INTERGENERATIONAL MOBILITY AND OPTIMAL INCOME TAXATION

Maximilian Kasy *

While the empirical literature on intergenerational mobility is politically controversial, it is not obvious what the implications of intergenerational status transmission for optimal policy are. Addressing this question, this paper studies the local comparative statics of optimal income taxes with respect to parameters of intergenerational transmission. The model used extends standard models of optimal linear income taxation, adding a parental preference for child earnings capability, an educational investment opportunity, and credit constraints.

We find that the optimal degree of redistribution, everything else equal, is increasing in the curvature of intergenerational transmission. This is because the non-linearity in the household budget set affects the curvature of household indirect utility as a function of virtual income. In contrast, the implications of stronger transmission for redistribution are ambiguous. The strength of transmission matters, however, for the optimal government budget deficit.

KEyWORDS: Optimal income taxation, intergenerational mobility.

1. INTRODUCTION

A large empirical literature in economics studies the intergenerational transmission of characteristics such as education or earnings capability. Some authors estimate intergenerational correlations, others variance decompositions, others again the causal effect of a particular parental characteristic on their children. Black and Devereux (2011) survey this literature; recent contributions include Lee and Solon (2009), Björklund, Jäntti, and Solon (2005), Behrman and Rosenzweig (2002), and Black, Devereux, and Salvanes (2005). Although this literature contributes to controversial political debates, it is not always clear what are the implications of the various estimates for (optimal) policy. For instance, as Goldberger (1979) pointed out, variation in eyesight might be largely genetically determined. One might conclude that this implies a limited scope for interventions affecting eyesight, yet a hypothetical policy distributing glasses compensating for innate eyesight could be Pareto improving, fully equalizing outcomes, and cheap.

This paper discusses the implications of intergenerational transmission of earnings capability for optimal income taxation. The relevance of particular empirical parameters to optimal policy depends on (i) the underlying model, (ii) the set of policies under consideration, as well as (iii) the evaluation criterion. Graham, Imbens, and Ridder (2008), Chetty (2009), and others have emphasized this point.

Assistant professor, Department of Economics, Harvard University, and junior associate faculty, IHS Vienna. Address: Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138. E-Mail: maximiliankasy@fas.harvard.edu.

*I thank Alan Auerbach, Raj Chetty, Pirmin Fessler, Bryan Graham, Nathaniel Hendren, Susanne Kimm, Zachary Liscow, Dayanand Manoli, Emmanuel Saez, Sarath Sanga, and Juan Carlos Suarez Serrato for valuable comments.
(i) The model developed in this paper extends the model of optimal linear income taxation introduced by Sheshinski (1972), who simplified the model that was introduced by Mirrlees (1971) and discussed more recently by Saez (2001). This model is extended here by allowing child earnings capability to enter parental utility, and by adding an educational production technology which allows parents to invest in their child’s earnings capability. Households are assumed to be credit constrained, similarly to the model in Becker and Tomes (1994). (ii) The policies under consideration are linear income taxes. (iii) The evaluation criterion used is a standard social welfare function, a weighted sum of household utility. In this setup, the empirical intergenerational transmission function, which describes the causal dependence of child earnings capability on parental income, depends on the educational production function. In particular, all else equal, the slope and curvature of the transmission function are increasing in the slope and curvature of the production function.

Our results imply that the optimal degree of redistribution is unambiguously increasing in the concavity (returns to scale) of the educational production function, conditional on household preferences. The slope of the production function has ambiguous implications for taxation, however.\footnote{This result has some similarity to one of the results of Graham, Imbens, and Ridder (2008): They show that mean peer effects are irrelevant when we are interested in the effect of real-locative policies on average outcomes. What matters is the curvature of equilibrium average outcomes as a function of group composition.} As a consequence, learning that the intergenerational production function is more concave then expected should induce a policy maker to increase the progressivity of the tax system, everything else equal, as shown in Proposition 1 in section 4 below. This suggests that empirical researchers should focus on the curvature of transmission in addition to the strength of transmission. Loken, Mogstad, and Wiswall (2012) provide an example of such a study.

The intuition for these results can be summarized as follows. If there are larger returns to educational investments for everyone, this increases the marginal value of additional income for all households. It does not affect the relative marginal value of income, however, which matters for the optimal degree of redistribution. In contrast, if the educational production function shows decreasing returns to scale, this implies that the relative marginal value of income is larger for poorer households compared to the constant returns to scale case, since increasing child earnings capability is relatively cheaper for them. This implies a higher optimal degree of redistribution. The average strength of transmission does matter for taxation, however, if we allow the government deficit to be endogenously chosen in a trade-off between debt and household welfare. Stronger transmission implies a higher optimal deficit, since it increases the marginal value of income to households relative to the marginal cost of government debt.

The crucial feature of the setup studied here is that households are both credit constrained and have an investment opportunity with decreasing returns. This results in a non-linear household budget set, which in turn affects the concavity
of household indirect utility. The results of this paper are not based on an assumption of externalities to parental choices, in contrast to Farhi and Werning (2007), who discuss optimal allocations of consumption across generations and dynasties when social discount rates are lower than the private rates.

The framework of this paper is consistent with a number of different mechanisms of intergenerational transmission which have been suggested in the theoretical literature; Piketty (2000) provides an overview. One of the mechanisms proposed is local segregation in combination with peer effects, discussed for instance by Benabou (1993). Credit constraints affecting human capital investments are reviewed in Bardhan, Bowles, and Gintis (2000). In the sociological literature, Bourdieu and Passeron (1966) have argued for the importance of transmission of cultural capital, such as “speaking the lingo”, and social capital, such as “connections”. All of these can be potentially subsumed under the present framework if we think of parental investment opportunities as including not only child education, but also residential location, or the social and cultural environment of a child.

Comparative statics of optimal linear taxes have been discussed in Helpman and Sadka (1978), comparative statics of nonlinear taxes in Weymark (1987). Loury (1981) discusses a macroeconomic model of intergenerational transmission and earnings inequality in which the absence of credit allows for interventions increasing both average outcomes and equality. Gelber and Weinzierl (2012) consider issues similar to the present paper in the context of a calibrated macro model.

