Pareto-Improving Education Subsidies with Financial Frictions and Heterogenous Ability

Dilip Mookherjee and Stefan Napel[1]

Version: August 28, 2017

Abstract

Previous literature has shown that missing credit or insurance markets do not necessarily provide an efficiency-based argument for educational subsidies, starting with a laissez faire competitive equilibrium. With heterogeneity in learning ability, we show that an efficiency rationale does exist. In an OLG economy with ability heterogeneity and missing financial markets, we prove: (a) every competitive equilibrium is interim Pareto dominated (and ex post Pareto dominated with sufficient parental altruism) by a policy providing education subsidies financed by income taxes, and (b) transfers conditional on educational investments similarly dominate unconditional transfers. The policies also result in macroeconomic improvements (higher per capita income and upward mobility).

[1]Boston University and University of Bayreuth respectively. We acknowledge helpful comments by Roland Bénabou, participants of the Oslo THRED conference in June 2013, the Ascea Summer School in Development Economics 2014, the Verein für Socialpolitik annual conference 2014 and the Public Economic Theory conference 2015 as well as seminar participants at Boston University, Michigan State University, University of Bayreuth, University of Münster and University of Turku. We thank Brant Abbott for providing us with details of the nature of inter vivos transfers in Abbott *et al.* (2013). We are also grateful to Marcello d'Amato for numerous discussions and ideas concerning occupational choice and role of the welfare state.

1 Introduction

A central issue in discussions of a welfare state concerns its normative rationale, when incomplete financial markets restrict the ability of households to finance education of their children, or insure against idiosyncratic risks. Arguments for provision of social insurance by the government have frequently been justified on the basis of notions of ex ante efficiency in an 'original' position prior to the realization of idiosyncratic ability or income shocks. Since such social insurance involves redistributive income or wealth taxes, such a consensus no longer holds once ability and income shocks have been realized. However, we typically observe relatively less disagreement across income classes or political parties with varying ideologies concerning the need for public education subsidies, particularly for children from poor families.

Existing models of human capital accumulation in environments with financial frictions however do not provide a clear rationale for interim or ex post efficiency enhancing role of educational subsidies. Mookherjee and Ray (2003) show the existence of fully (i.e., first-best) efficient equilibrium steady states in an OLG occupational choice model without ability heterogeneity or income risk, where parents have a Barro-Becker bequest motive, and cannot borrow to finance their children's education. With a continuum of occupations, their model has a unique steady state which turns out to be fully efficient. In a two period model with uninsurable income risk but no borrowing constraints or ability heterogeneity, Jacobs, Schindler and Yang (2012) show that the ex ante efficient optimal subsidy on education is zero, starting from a laissez faire competitive equilibrium.

In this paper we show there is a Pareto-improving role for educational subsidies in the presence of ability heterogeneity. Starting with a laissez faire competitive equilibrium of an OLG economy with missing credit and insurance markets, we show that there exists a policy of educational subsidies financed by income taxes which generates an interim Pareto improvement at the stage where parents have not yet learnt the ability realization of their children. If parents exhibit sufficient altruism, such a policy is also ex post Pareto improving (i.e., even after children abilities have been observed). These policy interventions induce macroeconomic improvements, i.e., higher per capita education, income and upward mobility. We also show that any progressive fiscal policy not involving educational subsidies is dominated (both in the Pareto and macro sense) by a policy involving educational subsidies. This implies transfers conditioned on educational enrollment of children dominate unconditional transfers, and educational subsidies are a necessary feature of any fiscal policy which maximizes a utilitarian welfare function, given ability heterogeneity and financial frictions.

The basic logic is illustrated most simply in a model with two occupations (skilled and unskilled), and two levels of education (zero or one). Agents either do or do not receive an education, which is necessary to enter the skilled occupation. Parents pay for their children's education and cannot borrow for this purpose. The amount they need to spend is decreasing in their child's learning ability, which is drawn from an iid distribution. They decide whether or not to educate their child after observing the latter's ability. Wages in each occupation in any given generation depend on the relative supply of skilled and unskilled adults, i.e., on the aggregate of educational decisions made by parents of the previous generation. In the base model, parents are assumed to supply labor inelastically, not bear any income risk, have perfect foresight over future wages, exhibit Barro-Becker paternalistic altruism towards their children, and unable to transfer wealth via financial bequests. Starting with an initial distribution of skills, a competitive equilibrium is characterized by a threshold ability above which parents in any given occupation and generation decide to invest in their children's education. Those that invest end up consuming less than those in the same generation-occupation pair who do not invest.

