If in addition early childhood investments had a higher marginal product in terms of the production function f_2, then there would be an even stronger argument for subsidizing them further.

Hence, the general model in this paper can be nicely extended to capture important structural issues such as dynamic complementarities and early childhood investments. It would be very fruitful to use the detailed estimates in the aforementioned papers in order to calibrate formula 38 and to obtain the value of optimal early childhood subsidies.

8 Multidimensional Skills: Cognitive and non-cognitive skills

Another crucial issue discovered in the aforementioned papers is the distinction between cognitive and non-cognitive skills. Skills are not unidimensional and a good model needs to account for this. In this section, I show how to extend the model to allow for multidimensional skills. The argument is valid for any numbers of skills, but I illustrate it here with a "cognitive" skill denoted by s^c and a non-cognitive skill, denoted by s^n.

We can allow these skills to enter the ultimate wage earned in the labor market differently, i.e., the wage will now be a function

$$w_t(s_t^c, s_t^n, \theta_t) \tag{38}$$

of both types of skill, as well as the stochastic shock θ_t. Correspondingly, define the Hicksian coefficient of complementarity for each type of skill (cognitive and non-cognitive) as in formula (20) earlier on:

$$\rho_{\theta s,t}^c \equiv \frac{w_{\theta s^c,t} w_t}{w_{s^c,t} w_{\theta,t}} \qquad \rho_{\theta s,t}^n \equiv \frac{w_{\theta s^n,t} w_t}{w_{s^n,t} w_{\theta,t}} \tag{39}$$

It has the same interpretation as before: a larger Hicksian coefficient of complementarity on one type of skill (e.g., cognitive skills) means that that type of skill is more strongly complementary with the stochastic ability θ. A Hicksian coefficient of zero would mean there is no complementarity between the skill type and stochastic ability. On the other hand, a negative Hicksian coefficient means that investments in that skill type will level the playing field: they will benefit most low ability children and improve their outcomes.

Proposition 8 *At the optimum, the following relation needs to be satisfied for investment in cognitive skills:*

$$R = E\left(\frac{\partial w_{t+1}}{\partial s^c_{t+1}} l_{t+1}(1 + \tau_{Lt+1}\frac{\varepsilon^c_{t+1}}{1+\varepsilon^u_{t+1}}\left(1 - \rho^c_{\theta s,t+1}\right))\right)$$

and the same relation needs to hold for non-cognitive skills:

$$R = E\left(\frac{\partial w_{t+1}}{\partial s^n_{t+1}} l_{t+1}(1 + \tau_{Lt+1}\frac{\varepsilon^c_{t+1}}{1+\varepsilon^u_{t+1}}\left(1 - \rho^n_{\theta s,t+1}\right))\right)$$

Inspecting these formulas yields the general rule: It is optimal to subsidize more the skill type that levels the playing field the most. I.e., it is optimal to subsidize more the skill type that benefits most low ability children and brings them up to speed. This means, subsidize more the skill type with the lowest Hicksian coefficient of complementarity ($\rho^c_{\theta,s}$ or $\rho^n_{\theta,s}$). This is very intuitive: if we are seeking to reduce inequality and improve social mobility, the subsidy should go where it gets the best bang for the buck, namely for the skill level that improves the chances of low ability and low background children the most.

Note that if cognitive and non-cognitive skills go hand in hand, i.e., they are very complementary in the wage function w_t, and if they both level the playing field, then it is optimal to subsidize both equally.

The optimal level of the subsidy for cognitive and non-cognitive skills can be nicely determined from the estimates in the series of aforementioned papers, in particular from Cunha, Heckman, Lochner, and Masterov (2006).

9 Conclusion

This paper studies optimal dynamic income and bequest taxes and education subsidies in a dynastic intergenerational model. Parents can invest in the education of their children and also leave financial bequests. Each generation is subject to idiosyncratic ability and preference shocks. The government aims to provide redistribution and insurance to maximize the expected discounted welfare of the dynasties, from the point of view of the current generation. I derive formulas for the optimal linear and history-independent taxes and subsidies as functions of estimable behavioral elasticities and redistributive factors that are robust to heterogeneities and preferences. I also show how one could make use of existing empirical estimates by deriving the optimal formulas based on variation from existing reforms and that depend on "reform elasticities."