The rest of this paper is structured as follows. Section 2 introduces the model and the planner’s problem. Section 3 illustrates using a parametric example with Cobb-Douglas household utility and a log-linear educational production function. Section 4 presents the main results of this paper: Proposition 1 states that a policy maker should increase redistribution if she learns that transmission is more concave than expected. Subsection 4.1 discusses the local comparative statics of optimal linear taxes with respect to the shape of household indirect utility. Subsection 4.2 characterizes the dependence of household indirect utility on the shape of the educational production function. Subsection 4.3 discusses the relationship between the educational production function and the intergenerational transmission function. It also provides a decomposition of parental valuation of increasing child earnings capacity in a dynastic utility framework. Section 5 concludes; all proofs are relegated to appendix A.

2. THE MODEL

This section discusses first the household problem, which involves an optimal choice of parental consumption, labor supply, and educational investments in their children. This choice is made subject to credit constraints and linear income taxes. Subsection 2.2 then presents the planner’s problem, which involves a choice of the intercept and slope of the linear income tax schedule in order
to maximize the (weighted) average household utility, subject to a government budget constraint.

Throughout this paper, subscripts are used to denote partial derivatives, e.g., $u_c$ denotes the partial derivative of $u$ with respect to $c$. Expectations denote an average across the population of households, similar for covariances.

### 2.1. Households

Consider a population of two-generation households with one working parent. Households are distinguished only by their parental wage $w$. Households choose their level of parental consumption $c$, parental labor supply $1 - l$ (corresponding to leisure time $l$), and the level of educational investment in the child $e$.

Parental disposable income $d$ is given by their labor income $z = w \cdot (1 - l)$ net of linear taxes $t(z) = \alpha + \beta z$:

$$d = z - t(z) = -\alpha + (1 - \beta) \cdot w \cdot (1 - l)$$

Parents allocate their income between their consumption $c$ and the educational investment $e$, where the price of either is normalized to 1. They cannot take credit or save, thus

$$c + e \leq d.$$  

We can rewrite these constraints as

$$c + e + nl \leq y,$$

where $n$ denotes net wages, $n = (1 - \beta) \cdot w$, and $y$ is virtual income (income at zero leisure), $y = n - \alpha$. Child wages $h$ are a function of the educational investment $e$,

$$h = g(e),$$

where $g \in \mathcal{C}^2$ is concave.

Households care about parental consumption $c$, leisure $l$, and their child’s wage $h$, i.e., they maximize the strictly concave utility function $u(c, l, h)$, where $u \in \mathcal{C}^2$. Household indirect utility $v$ is thus equal to

$$v(y) = \max_{c,l,e} u(c, l, g(e))$$

s.t. $c + nl + e \leq y$.

Note that the absence of credit and the possibility of educational investments imply a nonlinear budget set in terms of $(c, l, h)$ for the household.

This setup is the same as the one introduced in Sheshinski (1972) / Mirrlees (1971), and discussed in the subsequent literature, except that (i) household
utility also depends on child welfare, (ii) parents can invest in child earnings capacity, and (iii) households are credit constraint in that they cannot borrow against their child’s future income. Furthermore, we are restricting our discussion to linear taxes, both for ease of exposition and for analytical tractability. Extensions to non-linear taxes are discussed in the conclusion.

2.2. The planner’s problem

The policy objective is given by the utilitarian social welfare function

\[ SWF = E[v] \]

In general, one would want to rescale \( v \) by a (concave) function to reflect distributional preferences. For simplicity, we leave distributional preferences implicit in the definition of \( u \) and \( v \). The government budget constraint is given by

\[ T = E[t(z)] = \alpha + \beta E[z] \geq T_{\text{min}}. \]

Social welfare and the budget constraint are expressed here in terms of averages across households, which is equivalent to summing over households. We will study optimal linear taxes, which maximize social welfare (5) subject to the constraint (6). Subject to (6), we can write the intercept \( \alpha \) of the tax schedule as a function of the marginal tax rate \( \beta \), where the trade-off between these two parameters is given by

\[ \gamma := -\alpha \beta = \frac{T_\beta}{T_\alpha} = \frac{E[z] + \beta E[z_\beta]}{1 + \beta E[z_\alpha]}. \]

If there are no income effects in labor supply, that is if \( l_\alpha = 0 \), then \( E[z_\alpha] = 0 \) and the expression for \( \gamma \) simplifies to \( \gamma = E[z] + \beta E[z_\beta] \). The parameter \( \gamma \) reflects the elasticity of the tax base with respect to tax rates, and thus determines the “cost of redistribution”.

Under the assumption that \((c, l, e)\) are chosen by households to maximize \( u \), household behavioral reactions to changes in tax rates have no first order effect on utility. We thus get \( v_\alpha = -v_y \) and \( v_\beta = -v_y \cdot z \).

The first order condition for optimal linear taxes under the government budget constraint is given by

\[ \frac{T_\beta}{T_\alpha} = \frac{SWF_\beta}{SWF_\alpha}, \]

i.e.,

\[ \gamma = \frac{E[v_y \cdot z]}{E[v_y]}. \]

If there are no income effects on labor supply, this can also be rewritten as

\[ \text{Cov}(v_y, z) = \beta E[z_\beta] \cdot E[v_y]. \]
3. A PARAMETRIC EXAMPLE

In this section, a parametric example of the general model introduced in section 2 is presented. This example allows to derive more explicit expressions for optimal taxes, and provides some intuition for the general results to be discussed in section 4.