Starting with a laissez faire equilibrium, the government can offer a small subsidy to any parent that educates its child, which is financed by a tax paid by all parents in the same generation and occupation. This effectively redistributes consumption within a given occupation-generation pair from parents that do not invest in education, to parents that invest. This smooths parental consumption with respect to ability shocks of their children, besides lowering the ability threshold for education. *Ceteris paribus* this raises interim utility of parents in the given occupation-generation pair, as well as the proportion of the population in the next generation that are educated. In other words, the policy results simultaneously in higher education incentives and insurance of parental consumption against ability risk. This forms the primary source of the efficiency improvement.

What about effects on educational investments and utilities in other generations? Anticipation of this intervention will affect educational incentives in previous generations. We show the latter can be neutralized at zero first order cost by a set of corresponding interventions for every occupation-generation pair. For instance, if the primary intervention involved educational transfers for unskilled parents in generation t, a similar intervention for skilled parents in generation t would generate an increase in interim utility of skilled parents of the same magnitude for unskilled parents at generation t, leaving the difference in utilities between skilled and unskilled parents unchanged. Hence educational incentives in preceding generations would be unaffected. At the same time interim utilities and educational incentives of skilled parents would also increase.

Moreover, the intervention will generate ripple effects for succeeding generations, by increasing the supply of skilled agents at $t+1$. Owing to general equilibrium effects, this would lower skilled wages at $t+1$, which leave skilled agents at $t+1$ worse off. We show these can be neutralized by occupation-specific transfers that leave post-tax wages unchanged from $t+1$ onwards, at zero first order budgetary cost. This would involve regressive income taxes which would offset the reduction in wage differences between the skilled and unskilled occupations. The eventual fiscal intervention overlays all these different interventions together, in a way that raises interim utilities of parents in every occupation-generation pair. *Ex post* parents who do not invest in their children's education end up cross-subsidizing the consumption of those who do. With sufficient altruism, the costs of this are outweighed by the gain in expected utilities of their descendants, resulting in an *ex post* Pareto improvement.

A similar argument works for any tax-distorted competitive equilibrium which does not involve educational subsidies, provided the initial taxes are progressive (which is required to ensure that an increase in the supply of skilled agents does not increase fiscal deficits). Note that the argument does not require the government to borrow or lend across generations; nor does it require it to observe the ability realizations of children in the population. We subsequently show the argument is robust to incorporation of endogenous labor supply, paternalistic altruism, and an arbitrary number of occupations.

However, the results do not generalize quite as straightforwardly when human capital investments can be supplemented by financial bequests. In particular, they do not apply to households wealthy (and altruistic) enough that they *always* make financial bequests, irrespective of how much they invest in education. Within such a wealth class, those who do not invest in education end up spending more on their children overall, and thus consume less than parents who do invest in education. This reverses the pattern of consumption variation with respect to the realization of children's ability risk – the educational subsidy policy described above would now impose additional consumption risk, and thereby create a welfare loss. Laissez faire competitive equilibria continue to be constrained Pareto inefficient, however. The nature of a Pareto improving policy is now reversed: requiring education for the wealthy (as defined above) to be taxed, and these taxes to fund income transfers to the same class. The nature of the Pareto improving policy is unchanged for poor households who invest if at all only in education and leave no financial bequests. The aggregate macroeconomic effects of such a Pareto improving policy are unclear, as the increased educational investments among the poor will be countered by falling investments among the wealthy.

This paper relates to a wide literature spanning macroeconomics, public finance, occupational choice and development economics on the policy implications of financial frictions. We present a detailed review of related literature in Section 5. We argue there that the distinguishing characteristic of this paper is a qualitative point which is simple but has not yet been made in this literature: ability heterogeneity implies that educational subsidies are interim Pareto improving under a wide range of circumstances. A similar argument would imply an efficiency enhancing role of investment subsidies for small firms that are subject to productivity shocks and face borrowing constraints.

The remainder of the paper is organized as follows. Section 2 introduces the baseline model, followed by the main results for this model in Section 3. Extensions are discussed in Section 4. The relation to existing literature is described in Section 5, and Section 6 concludes.

2 Baseline Model

We first describe the dynamic economy in the absence of any government intervention.

There are two occupations: unskilled and skilled (denoted 0 and 1 respectively). There is a continuum of households indexed by $i \in [0,1]$. Generations are denoted $t = 0, 1, 2, \ldots$ Each household has one adult and one child in each generation. The utility of the adult in household i in generation t is denoted $V_{it} = u(c_{it}) + \delta V_{i,t+1}$ where c_{it} denotes consumption in household i in generation t, $\delta \in (0,1)$ is a discount factor and measure of the intensity of parental altruism, and u is a strictly increasing, strictly concave and C^2 function defined on the real line. There is no lower bound to consumption, while u tends to $-\infty$ as c tends to $-\infty$.