It is in general not optimal to make education expenses fully tax deductible, as education subsidies have differential distributional impacts. Education subsidies, bequest taxes and income taxes can, but need not, optimally move together. Because of their distributional values and finite elasticities to taxes, the optimal education subsidy and bequest tax are generically non-zero. The presence of credit constraints tends to increase optimal education subsidies, reduce optimal income taxes, and leave optimal bequest tax formulas unchanged.

To extend the analysis, the model could naturally be reformulated as a lifecycle model in which a single agent invests in his human capital throughout life. Then, the formulas obtained give us the optimal linear and history-independent lifecycle taxes and subsidies. Another possible reformulation applies to entrepreneurs investing in their business productivity and being subject to linear, history-independent income, savings and investment taxes.

With fully unrestricted, history-dependent taxes, if education is highly complementary to ability, the return to education investments has to be reduced below the return on bequests, or, put differently, education investments by parents will be taxed relative to bequests. It is hence in general optimal to distort the trade-off between parental bequests and education investments because of the redistributive and insurance values of education.

This theoretical research points to two important empirical explorations. First, do parents value the educational achievements of their children *per se*, i.e., do they have a type of "warm glow" preferences for education, or do they rather value the utility of their offspring? Second, it has not yet been convincingly documented how strongly parents react to bequest taxation and education subsidies when choosing between alternative ways of transferring resources to their children.

References

Abramitzky, R., and V. Lavy (2014): "How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?," *Econometrica*, 82(4), 1241–1272.

Benabou, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

Bovenberg, L. A., and B. Jacobs (2005): "Redistribution and education subsidies are Siamese twins," *Journal of Public Economics*, 89(11).

——— (2011): "Optimal taxation of human capital and the earnings function," *Journal of Public Economic Theory*, 13(6), 957–971.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191–261.

Cunha, F., and J. J. Heckman (2007): "Identifying and estimating the distributions of ex post and ex ante returns to schooling: A survey of recent developments," *Labour Economics*, 43(4), 738–782.

——— (2008): "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation," *Journal of Human Resources*, 43(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006): "Interpreting the evidence on life cycle skill formation," *Handbook of the Economics of Education*, 1, 697–812.

Farhi, E., and I. Werning (2010): "Progressive Estate Taxation," *Quarterly Journal of Economics*, 125(2), 635–673.

——— (2013a): "Estate Taxation with Altruism Heterogeneity," *American Economic Review, Papers and Proceedings*, 103(3), 489–495.

Farhi, E., and I. Werning (2013b): "Insurance and taxation over the life cycle," *The Review of Economic Studies*, 80(2), 596–635.

Feldstein, M. (1972): "Distributional Equity and the Optimal Structure of Public Prices," *American Economic Review*, 62(1/2), 32–36.

Findeisen, S., and D. Sachs (2012): "Education and optimal dynamic taxation: The role of income-contingent student loans," Discussion paper, Working Paper Series, Department of Economics, University of Zurich.

——— (2015): "Redistribution and Insurance with Simple Tax Instruments," Working paper, University of Mannheim.

Golosov, M., A. Tsyvinski, and I. Werning (2006): "New dynamic public finance: a user's guide," in *NBER Macroeconomics Annual 2006, Volume 21*, pp. 317–388. MIT Press.

Golosov, M., A. Tsyvinski, and N. Werquin (2014): "A Variational Approach to the Analysis of Tax Systems," Working Paper 20780, NBER.

Hader, S., and K. McGarry (2012): "Parental Investments in College and Later Cash Transfers," Working Paper 18485, NBER.

Heckman, J., L. Lochner, and C. Taber (1998): "Tax Policy and Human Capital Formation," *American Economic Review*, 88(2), 293–297.

Heckman, J. J. (1976a): "Estimates of a human capital production function embedded in a life-cycle model of labor supply," in *Household production and consumption*, pp. 225–264. NBER.

——— (1976b): "A life-cycle model of earnings, learning, and consumption," *Journal of Political Economy*, 84(4), S9–S44.

——— (2011): "The Economics of Inequality: The Value of Early Childhood Education," *American Educator*, pp. 31–47.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond," *Handbook of the Economics of Education*, 1, 307–458.

Heckman, J. J., P. R., and S. P. (2013): "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 2052–2086.