Suppose household utility takes the Cobb-Douglas form,

\[ u = c^{\mu} h^{\nu} l^{\xi} \]  

with \( \mu + \nu + \xi = 1 \). Suppose furthermore that the educational production function is log-linear, i.e.,

\[ h = g(e) = \theta_1 e^{\theta_2}, \]

where \( \theta_1 > 0 \) and \( 0 < \theta_2 < 1 \). Under these assumptions, we can rewrite the household problem (4) as

\[
\max_{c,l,e} c^{\mu} (\theta_1 e^{\theta_2})^{\nu} l^{\xi} \\
\text{s.t. } c + e + nl \leq y.
\]

Define the new parameter \( \tilde{\theta} := \mu + \theta_2 \nu + \xi < 1 \). The solution to the household problem is given by

\begin{align*}
    c &= \frac{\mu}{\tilde{\theta}} \cdot y \\
    l &= \frac{1}{n \tilde{\theta}} \cdot y \\
    e &= \frac{\nu \theta_2}{\tilde{\theta}} \cdot y,
\end{align*}

implying that the transmission function and household market income are given by

\[ h = \theta_1 e^{\theta_2} = \theta_1 \left( \frac{\nu \theta_2}{\tilde{\theta}} \right)^{\theta_2} \cdot y^{\theta_2}, \]

\[ z = w(1 - l) = w \left( 1 - \frac{1}{n \tilde{\theta}} \right) = \frac{\alpha}{1 - \beta \tilde{\theta}} + \frac{\mu + \nu \theta_2}{\tilde{\theta}} w, \]

and indirect utility equals

\[ v = \left( \frac{\mu}{\tilde{\theta}} y \right)^{\mu} \left( \theta_1 \left( \frac{\nu \theta_2}{\tilde{\theta}} y \right)^{\theta_2} \right)^{\nu} \left( \frac{\nu}{\tilde{\theta}} y \right)^{\xi} = C_1 \cdot y^{\tilde{\theta}}, \]

where \( C_1 \) is constant across households. Recall that virtual income \( y \) is given by market income at zero leisure, net of taxes: \( y = n - \alpha = (1 - \beta)w - \alpha. \)
The dependence of tax revenues per household, $T = \alpha + \beta E[z]$, on the taxation parameters $\alpha, \beta$ is given by
\[
T_\alpha = 1 + \frac{\beta}{1 - \beta} \xi
\]
\[
T_\beta = E[z] + \beta \frac{\alpha}{(1 - \beta)^2} \xi
\]
This implies a dependence of these two parameters, under the government budget constraint, of
\[
\gamma := -\alpha \beta = \frac{E[z] + \beta E[z\alpha]}{1 + \beta E[z\alpha]} = \frac{E[z] + \beta \frac{\alpha}{1 - \beta^2} \xi}{1 + \beta \frac{\xi}{1 - \beta^2}}.
\]
The effect on a household’s utility of a change in the tax parameters is given by
\[
v_\alpha = -\frac{\mu}{c_1} = -C_2 \cdot \bar{y}^{-1} = -C_2 \cdot y^{(\theta_2 - 1)\nu}
\]
\[
v_\beta = v_\alpha \cdot z = -C_2 \cdot y^{(\theta_2 - 1)\nu} \cdot z,
\]
where $C_2$ is again constant across households.

Suppose the social planner solves
\[
\max SWF = E[v^\delta]
\]
s.t. $T > T^{\text{min}}$.

We have
\[
SWF_\alpha = \gamma E[v_\gamma^{\gamma^{-1}}v_\alpha] = C_3 \cdot E[y^{\delta \delta^{-1}}]
\]
\[
SWF_\beta = \gamma E[v_\gamma^{\gamma^{-1}}v_\beta] = C_3 \cdot E[y^{\delta \delta^{-1}} \cdot z].
\]
Recall that the parameter $\hat{\theta}$ was defined as $\hat{\theta} = \mu + \theta_2 \nu + \xi$, which describes the concavity of indirect utility in this model, and thus
\[
\delta := \hat{\theta} \delta - 1 = (\delta - 1) - \delta \nu (1 - \theta_2).
\]
The first order condition for optimal taxes is now given by
\[
(12) \quad \gamma = \frac{SWF_\beta}{SWF_\alpha} = \frac{E[y^\delta \cdot z]}{E[y^\delta]}.
\]
\[\text{Here we explicitly take distributional preferences into account; } 0 < \delta \leq 1 \text{ is a parameter reflecting increasing relative weight on lower income households as } \delta \text{ decreases.}\]
This expression allows us to gain some intuition for the more general results to follow. Consider comparative statics of optimal taxes with respect to the parameters of intergenerational transmission, $\theta_1, \theta_2$, given the elasticity of the tax base with respect to tax rates, i.e., given $\gamma$, and given household preferences. Equation (12) shows us that optimal taxes are constant in $\theta_1$ given all other model parameters, i.e., optimal taxes are invariant to a proportional rescaling of the educational production function $h = \theta_1 e^{\theta_2}$. This implies that optimal taxes are invariant to a proportional rescaling of the intergenerational transmission function

$$h(y) = \theta_1 \left( \frac{\nu \theta_2}{\theta} \right)^{\theta_2} \cdot y^{\theta_2}.$$ 

In contrast, the right hand side of equation (12) does depend on $\theta_2$, which affects the curvature of the educational production function $g(e)$, the intergenerational transmission function $h(y)$, and thus of the indirect utility function $v(y)$. In fact, a decrease in $\theta_2$ (more concave transmission) has a similar effect to a decrease in $\delta$ (a stronger taste for redistribution).

The reason for this difference between $\theta_1$ and $\theta_2$ becomes apparent when looking at the expression giving a household’s marginal utility of wealth, $v_y = C_2 \cdot y^{(\theta_2 - 1)\nu}$. This expression shows that the relative marginal utility of wealth between different households does depend on $\theta_2$, but not on $\theta_1$. The average marginal utility of wealth does depend on $\theta_1$, through $C_2$, however. The average slope of intergenerational transmission might thus have an impact on a potential trade-off between government debt and household welfare.

This discussion implies that a policymaker who believes in the parametric model of this section should increase redistribution when she learns that the curvature of transmission is larger than expected. She should not, however, react to news about the average strength of transmission. Proposition 1 below formalizes and generalizes these claims. There are some features of the Cobb-Douglas setup which considerably simplify the relationship between transmission and taxes; in particular the share of income devoted to education is constant in income. That said, the main claims we made do generalize as shown in the following section.

4. COMPARATIVE STATICS

In this section the main results of the present paper are discussed. We are interested in the dependence of optimal income taxes on parameters of intergenerational transmission. Analysis of this dependence can be decomposed into several parts: (i) The dependence of optimal taxes on the shape of the indirect utility function. (ii) The dependence of the shape of the indirect utility function on the shape of the household budget set, i.e., the educational production function. (iii) The relationship between the shape of the educational production function and the intergenerational transmission function. These questions are
discussed in turn in the following three subsections.