Household i earns y_{it} in generation t, and divides this between consumption at t and investment in child education. Education investment I_{it} is indivisible, either 1 or 0. An educated adult has the option of working in either occupation, while an uneducated adult can only work in the unskilled occupation. The ability of the child in household i is represented by how little its parent needs to spend in order to educate it. The cost of education x_{it} in household i in generation t is drawn randomly and independently according to a common distribution function F defined on the nonnegative reals. F is C^2 and strictly increasing; its density is denoted f. The household budget constraint is $y_{it} = c_{it} + x_{it} I_{it}$. Every parent privately observes the realization of education cost of its child before deciding on whether to invest in education.[2]

The key market incompleteness is that parents cannot borrow to finance their children's

[2] Findings would not change if we assumed that parents receive a noisy signal $\hat{x}_{it}$ of true education costs provided that the signal's precision is non-decreasing in parental status and u exhibits non-increasing absolute risk aversion.

education. Neither can they insure against the risk that their child has low learning ability, the main source of (exogenous) heterogeneity in the model. The former arises owing to inability of parents to borrow against their children's future earnings. The latter could be due to privacy of information amongst parents regarding the realization of their children's ability.

Household earnings are defined by occupational wages: $y_{it} = w_{0t} + I_{i,t-1} \cdot (w_{1t} - w_{0t})$, where w_{ct} denotes the wage in occupation c in generation t obtaining in a competitive labor market.

Wages are determined as follows at any given date (so we suppress the t subscript for the time being). There is a CRS production function $G(\lambda, 1-\lambda)$ which determines the per capita output in the economy in any generation t if the proportion of the economy that works in the skilled and unskilled occupations equal λ and $1-\lambda$ respectively. We assume G is a C^2, strictly increasing, linearly homogenous and strictly concave function. Let $g_c(\lambda)$ denote the marginal product of occupation $c = 0, 1$ workers when λ proportion of adults work in the skilled occupation. So g_1 is decreasing and g_0 is an increasing function. Moreover, $g_1(0) > g_0(0)$ while $g_1(1) < g_0(1)$. To avoid some technical complications we assume the functions g_i are bounded over $[0, 1]$. In other words, the marginal product of each occupation is bounded above even as its proportion in the economy becomes vanishingly small.[3]

Let $\bar{\lambda}$ denote the smallest value of λ at which $g_1(\lambda) = g_0(\lambda)$. Then in any given generation t, all educated workers will prefer to work in the skilled occupation, with $w_{1t} = g_1(\lambda_t)$ and $w_{0t} = g_0(\lambda_t)$, if the proportion of educated adults is $\lambda_t < \bar{\lambda}$. And if $\lambda_t \geq \bar{\lambda}$, equilib-

[3]When the production function satisfies Inada conditions, i.e., marginal products are unbounded, we obtain the same results if every household is able to resort to a subsistence self-employment earnings level $\underline{w}$ which is positive and exogenous. As the proportion of unskilled workers tends to one, the labor market will clear at an unskilled wage equal to $\underline{w}$, and the proportion of skilled households working for others will be fixed at a level where the marginal product of the unskilled equals this wage. The only difference is that wages in either occupation as a function of the skill ratio become kinked at the point where the marginal product of the unskilled equals $\underline{w}$. Except at this single skill ratio, the wage functions are smooth, and our results continue to apply with an 'almost everywhere' proviso.