Hendren, N. (2014): "The Policy Elasticity," Working paper, Harvard University.

Hicks, J. (1970): "Elasticity of substitution again: substitutes and complements," *Oxford Economic Papers*, 25, 289–296.

Jacobs, B. (2007): "Optimal Redistributive Tax and Education Policies in General Equilibrium," *International Tax and Public Finance*, 20(2), 312–337.

Jacobs, B., D. Schindler, and H. Yang (2011): "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics*, 114(3), 908–931.

Jacobs, B., and H. Yang (2014): "Second-Best Income Taxation and Education Policy with Endogenous Human Capital and Borrowing Constraints," Working paper, Erasmus University Rotterdam.

Kindermann, F., and D. Krueger (2014): "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk," Discussion paper.

Krueger, D., and A. Ludwig (2013): "Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous," *American Economic Review, Papers and Proceedings*, 103(3), 496 –501.

——— (2014): "Optimal Capital and Progressive Labor Income Taxation with Endogenous Schooling Decisions and Intergenerational Transfers," Working paper, University of Pennsylvania.

Lochner, L. J., and A. Monge-Naranjo (2011): "The Nature of Credit Constraints and Human Capital," *The American Economic Review*, 101(6), 2487–2529.

——— (2012): "Credit Constraints in Education," *Annual Review of Economics*, 4, 225–256.

Mirrlees, J. (1971): "An exploration in the theory of optimum income taxation," *The review of economic studies*, 38, 175–208.

Nordblom, K., and H. Ohlsson (2011): "Bequests, gifts, and education: links between intergenerational transfers," *Empirical Economics*, 40, 343–358.

Piketty, T., and E. Saez (2013a): "Optimal Labor Income Taxation," in *Handbook of Public Economics, Vol. 5*, ed. by M. F. Alan J. Auerbach, Raj Chetty, and E. Saez, pp. 391–474. Amsterdam: North-Holland.

——— (2013b): "A Theory of Optimal Inheritance Taxation," *Econometrica*, 82(4), 1241–1272.

Rogerson, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

Saez, E. (2001): "Using elasticities to derive optimal income tax rates," *The review of economic studies*, 68(1), 205–229.

Saez, E., and S. Stantcheva (2015): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *American Economic Review, forthcoming.*

Sallie Mae (2012): "How America Pays for College 2012," Report, Sallie Mae.

Samuelson, P. A. (1974): "Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," *Journal of Economic Literature*, 12(4), 1255–1289.

Sheshinski, E. (1972): "The Optimal Linear Income Tax," *Review of Economic Studies*, 39(3), 297–302.

Stantcheva, S. (2012): "Optimal Taxation and Human Capital Policies over the Life Cycle," Working paper, Harvard University.

Werquin, N. (2014): "Income Taxation with Frictional Labor Supply," Working paper, Yale University.

Appendix

A Derivation of the fully unrestricted mechanism

Using the notation of the text, let the expenditure function: $\tilde{c}(l, \omega - \beta v, \theta)$ define consumption indirectly as a function of labor l, current period utility ($\tilde{u} = \omega - \beta v$), and the current realization of the type of the dynasty (note that conditional on these variables, consumption does not depend on the human capital of the generation s). Then, $\omega(\theta) = u_t(c(\theta)) - \phi_t(l(\theta)) + \beta v(\theta)$ becomes redundant as a constraint, and the choice variables are $(l(\theta), s(\theta), \omega(\theta), v(\theta), \Delta(\theta))$. Let the multipliers on constraints (27), (28), and (29) be, respectively, $\mu(\theta)$, λ_-, and γ_-. The problem is solved using the optimal control approach where the "types" play the role of the running variable, $\omega(\theta)$ is the state (and $\dot{\omega}(\theta)$ its law of motion), and the controls are $l(\theta), v(\theta), s(\theta)$ and $\Delta(\theta)$. The Hamiltonian is:

$$
\begin{aligned}
&(\tilde{c}(l(\theta), \omega(\theta) - \beta v(\theta), \theta) + s(\theta) - w_t(\theta, s_{-}) l(\theta)) f^t(\theta|\theta_-) \\
&+\frac{1}{R} K(v(\theta), \Delta(\theta), \theta, s(\theta), t+1) f^t(\theta|\theta_-) \\
&+\lambda_- \left[v - \omega(\theta) f^t(\theta|\theta_-)\right] + \gamma_- \left[\Delta - \omega(\theta) \frac{\partial f^t(\theta|\theta_-)}{\partial \theta_-}\right] + \mu(\theta) \left[\frac{w_{\theta,t}}{w_t} l(\theta) \phi_{l,t}(l(\theta)) + \beta \Delta(\theta)\right]
\end{aligned}
$$

with boundary conditions:

$$\lim_{\theta \to \bar{\theta}} \mu(\theta) = \lim_{\theta \to \underline{\theta}} \mu(\theta) = 0$$

Note that taking the first order conditions (hereafter, FOC) of the recursive planning problem yields:

$$[\omega(\theta)] : \left(-\frac{1}{u_t'(c(\theta))} + (\lambda_-) + (\gamma_-) \frac{\partial f^t(\theta|\theta_-)}{\partial \theta_-} \frac{1}{f^t(\theta|\theta_-)}\right) f^t(\theta|\theta_-) = \dot{\mu}(\theta) \tag{40}$$

Integrating this and using the boundary condition $\mu(\bar{\theta}) = 0$, yields:

$$\mu(\theta) = \int_\theta^{\bar{\theta}} \left(\frac{1}{u_t'(c(\theta))} - (\lambda_-) - (\gamma_-) \frac{\partial f^t(\theta|\theta_-)}{\partial \theta_-} \frac{1}{f^t(\theta|\theta_-)}\right) f^t(\theta|\theta_-) \tag{41}$$

Integrating and using both boundary conditions yields:

$$\lambda_- = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta))} f^t(\theta|\theta_-) \, d\theta \tag{42}$$

Using the envelope conditions $\frac{\partial K(v(\theta),\Delta(\theta),\theta,s(\theta),t+1)}{\partial v(\theta)} = \lambda(\theta)$ and $\frac{\partial K(v(\theta),\Delta(\theta),\theta,s(\theta),t+1)}{\partial \Delta(\theta)} = -\gamma(\theta)$, the first-order conditions with respect to $v(\theta)$ and $\Delta(\theta)$ respectively lead to:

$$[v(\theta)] : \frac{1}{u'(c)} = \frac{\lambda(\theta)}{R\beta} \tag{43}$$

and

$$[\Delta(\theta)] : -\frac{\gamma(\theta)}{R\beta} = \frac{\mu(\theta)}{f^t(\theta|\theta_-)} \tag{44}$$

Taking integral of $\dot{\mu}(\theta)$ in equation (40) between the two boundaries, $\bar{\theta}$ and $\underline{\theta}$, and using the boundary conditions $\mu\left(\bar{\theta}\right) = \mu\left(\underline{\theta}\right) = 0$, as well as the expression for λ_{-} from (43), lagged by one period, yields the inverse Euler equation.

The first-order condition with respect to l yields the expression for the optimal labor wedge:

$$[l(\theta)] : \frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} = \frac{\mu(\theta)}{f^t(\theta|\theta_-)} \frac{w_{\theta,t}}{w_t} u_t'(c(\theta)) \left(1 + \frac{l(\theta)\,\phi_{ll,t}(l(\theta))}{\phi_{l,t}(l(\theta))}\right)$$

using the definitions of ε^c, ε^u and ε in the text:

$$\frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} = \frac{\mu(\theta)}{f^t(\theta|\theta_-)} \frac{\varepsilon_{w\theta}}{\theta} u_t'(c(\theta)) \frac{1+\varepsilon^u}{\varepsilon^c}$$

$$[s(\theta)] : 1 - \frac{1}{R} E\left(\frac{\partial w_{t+1}}{\partial s} l_{t+1}\right) + \frac{1}{R} E\left(\mu(\theta_{t+1}) \frac{y_{t+1}}{w_{t+1}} \phi'(l_{t+1}) \frac{\partial w_{t+1}}{\partial \theta_{t+1}} \frac{\partial w_{t+1}}{\partial s} \frac{1}{w_{t+1}^2} (\rho_{\theta s,t+1} - 1)\right) = 0$$

Using the expression for the labor wedge above to substitute for μ, and rearranging, yields the result in Proposition 6.