Proposition 1 summarizes some of following results. This proposition generalizes the conclusions of the parametric model of section 3.

**Proposition 1 (Transmission and a policymaker’s choices)**

Consider a policymaker who maximizes expected social welfare under the assumptions of section 2. Suppose furthermore that her beliefs about household preferences and about labor supply are not affected by learning about intergenerational transmission. Then, if she learns that

1. the curvature of educational production $g_{ee}$ is more negative than expected, her choice of $\beta$ will increase.
2. the slope of educational production $g_e$ is larger than expected, the effect on $\beta$ is ambiguous.
3. the curvature of intergenerational transmission $h_{yy}$ is more negative than expected, while her believes about $h_y$ are unchanged, her choice of $\beta$ will increase.
4. the slope of transmission $h_y$ is larger than expected, the effect on $\beta$ is ambiguous.

The following subsections provide the formal basis for these claims. Subsection 4.1 characterizes the local comparative statics of optimal linear income taxes with respect to the indirect utility function $v(y)$. It is shown that the optimal marginal tax rate $\beta$ is decreasing in $\text{Cov}(v_y, z)$, the covariance of the marginal utility of wealth and of market income. A proportional rescaling of $v$ does not affect optimal taxes. The optimal budget deficit, in contrast, is increasing in $E[v_y]$, the average marginal utility of wealth.

Subsection 4.2 characterizes the dependence of $v_y$, the marginal utility of wealth, and of $v_{yy}$, the curvature of the indirect utility function, on the educational production function $g$. The results are specialized, in particular, to the case of separable household utility, $u(c, l, h) = u^\rho(c, l) + u^\zeta(h)$, which yields more tractable and interpretable expressions. Separability holds in particular in the context of the dynastic utility setup discussed in subsection 4.3. Under this separability assumption, it is shown in subsection 4.2 that $v_{yy}$ is an increasing function of $(u^\zeta \circ g)_{ee}$, the curvature of utility in educational investments. This curvature (concavity) in turn is increasing in $g_{ee}$, the curvature of the educational production function. The concavity of $v$ also increases in $g_e$, the returns to education. It is shown furthermore that a decrease in $v_{yy}$ for all values of $y$ decreases $\text{Cov}(v_y, z)$, which in turn increases the optimal marginal tax rate by the results of subsection 4.1. Put differently, more strongly decreasing returns to scale in educational investments imply higher optimal redistribution.

Subsection 4.3, finally, relates the intergenerational transmission function $h(y)$, giving child wages as a function of parental income, to the educational production function $g(e)$. Furthermore, subsection 4.3 presents a decomposition of $u^\zeta(h)$, the value of child wages, in a dynastic utility framework. This framework assumes households maximize the sum of the discounted utility of all their descendants.
4.1. Optimal linear taxes and indirect utility

The following lemmas characterize the local comparative statics of the optimal linear tax parameters $\alpha^*, \beta^*$ with respect to the indirect utility function $v$. Since $v$ is an infinite-dimensional parameter, derivatives with respect to $v$ are potentially technically tricky objects. This difficulty is easily circumvented by considering directional derivatives: Index $v$ by a one-dimensional parameter $\theta$, so that $v = v(y, \theta)$, and consider derivatives with respect to $\theta$. By the chain rule, this gives the directional derivative with respect to $v$ in the direction $v_\theta$.

The partial derivatives in lemma 1 and 2 are taken with respect to $v$, holding the labor supply function, and thus the elasticity of the tax base with respect to tax changes, constant. Put differently, we will characterize the derivative, with respect to $\theta$, of the solution of

$$\gamma = \frac{E[v_y(y, \theta) \cdot z]}{E[v_y(y, \theta)]}$$

$$T_{\min} = \alpha + \beta E[z]$$

taking $\gamma$, and $z$ as a function of $w$, as constant in $\theta$. This makes sense in the context of the question motivating this paper: The labor supply elasticity, and thus $\gamma$, are fairly well studied objects in the empirical public finance literature. We are ultimately interested in what knowledge of the intergenerational transmission function would imply for optimal taxation, given knowledge of $\gamma$.

**Lemma 1 (Local comparative statics of optimal linear taxes)**

Let $(\alpha^*, \beta^*)$ be the tax schedule maximizing SWF (5) subject to the budget constraint (6), and assume the second order condition is fulfilled at the optimum. The local comparative statics of optimal taxes with respect to a parameter $\theta$ indexing indirect utility $v$ are given by

$$\alpha_{\theta}^* = -\gamma \cdot C \cdot E[v_y(y, \theta) \cdot (-z + \gamma)]$$

$$\beta_{\theta}^* = C \cdot E[v_y(y, \theta) \cdot (-z + \gamma)]$$

where $C > 0$. If $E[z_\alpha] = 0$, equation (14) can be rewritten as

$$\beta_{\theta}^* = C \cdot (\text{Cov}(v_y(y, \theta), -z) + \beta E[z_\beta] \cdot E[v_y(y, \theta)])$$

This lemma tells us in particular that the optimal marginal tax rate $\beta^*$ is increasing in the covariance Cov($v_y, -z$), given the average strength of the effect of income on the welfare of future generations, $E[v_y]$. It also tells us that taxes are invariant under proportional increases of $v$: At the optimum,

---

3Econometricians are familiar with this approach from the literature on semi-parametric efficiency bounds, where influence functions are derived as the dual representations of directional derivatives of parameters of interest with respect to the (infinite dimensional) observable-data distribution. See, for instance, Tsiatis (2006) or Newey (1994).
MOBILITY AND TAXATION

11

\[ E[y(-z+\gamma)] = 0. \] Thus, if \( v_{y,\theta} \) is proportional to \( v_y \), we have \( \alpha^*_\theta = \beta^*_\theta = 0 \)