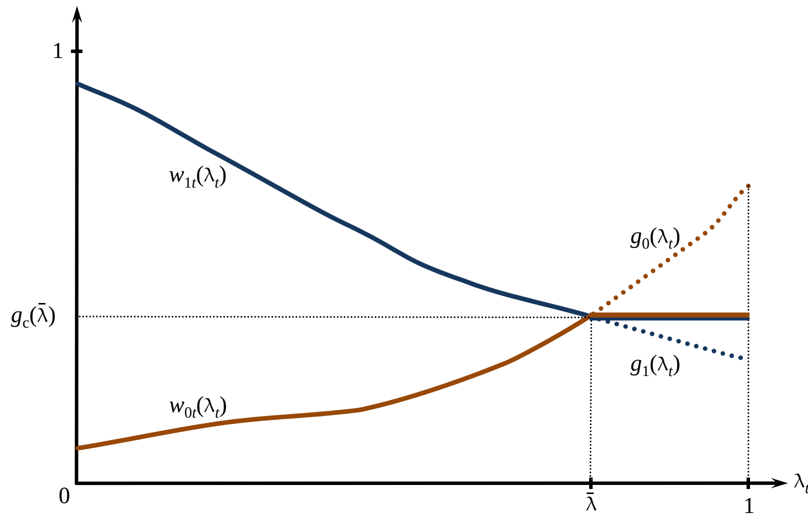

Figure 1: Labor Market

rium in the labor market at t will imply that exactly $\bar{\lambda}$ fraction of adults will work in the skilled occupation, as educated workers will be indifferent between the two occupations, and $w_{1t} = w_{0t} = g_1(\bar{\lambda}) = g_0(\bar{\lambda})$. See Figure 1. When more than $\bar{\lambda}$ fraction of adults in the economy are educated, the returns to education are zero. Since education is costly, education incentives vanish if households anticipate that more than $\bar{\lambda}$ proportion of adults in the next generation will be educated. Hence the proportion of educated adults will always be less than $\bar{\lambda}$ in any equilibrium with perfect foresight. We can identify the occupation of each household i in generation t with its education status $I_{i,t-1}$, and refer to λ_t as the skill ratio in the economy in generation t.

2.1 Dynamic Competitive Equilibrium under Laissez Faire

Definition 1 *Given a skill ratio* $\lambda_0 \in (0, \bar{\lambda})$ *in generation* 0*, a* dynamic competitive equilibrium under laissez faire (DCELF) *is a sequence* $\{\lambda_t\}_{t=0,1,2,...}$ *of skill ratios and investment strategies* $\{I_{ct}(x)\}_{t=0,1,2,...}$ *for every household in occupation* c *in generation* t *when its child's education cost happens to be* x *such that:*

(a) For each household and each t: $I_{ct}(x) \in \{0,1\}$ *maximizes*

$$u\big(w_{ct} - I_{ct}x\big) + \delta \mathsf{E}_{\tilde{x}} V_{t+1}(I_{ct}, \tilde{x}) \tag{1}$$

and the resulting value is $V_t(c, x)$*.*

(b)

$$\lambda_t = \lambda_{t-1}\mathsf{E}_x[I_{1t}(x)] + (1 - \lambda_{t-1})\mathsf{E}_x[I_{0t}(x)]. \tag{2}$$

(c) Every household correctly anticipates $w_{ct} = g_c(\lambda_t)$ *for occupation* $c = 0, 1$ *in generation* t*.*

It is useful to note the following features of a DCELF.

Lemma 1 *In any DCELF and at any date t:*

(i) $V_t(1, x) > V_t(0, x)$ *for all* x *if and only if* $\lambda_t < \bar{\lambda}$*.*

(ii) $\lambda_t < \bar{\lambda}$, $w_{1t} > w_{0t}$*.*

(iii) $I_{ct}(x) = 1$ *iff* $x < x_{ct}$*, where threshold* x_{ct} *is defined by*

$$u(g_c(\lambda_t)) - u(g_c(\lambda_t) - x_{ct}) = \delta[W_{1,t+1} - W_{0,t+1}] \tag{3}$$

and $W_{ct} \equiv \mathsf{E}_x V_t(c, x)$

(iv) The investment thresholds satisfy $x_{0t} < x_{1t}$*, are uniformly bounded away from 0, and uniformly bounded above, while* λ_t *is uniformly bounded away from* 0 *and* $\bar{\lambda}$ *respectively. Consumptions of all agents are uniformly bounded.*

This Lemma shows that skilled wages always exceed unskilled wages, and those agents in skilled occupations always have higher utility. There is inequality of educational opportunity: children born to skilled parents are more likely to be educated. There is also upward and downward mobility: some talented children born to unskilled parents receive an education, while some untalented children born to skilled parents fail to receive an education. Finally, equilibrium consumptions and utility differences are bounded, which will be useful in our subsequent analysis.

2.2 Competitive Equilibrium with Taxes

We now extend the model to incorporate fiscal policies. The government observes the occupation/income of parents as well as the education decisions they make for their children. Transfers can accordingly be conditioned on these. Fiscal policy is represented by four variables in any generation t: income transfers τ_{1t}, τ_{0t} based on parental occupation, and transfers e_{1t}, e_{0t} based additionally on the parent's education investment decision. In particular, the government does not observe directly nor indirectly the ability realization of any given child.[4] This is the key informational constraint that prevents attainment of a first-best utilitarian optimum. We are also focusing on transfers that depend only on the current status of the household, thus ruling out educational loans and schemes which condition on a family's transfer or decision history. Similar to private agents, the government will also not be able to lend or borrow across generations, and will hence have to balance its budget within each generation.

Definition 2 *Given a skill ratio $\lambda_0 \in (0, \bar{\lambda})$ in generation* 0*, a* dynamic competitive equilibrium (DCE) given fiscal policy $\{\tau_{0t}, \tau_{1t}, e_{0t}, e_{1t}\}_{t=0,1,2,...}$ *is a sequence $\{\lambda_t\}_{t=0,1,2,...}$ of skill ratios and investment strategies $\{I_{ct}(x)\}_{t=0,1,2,...}$ for every household in occupation c in generation t when its child's education cost happens to be x such that for each c, t:*

(a) $I_{ct}(x) \in \{0, 1\}$ maximizes

[4] Indirect observability of children's abilities from the parental education expenses or test results would allow policy to realize efficiency gains from explicit improvements in the talent composition of investors. We think of education costs x as having a major unverifiable component, possibly also reflecting parental time that would be dedicated to a child's education and training in a more sophisticated model of labor supply.

$$u\big(w_{ct} + \tau_{ct} - I_{ct} \cdot (x - e_{ct})\big) + \delta \mathsf{E}_{\tilde{x}} V_{t+1}(I_{ct}, \tilde{x}) \tag{4}$$

and the resulting value is $V_t(c, x)$*.*

(b)

$$\lambda_t = \lambda_{t-1} \mathsf{E}_x I_{1t}(x) + (1 - \lambda_{t-1}) \mathsf{E}_x I_{0t}(x). \tag{5}$$

(c) Every household correctly anticipates $w_{ct} = g_c(\lambda_t)$ *for occupation* $c = 0, 1$ *in generation* t*.*

The government has a balanced budget *if at every* t *it is the case that*

$$\lambda_t \big\{ \tau_{1t} + e_{1t} \mathsf{E}_x[I_t(1, x)] \big\} + (1 - \lambda_t) \big\{ \tau_{0t} + e_{0t} \mathsf{E}_x[I_t(0, x)] \big\} \leq 0. \tag{6}$$

A DCELF with a (trivially) balanced budget obtains as a special case of a DCE when the government selects zero income transfers and educational subsidies.

It is easy to check that a DCE can also be described by investment thresholds x_{ct} satisfying the following conditions. Define the interim expected utility of consumption of a parent in occupation c in generation t as follows:

$$\mathcal{U}_{ct} \equiv u(w_{ct} + \tau_{ct})[1 - F(x_{ct})] + \int_0^{x_{ct}} u(w_{ct} + \tau_{ct} + e_{ct} - x) dF(x) \tag{7}$$

The thresholds must then satisfy

$$u(w_{ct} + \tau_{ct}) - u(w_{ct} + \tau_{ct} + e_{ct} - x_{ct}) = \delta \cdot \Delta W_{t+1} \tag{8}$$

where W_{it} denotes $\mathsf{E}_x V_t(i, x)$,

$$\Delta W_t \equiv W_{1,t} - W_{0,t} = \sum_{k=0}^{\infty} \nu_k [\mathcal{U}_{1,t+k} - \mathcal{U}_{0,t+k}] \tag{9}$$

with $\nu_0 = 1$, $\nu_k = \delta^k \Pi_{l=0}^{k-1} [F(x_{1,t+l}) - F(x_{0,t+l})]$ for $k \geq 1$. A DCE is then described by a sequence $\{\lambda_t, w_{1t}, w_{0t}, x_{1t}, x_{0t}, \mathcal{U}_{1t}, \mathcal{U}_{0t}\}_{t=0,1,2,\ldots}$ which satisfies equalities (5) and (7)–(9).

3 Results for the Baseline Model

Our first result is an efficiency as well as a macroeconomic role for fiscal policy. The efficiency criterion is *interim Pareto dominance*, which requires parental expected utility $W_{ct} \equiv \mathsf{E}_x V_t(c,x)$ to be higher for every c,t. The criterion of *macroeconomic dominance* is that the skill ratio λ_t must be higher at every t, and the investment threshold x_{ct} must be higher for every c,t. This ensures higher per capita skill and output at every date, as well as greater educational opportunity in the sense of a higher probability for every child to become educated (both conditional on parent's occupation, and unconditionally).

Theorem 1 *Consider any DCELF starting from an arbitrary skill ratio $\lambda_0 \in (0,\bar{\lambda})$ at $t=0$. There exists a balanced budget fiscal policy with educational subsidies for each occupation funded by income taxes, and an associated DCE which interim Pareto as well as macroeconomically dominates the original DCELF.*

A more general version of this result is the following: any fiscal policy involving income transfers alone is Pareto dominated by a policy with educational subsidies.