Lemma 1 takes government revenues \( T^{\text{min}} \) as given. Let now \( \lambda \) be the marginal value of government revenues, expressed in the same units as household utility. Consider the modified problem of optimizing the following social welfare function, which endogenizes the choice of the budget deficit or surplus, \( T^{\text{min}} \):

\[ SWF^\lambda = SWF + \lambda T = \mathbb{E}[v(y)] + \lambda \mathbb{E}[t(z)] \]

The trade-off embodied in the parameter \( \lambda \) involves the relative value of directly increasing current generation household income relative to the value of other government expenditures, or to reduction of government debt. Conditional on \( T \), the problem of maximizing \( SWF^\lambda \) is again the same as that of maximizing \( SWF \) (5) subject to the government budget constraint (6). The first order condition for optimal marginal taxes \( \beta \) when maximizing \( SWF^\lambda \), \( SWF^\lambda_\beta = 0 \), implies

\[ \mathbb{E}[v_{y} \cdot z] = \lambda \cdot (\mathbb{E}[z] + \beta \mathbb{E}[z_\beta]) . \]

The optimal minimum guaranteed income (or poll tax) \( \alpha \) satisfies \( SWF^\lambda_\alpha = 0 \), and hence

\[ \mathbb{E}[v_y] = \lambda \cdot (1 + \beta \mathbb{E}[z_\alpha]) . \]

Taking ratios, we again get the first order condition (7). The following lemma characterizes the comparative statics of the optimal budget surplus (deficit) \( T^* \) with respect to \( v \).

**Lemma 2 (Local comparative statics with endogenous budget deficit)**

Let \( T^* \) be the optimal budget surplus corresponding to the tax schedule maximizing \( SWF^\lambda \), and assume the second order condition is fulfilled at the optimum. Let \( (\alpha^*, \beta^*) \) be the tax schedule maximizing \( SWF^\lambda \) given \( T \). The local comparative statics of the optimal budget surplus \( T^* \) with respect to a parameter \( \theta \) indexing changes in indirect utility \( v \) are given by

\[ T^*_\theta = C \cdot [\text{Cov}(v_{y,\theta}, -z) \cdot \beta^*_T - \mathbb{E}[v_{y,\theta}] \cdot (\alpha^*_T + \beta^*_T \mathbb{E}[z])] , \]

where \( C > 0 \), \( \alpha^*_T > 0 \), and \( \beta^*_T > 0 \). This lemma tells us that the optimal \( T^* \) is decreasing in \( \mathbb{E}[v_y] \) given \( \text{Cov}(v_y, -z) \), and that it is increasing in \( \text{Cov}(v_y, -z) \) given \( \mathbb{E}[v_y] \). In particular, a higher average marginal utility of wealth implies a higher optimal budget deficit. Put differently, income subsidies become more valuable relative to other government expenditures if \( \mathbb{E}[v_y] \) increases.
4.2. Indirect utility and nonlinear budget constraints

The last subsection characterized the dependence of optimal taxes on the indirect utility function \( v \). This subsection characterizes the dependence of the indirect utility function on the underlying structural objects, and in particular on the educational production function \( g \).

Lemma 3 gives explicit expressions for \( v_{yy} \) and \( v_{y} \). Lemma 4 then characterizes the dependence of \( v_{yy} \) and \( v_{y} \) on the educational production function \( g \). Similar to the last subsection, we are indexing \( g \) by some parameter \( \theta \), i.e., \( h = g(e, \theta) \), and then take the derivative of \( v_{y} \) and \( v_{yy} \) with respect to \( \theta \). Both lemma 3 and lemma 4 give results first for a general utility maximization problem with linear budget sets, second for a utility maximization problem with non-linear budget sets, and finally specialize to the setup introduced in section 2, imposing the additional separability assumption \( u(c, l, h) = u^{p}(c, l) + u^{c}(h) \). Lemma 5, finally, connects the results on the comparative statics of \( v_{yy} \) to the covariance \( \text{Cov}(v_{y}, z) \), showing that an increase in the concavity of \( v \) decreases this covariance.

**Lemma 3 (Concavity of indirect utility)**

Assume that \( u \in C^2 \) is a strictly concave function from \( \mathbb{R}^k \) to \( \mathbb{R} \).

1. Let \( v(y) = \max_{x} u(x) \) subject to \( x \cdot p \leq y \), where \( y \in \mathbb{R} \) and \( x, p \in \mathbb{R}^k \). Then

\[
v_{y} = u_{x} x_{y} = u_{x1}/p^{1} = u_{x2}/p^{2} = \ldots = u_{xk}/p^{k}\]

and

\[
v_{yy} = \frac{1}{p' u_{xx}^{-1} p}.
\]

2. Let \( v(y) = \max_{x} u(g(x)) \) subject to \( x \cdot p \leq y \), where \( y \in \mathbb{R} \), \( x, p \in \mathbb{R}^k \), and \( g \) is a strictly concave function from \( \mathbb{R}^k \) to \( \mathbb{R}^k \). Then

\[
v_{y} = u_{g} g_{x} x_{y} = u_{g} g_{x1}/p^{1} = u_{g} g_{x2}/p^{2} = \ldots = u_{g} g_{xk}/p^{k}\]

and

\[
v_{yy} = \frac{1}{p' (g_{x} u_{g} g_{x} + u_{gg} g_{xx})^{-1} p}.
\]

Note that in the last expression \( g_{xx} \) is a three dimensional array.

3. Suppose \( v(y) = \max_{x, e} u^{p}(x) + u^{c}(g(e)) \) subject to \( x \cdot p + e \leq y \), where \( y, e \in \mathbb{R} \), \( x, p \in \mathbb{R}^k \), and \( g, w \) are strictly concave functions from \( \mathbb{R} \) to \( \mathbb{R} \). We could take, for instance, \( x = (c, l) \), and \( p = (1, n) \). Then

\[
v_{y} = u_{x}^{p} x_{y} + u_{g}^{c} g_{e} e_{y} = u_{x1}^{p}/p^{1} = u_{x2}^{p}/p^{2} = \ldots = u_{g}^{c} g_{e}\]

and

\[
v_{yy} = \frac{1}{p' u_{xx}^{-1} p + 1 / (u^{c} g_{ee})}.
\]
MOBILITY AND TAXATION

13

The proof of this lemma is based on the fact that the curvature of $v$ is given by the curvature of $u \circ g$ in the direction of the Engel curve $x(y)$. The latter is given by $x_y = v_{yy} u_{xx}^{-1} p$ in the general consumer problem.