Theorem 2 *Consider any DCE given an initial skill ratio $\lambda_0 \in (0,\bar{\lambda})$ and a balanced budget fiscal policy consisting of income transfers alone ($e_{ct}=0$ for all c,t), satisfying the following conditions:*

(a) $\tau_{0t} \geq \tau_{1t}$ for all t;

(b) there exists $\kappa > 0$ such that $-[\tau_{1t} - \tau_{0t}] < [g_1(0) - g_0(0)] - \kappa$ for all t;

(c) τ_{ct} is uniformly bounded.

Then there exists another balanced budget fiscal policy consisting of income transfers combined with educational subsidies ($e_{ct} > 0$ for all c,t) and an associated DCE which interim Pareto as well as macroeconomically dominates the original DCE.

Condition (a) of Theorem 2 requires the income transfers to be progressive in the weak sense that unskilled parents receive a higher transfer (or pay a lower tax), while (b) restricts the marginal tax rate to be less than (and bounded away from) 100%. Condition (c) is a technical restriction needed to ensure that competitive equilibria always involve bounded consumptions and investment thresholds. The role of (b) is to ensure that skilled households earn more both before and after government transfers, so agents always have investment incentives (that are bounded away from zero). Condition (a) ensures that the direct effect of any reduction in the proportion of unskilled households is to weaken the government budget balance constraint. These conditions imply that equilibria with fiscal policy continue to satisfy the same properties as equilibria under laissez faire that were shown in Lemma 1.

It is evident that Theorem 1 is a special case of Theorem 2. In both versions, the new policy provides educational subsidies to a given occupation which are funded by higher income taxes *levied on the same occupation*. In this sense, the policy does not redistribute across occupations. Instead, it redistributes across parents within each occupation class, between those that do and do not invest in their children's education. The key underlying idea is to provide insurance against the uncertain realizations of children's ability. The 'accident' in question is that one's child is born with enough talent that the parent invests in education, resulting in lower consumption than other parents in the same occupation who do not invest (i.e., whose children are not so talented). The nature of variation of parental consumption with the realization of their child's ability is illustrated in Figure 2. If the child is a genius and can be costlessly educated, the parent's consumption equals his earning. The same is true when the child has low enough ability that it is not educated. For intermediate abilities where the child is educated, the parent invests a positive amount, lowering consumption. Hence parental consumption varies non-monotonically with respect to the cost necessary to educate the child.

An educational subsidy increase $\epsilon(1-\mu_t)$ raises the consumption of the investors, while financing it by income taxes on the same occupation lowers the consumption of the non-

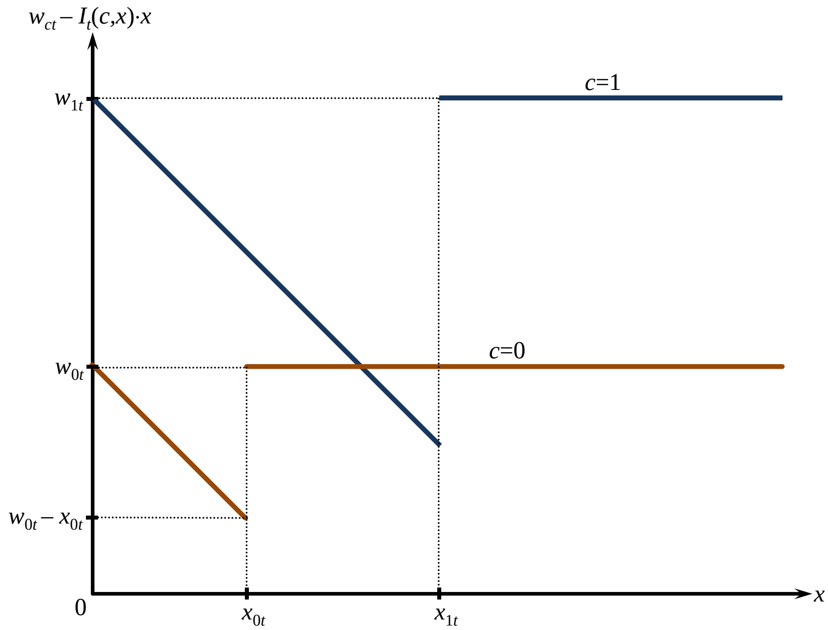

Figure 2: Variation of Parental Consumption with Education Cost

investors.[5] See Figure 3. If μ_t were zero, average consumption would be unaffected but differences in consumption associated with heterogeneity of the children's education costs would be reduced (assuming that education is not subsidized to start with and hence non-investing households consume more than every household in the same occupation that does invest). This would result in a mean-preserving reduction of the variation in parental consumption, thus raising the interim expected utility of current consumption in occupation c.

The parameter μ_t, however, is set so as to reduce the mean consumption enough that there is no change in the expected utility of current consumption at date t for each occupation. Assuming wages are unchanged, this implies that dynastic utilities of both occupations are unchanged. Hence the future benefit of investment is unchanged. The subsidization of

[5] Lowered consumption utility of non-investors may – despite a universal increase in the dynastic utility component – preclude the policy from achieving also an *ex post* Pareto improvement. By restricting tax-funded education subsidies to the minority occupation defined by $\lambda_t \gtrless \frac{1}{2}$, a government could prevent any political commitment problems which might derive from having too many ex post losers.

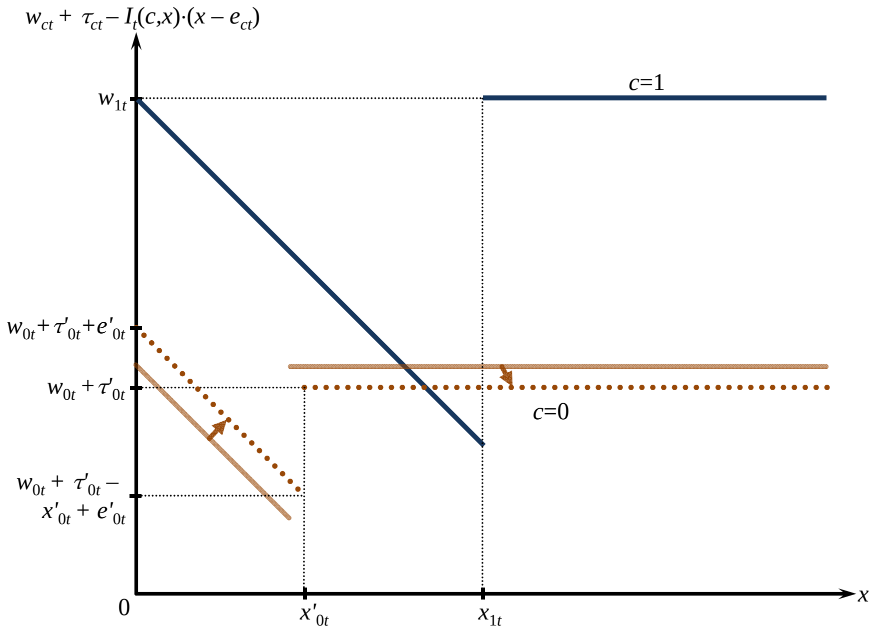

Figure 3: Effects of Steps 1 and 2 of Fiscal Policy Variation on Parental Consumption

education in occupation c on the other hand lowers the sacrifice parents must endure to educate their children. Hence households invest more often.

Aggregate investment in the economy will then rise, which will tend to lower skilled wages and raise unskilled wages. These general equilibrium changes would reduce the benefits of investment and therefore fiscal policy is adjusted further to neutralize the wage changes. This results in a new competitive equilibrium sequence with a higher skill ratio at every date, and a zero first order effect on interim utilities. However, the government has a first order improvement in its surplus, owing to the rise in the skill ratio and the extraction of resources from households by setting $\mu_t > 0$. The progressivity of the original fiscal policy implies that the government budget surplus also improves as a result of the decline in the proportion of unskilled households.

In the last step of the argument the government constructs another variation in its tax-subsidy policy. It distributes the additional revenues so as to achieve a strict interim Pareto improvement, while preserving investment incentives. Note that by construction the dispersion in utility of consumption between occupations is unchanged, while a fraction of agents move up from the unskilled to the skilled occupation in every generation.

The consumption losses which the policy imposes on non-investing parents stay bounded while the dynastic gains which are created for all parents grow without bound as $\delta \to 1$. So with a sufficiently high degree of parental altruism, parameterized by δ, the policy-induced gains in expected utility of descendants outweigh any loss in own consumption relative to laissez faire. The constructed policy then achieves an ex post Pareto improvement. Formally, we can show:

Theorem 3 *Let a collection of economies with identical consumption utility function u, production function G and ability distribution F but different parental discount factors $\delta \in (0,1)$ be given. For each corresponding DCELF that starts from skill ratio $\lambda_0 \in (0, \bar{\lambda})$ at $t = 0$, consider fiscal policies $\left\{\tilde{\tau}^{\delta}_{0t}(\epsilon), \tilde{\tau}^{\delta}_{1t}(\epsilon), \tilde{e}^{\delta}_{0t}(\epsilon), \tilde{e}^{\delta}_{1t}(\epsilon)\right\}_{t=0,1,2,\ldots}$ which induce an interim Pareto improvement according to Theorem 1. Then there exist $\underline{\delta} \in (0,1)$ and $\bar{\epsilon} > 0$ such that for any $\epsilon \in (0, \bar{\epsilon})$ and $\delta \in (\underline{\delta}, 1)$ the fiscal policy also ex post Pareto dominates the respective DCELF for all t.*