Lemma 4 (Comparative statics of the concavity of indirect utility)

1. Let $v(y) = \max_x u(x, \theta)$ subject to $x \cdot p \leq y$, where $y \in \mathbb{R}$, $x, p \in \mathbb{R}^k$, and $u \in \mathcal{C}^2$ is strictly concave in $x$. Then

$$v_y = u_{x \theta} x_y$$

and

$$v_{yy} = \frac{p' u_{xx}^{-1} u_{xxx \theta} u_{xx}^{-1} p}{(p' u_{xx}^{-1} p)^2}.$$  

2. Suppose $v(y) = \max_x u^p(x) + u^c(g(e, \theta))$ subject to $x \cdot p + e \leq y$, where $y, e \in \mathbb{R}$, $x, p \in \mathbb{R}^k$, and $u^p, g, u^c \in \mathcal{C}^2$ are strictly concave. We could have, for instance, $x = (c, l)$, and $p = (1, n)$. Then

$$v_y = (u^c \circ g)_{e \theta} \cdot e_y$$

and

$$(19) \quad v_{yy} = v_{yy}^2 \cdot \frac{(u^c \circ g)_{ee \theta}}{(u^c \circ g)_{ee}^2}.$$  

Comparing the results of this lemma to those of lemma 3, we see that the expressions for $v_{yy \theta}$ are obtained from those for $v_{yy}$ by direct differentiation of $u$ or $g$ with respect to $\theta$. The reason for this is that behavioral reactions to changes in $\theta$ can be ignored due to a differentiated version of the first order conditions for household optimal behavior.

The crucial result here for our purposes is equation (19). It shows that the curvature of household indirect utility $v$ is increasing in the curvature of $(u^c \circ g)$, which in turn is increasing in the curvature of the educational production function $g$.

Lemma 5 (Comparative statics of expectations and covariances)

1. Suppose $v(y, \theta) \in \mathcal{C}^2$, and $z$ is a strictly monotonically increasing function of $y$. Suppose $\theta$ is a constant, while $y$ has some cross-sectional distribution. Then, if $v_{yy} < 0$ for all $y$, then $\text{Cov}(v_{y \theta}, z) < 0$.

2. Suppose in particular that $v(y) = \max_x u^p(x) + u^c(g(e, \theta))$ subject to $x \cdot p + e \leq y$, where $y, e \in \mathbb{R}$, $x, p \in \mathbb{R}^k$, and $u^p, g, u^c \in \mathcal{C}^2$ are strictly concave.

Then, if $(u^c \circ g)_{ee \theta} < 0$ for all $e$, then $\text{Cov}(v_{y \theta}, z) < 0$.

If $g_{ee \theta} > 0$ for all $e$, then $E[v_{y \theta}] > 0$.  

This lemma tells us, first, that a more concave indirect utility function $v$ implies a more negative covariance between the marginal utility of wealth and household market income. It then tells us, building on lemma 4, that a more concave education production function implies a more negative such covariance, and that higher returns to education imply higher marginal utility of wealth.

4.3. Production function, transmission function, and the utility of future generations

The intergenerational transmission function is given by $h(y) = g(e(y))$, where $e(y)$ is the Engel curve describing the dependence of educational investments on virtual income. How does the shape of $h(y)$ relate to the shape of $g$? Differentiation yields

$$h_y = g_e e_y$$

$$h_{yy} = g_{ee} e_y^2 + g_e e_{yy}. \tag{20}$$

Given the Engel curve $e(y)$, the slope of the transmission function is increasing in the slope of the educational production function $g(e)$. The curvature of the transmission function is increasing in the curvature of $g$ and has an ambiguous dependence on the slope of $g$, depending on the curvature of $e(y)$. Recall that in the parametric example of section 3 the transmission function was given by $h(y) = \theta_1 \left( \frac{w_0}{\sigma} \right)^{\theta_2} \cdot g^{\theta_2}$, while the production function was $h = \theta_1 e^{\theta_2}$, and the Engel curve for $e$ was $e = \frac{\nu \theta_2}{\sigma} \cdot y$. In this example, the slope and curvature of the production function and of the transmission function are proportional because the Engel curve is linear. This is a special feature of models with homothetic preferences.

The following lemma characterizes the Engel curve for $e$ in the more general setup.

Lemma 6 (Engel curve for educational investments)

Suppose $e$ is part of the solution of (4), where

$$u(c, l, h) = u^p(c, l) + u^e(h).$$

Then the slope of the Engel curve for educational investments is given by

$$e_y = \frac{(u^e \circ g)_{ee}}{p' u^p_{xx} p + \frac{1}{(u^c \circ g)_{xx}}}, \tag{21}$$

where $p = (1, n)$ and $x = (c, l)$. 

So far we have taken the parental utility function \( u(c, l, h) \) as given; some of the results in the last subsection were imposing separability \( u(c, l, h) = u^p(c, l) + u^c(h) \). We can decompose this further, assuming that households maximize dynastic utility

\[
\begin{align*}
u(c, l, e) &= u^p(c, l) + u^c(g(e)) \\
&= \sum_{t=0}^{\infty} R^t \cdot u^p(c_t, l_t)
\end{align*}
\]

subject to the sequence of constraints

\[
\begin{align*}
c^t + (1 - \beta)w^t l^t + e^t &\leq (1 - \beta)w^t - \alpha \\
w^{t+1} &= g(e^t).
\end{align*}
\]

This yields the following Bellman equation:

\[
\begin{align*}
u^c(w) &= R \cdot \max_{c, l, e} \left( u^p(c, l) + u^c(g(e)) \right) \\
&\text{s.t. } c + (1 - \beta)w l + e \leq (1 - \beta)w - \alpha.
\end{align*}
\]

Differentiating, and using the household first order condition to eliminate some terms, yields the following lemma.

**Lemma 7 (The value of child wages)**

Suppose \( u^c \) is given by equation (24). Then

\[
\begin{align*}
u^c_w(w) &= R \cdot (1 - \beta) \cdot (u^p_c c_w + u^p_l l_w + u^c_w g_e e_w) \\
&= R \cdot (1 - \beta) \cdot u^p_c \\
&= R \cdot (1 - \beta) \cdot u^p_c / n \\
&= R \cdot (1 - \beta) \cdot u^c_w g_e
\end{align*}
\]

At a steady state value \( w \), satisfying \( w = h = g(e((1 - \beta)w - \alpha)) \), we have in particular

\[
\begin{align*}
g_e &= \frac{1}{R \cdot (1 - \beta)}
\end{align*}
\]

and

\[
\begin{align*}
u^c_w &= \frac{R \cdot (u^p_c c_w + u^p_l l_w)}{1 - R \cdot g_e e_w}.
\end{align*}
\]

Equation (27) is particularly instructive: The marginal value of an increase in household wage corresponds to the marginal value of the resulting increase in consumption and leisure for that household, times a multiplier. This multiplier
in turn is a function of the slope of the transmission function times the discount rate, where the multiplier is increasing in the strength of transmission. This implies that the value of interventions (beyond the scope of our model) increasing household wages is increasing in the strength of transmission.

Equation (26) states the households first order condition for the intertemporal tradeoff in steady state: a reduction in parental consumption by one unit increases child virtual income by \( g_e \cdot R \cdot (1 - \beta) \). Since in steady state the value of the two is equalized, we have \( g_e \cdot R \cdot (1 - \beta) = 1 \).

5. CONCLUSION

Methodological debates in microeconometrics tend to focus on questions of “what we can get” with given data and assumptions (identification), and “what we want” (parameters of interest). Both are particularly controversial in the politically charged field of intergenerational mobility. This paper focuses on the latter question: what empirical parameters are actually relevant? One way to give a systematic answer to that question is to consider what implications empirical results might have for a choice between policies.

The answer depends on the model assumptions we make, the policies we consider, and the evaluation criterion we use. This paper considers a model which is a modification of the model in Sheshinski (1972) / Mirrlees (1971), adding a parental preference for child earnings capability, an educational investment opportunity, and credit constraints. The policies considered are linear income taxes satisfying a government budget constraint, the evaluation criterion is a utilitarian social welfare function. In this setup, the comparative statics of the optimal degree of redistribution with respect to the mechanism of intergenerational transmission are considered.

The optimal degree of redistribution equalizes the relative cost of marginal income for different households with the relative social valuation of additional income for these households. The relative cost is a function of the elasticity of the tax base with respect to tax rates. The relative benefit depends on the planner’s preferences, i.e., how much weight she puts on additional utility of households of different levels of utility. It also depends on the shape of the household budget set, i.e., how households can translate additional income into valued goods. In particular, suppose households face an investment opportunity with decreasing returns to scale, say into child education, and are credit constraint. Then the relative marginal valuation of additional income for lower income households is increased, compared to the case where households face a linear budget set. This in turn implies a higher optimal degree of redistribution. This is the main finding of the present paper: That the optimal degree of redistribution depends on the curvature of the transmission function rather than its slope.

There are several promising ways for future research to build on the results of this paper. First, it would be interesting to conduct more empirical research on the relative strength of transmission for households of different income. Second,
one could extend the theoretical results of this paper to the case of non-linear income taxes, building on the methods of Weymark (1987). Third, the setup considered here could be extended to be more directly applicable to data, allowing in particular for more heterogeneity of utility, returns to education etc. Finally, there are many other relevant fields of policy that might relate to intergenerational transmission (educational policy, urban planning, financial regulation, ...). It would be interesting to see more formal studies of the implications of intergenerational transmission for optimal policy in these fields.

APPENDIX A: PROOFS

Proof of Proposition 1
The first two claims follow from the results of Lemma 1 and Lemma 5 below. Claim 3 and 4 follow from the first two claims, since $h_{yy} = g_y e_y^2 + g_y e_{yy};$ c.f. section 4.3. □

Section 4.1

Proof of Lemma 1:
Let $\widehat{SWF}(\beta)$ be the social welfare function $SWF$ as a function of $\beta$, taking $\alpha$ as a function of $\beta$ under the budget constraint. Thus

$$\widehat{SWF}(\beta) = E[v_y \cdot (-z - \alpha \beta)] = E[v_y \cdot (\gamma - z)].$$

By the implicit function theorem,

$$\beta_\theta = -\widehat{SWF}^{-1}_{\beta,\beta} \cdot \widehat{SWF}_{\theta,\beta}.$$

By assumption the second order condition of the planners problem is fulfilled, so that

$$\widehat{SWF}_{\beta,\beta} < 0.$$

Furthermore,

$$\widehat{SWF}_{\theta,\beta} = E[v_y,\theta (-z + \gamma)].$$

Setting $C := -\widehat{SWF}^{-1}_{\beta,\beta}$ yields the claim □

Proof of Lemma 2:
Let $(\alpha^*, \beta^*)$ be the tax schedule maximizing $SWF$ (5) subject to the budget constraint (6), that is given $T$. Let now $\widehat{SWF}_T^\lambda (T)$ be the result of substituting $(\alpha^*, \beta^*)$ into $SWF^\lambda$, defined in equation (15). Maximizing $\widehat{SWF}_T^\lambda (T)$ with respect to $T$ gives the tax-schedule maximizing (15).

We get the first order condition for optimal $T^*$,

$$\widehat{SWF}_T^\lambda = -E[v_y(z \cdot \beta_T^* + \alpha_T^*)] + \lambda,$$

and hence

$$T_\theta^* = -\widehat{SWF}^{-1}_{T,T} \cdot \widehat{SWF}_{T,\theta}.$$

By assumption, again, second order conditions for optimal taxes are satisfied, so that $\widehat{SWF}_{T,T}^\lambda < 0.$ Furthermore,

$$\widehat{SWF}_{T,\theta}^\lambda = -E[v_y,\theta (z \cdot \beta_T^* + \alpha_T^*)].$$
Rearranging gives
\[
T_\theta = -\text{SWF} E_{\theta,T}^{\lambda^{-1}} \cdot \left[ \text{Cov}(v_{y,\theta} - z) \cdot \beta_T^\tau - E[v_{y,\theta}] \cdot (\alpha_T^\tau + \beta_T^\tau E[z]) \right]
\]
The claim now follows if we can show that \(0 < \beta_T^\tau\) and \(0 < \alpha_T^\tau\). Note that \(\beta_T^\tau\) is characterized by
\[
-E[v_{y,\theta} - z] + \lambda \cdot (E[z] + \beta E[z_{\beta}]) = 0
\]
where \(\lambda\) is the Lagrange multiplier on the government budget constraint. Increasing \(T\) in this problem is equivalent to increasing the multiplier \(\lambda\). \(\beta_T^\tau > 0\) follows by again applying the implicit function theorem, assuming that the second order condition holds, and noting that \((E[z] + \beta E[z_{\beta}]) > 0\). A similar argument shows \(\alpha_T^\tau > 0\).

Section 4.2

Proof of lemma 3:
1. The first order condition for constrained optimization is \(u_x = \lambda p\), where \(\lambda\) is a Lagrange multiplier which depends on \(y\) as we vary the constraint. We get \(v_y = u_x y = \lambda p x y = \lambda\), since \(x p = y\) and thus \(p x y = 1\).
2. Differentiating \(v_y = u_x x y\) gives \(v_{yy} = x' y u_{xx} x y + u_x x y y = x' y u_{xx} x y\), because \(u_x x y y = \lambda p x y y = 0\).
3. Differentiating the first order condition gives \(u_x x y = \lambda y p = v_{yy} p\), hence \(x y = v_{yy}^{-1} p\). Plugging this into the first expression for \(v_{yy}\) gives \(v_{yy} = x' y u_{xx} x y = v_{yy}^2 p u_{xx}^1 u_{xx} u_{xx}^{-1} p = v_{yy}^2 p u_{xx}^{-1} p\). Solving for \(v_{yy}\) yields the claim.
4. This immediately follows from 2, noting that \((u o g)_{xx} = g'_{x} u_{xy} g_{x} + u_{xx} g_{xx}\), where the last term is a tensor product of a vector and a three dimensional array.
5. This again follows from 2, replacing \(x\) by \((x, e)\), \(u\) by \((u^p, u^c o g)\) and \(p\) by \((p, 1)\), and noting that \((u^p(x), u^c(g(e))(x,e))(x,e)\) is block diagonal due to the assumed additive separability.

Proof of lemma 4:
1. By differentiation, \(v_{y,\theta} = u_{y,\theta} + u_{x,\theta} x y\). By the constraint, \(x y p = 0\); by the first order condition for optimal \(x\), \(u_x = \lambda p\), thus \(v_y = u_{y,\theta}\). Differentiating by \(y\) yields the first claim, \(v_{y,\theta} = u_{x,\theta} x y\).
2. Differentiating \(y,\theta\) again yields \(v_{y,\theta} = x' y u_{xx} x y + u_{x,\theta} x y y = x' y u_{xx} x y\). The last equality holds because \(u_x x y y = (\lambda p) p x y y = \lambda p p x y y = \lambda p y y y = 0\).
3. Recall now from the proof of lemma 3 that \(x y = v_{yy}^{-1} p\); plug this in to get \(v_{y,\theta} = v_{y,\theta}^2 p u_{xx} u_{xx}^{-1} p\). Using the result \(v_{yy} = \frac{1}{y' u_{xx} p}\) yields the claim.
4. This is a special case of the previous result. In particular, \(v_{y,\theta} = (u^c o g)_{y,\theta}\), from which the first claim follows. As for the second claim, note that the second derivative of the objective function with respect to \((x, e)\) is block diagonal, which implies that the numerator (corresponding to \(p' u_{xx} u_{xx} u_{xx}^{-1} p\) in the previous result), is given by \(\frac{(u^c o g)_{y,\theta}}{(u^c o g)_{x,\theta}}\). The claim follows.

Proof of lemma 5:
1. Under the given assumptions, \(v_{y,\theta}\) is a strictly decreasing function of \(y\). We thus have to show the general result \(\text{Cov}(a(y), b(y)) > 0\) for strictly increasing functions \(a, b\) of \(y\) and nondegenerate distribution of \(y\). To see this, consider two independent draws \(y^1, y^2\) from the distribution of \(y\). Then \(\text{Cov}(a(y^1) - a(y^2), b(y^1) - b(y^2)) = 2 \text{Cov}(a(y), b(y))\) by independence. Furthermore, by monotonicity \(a(y^1) > a(y^2)\) if and only if \(b(y^1) > b(y^2)\), thus \(\text{Cov}(a(y^1) - a(y^2), b(y^1) - b(y^2)) > 0\).
2. This follows from the first result, using lemma 4. In particular, since \((u^c \circ g)_{ee\theta} < 0\), we have \(v_{yy\theta} = v_{yy}^2 (u^c \circ g)_{ee\theta} < 0\). Furthermore, if \(g_{e\theta} > 0\), then \(v_{y\theta} = (u^c \circ g)_{e\theta} \cdot e_y > 0\).

\[ \square \]

Section 4.3

**Proof of Lemma 6:**
Consider first the more general problem \(\max_x u(x)\) subject to \(xp \leq y\). Recall from the proof of lemma 3 that in this case \(u_x = v_y p\) at the optimum. Differentiating w.r.t. \(y\) yields \(u_{xx} x_y = v_{yy} p\), so that

\[ x_y = v_{yy} u_{xx} p = \frac{u_{xx} p}{p' u_{xx} p}. \]

The claim now follows from replacing \(u(x)\) by \(u^b(c, l) + u^c(g(c))\), plugging in the special-case expression for \(v_{yy}\) from lemma 3, and using again the block diagonality of \(u_{xx}\).

\[ \square \]

**Proof of Lemma 7:**
These claims follow immediately from differentiation of the Bellman equation (24), once we note that the first order condition for optimal household choices implies that changes of the choice variables consistent with the constraint have no first order effect on the maximized value.

\[ \square \]

REFERENCES