4 Extensions

4.1 Endogenous Labor Supply

A first extension of the baseline model could be to consider households who choose how many hours of labor they supply, together with the binary decision whether to invest in education or not. That is, each household in occupation c selects $I_{ct}(x) \in \{0,1\}$ and $l_{ct}(x) \geq 0$ to maximize

$$u\big(l_{ct}\, w_{ct} - I_{ct}\, x\big) - d(l_{ct}) + \delta \mathsf{E}_{\tilde{x}} V_{t+1}(I_{ct}, \tilde{x}) \tag{10}$$

for strictly increasing and convex disutility of labor d. The optimal investment strategy $I_{ct}(x)$ in this case is of the same threshold form as in the baseline model. Namely, if we define

$$v(w_{ct},x,I_{ct}) \equiv \max_{l_{ct}} \Big[u\big(l_{ct}\, w_{ct} - I_{ct}\, x\big) - d(l_{ct})\Big] \tag{11}$$

then a parent in occupation c in period t who faces education cost x will invest iff $x < x_{ct}$, where threshold x_{ct} is defined by

$$v(w_{ct},x_{ct},0) - v(w_{ct},x_{ct},1) = \delta[W_{1,t+1} - W_{0,t+1}] \tag{12}$$

and $W_{ct} \equiv \mathsf{E}_{\tilde{x}} V_t(c,\tilde{x})$. Parents in occupation c with cost $x = 0$ or cost $x \geq x_{ct}$ have identical (indirect) utilities of consumption $v(w_{ct},0,1) = v(w_{ct},x,0)$, while those with cost $x \in (0,x_{ct})$ consume less. In particular, from (11) and the envelope theorem, we have

$$\frac{\partial v(w_{ct},x,I_{ct}(x))}{\partial x} = -u'\big(l_{ct}(x)w_{ct} - x\big) < 0 \text{ for each } x \in (0,x_{ct}). \tag{13}$$

It follows that consumption utilities $v(w_{ct},x,I_{ct}(x))$ are decreasing on $[0,x_{ct})$, jump back to $v(w_{ct},0,1)$, and then stay at this level. That is, they exhibit a non-monotonic pattern with respect to education cost x just like in the baseline model (cf. Figure 2). A variation of the baseline policy intervention can therefore be applied in order to raise interim utility.

4.2 Paternalistic Altruism

Next suppose parents do not have Barro-Becker dynastic preferences. Instead, they value (only) the earnings of their children according to a given increasing function $Y(w_{t+1})$, as in Becker and Tomes (1979) or Mookherjee and Napel (2007) – perhaps incorporating parental concern for their own old age security. A parent in occupation $c \in \{0,1\}$ at date t with a child who costs x to educate then selects $I \in \{0,1\}$ to maximize $u\big(w_{ct} - I\,x) + I\,Y(w_{1,t+1}) + (1-I)\,Y(w_{0,t+1})\big)$. Theorems 1 and 2 continue to extend with this formulation of parental altruism. The wage neutralization policy preserves after-tax wages in each occupation, whence the altruistic benefit of investments remain unchanged. The costs of investing are lowered by providing educational subsidies, and at the same time the variation of parental consumption is lowered. So investment incentives continue to rise, while enhancing interim expected utilities.

4.3 Continuous Investment Choices

What if educational investments can be varied continuously, rather than being indivisible? Our results extend straightforwardly to this context, too, as we now explain.

Let the extent of education be described by a compact interval $E \equiv [0, \bar{e}]$ of the real line. Assume that the relation between wage earnings and education is given by a real-valued continuous function $w(e)$ defined on E. If the earnings function depends endogenously on the supply of workers with varying levels of education, the analysis can be extended using a similar strategy of following up on educational subsidy policies that increase the supply of more educated workers with a wage-neutralization policy that leaves the after-tax remuneration pattern unchanged. To illustrate how our results extend, it therefore suffices to take the earnings–education pattern in the status quo equilibrium as given.

Let $I(e'; x)$ denote the expenditure that must be incurred by a parent to procure education $e' \geq 0$ for its child whose learning ability gives rise to a learning cost parameter x. The latter varies according to a continuous distribution with full support on $[0, \infty)$, similar to the preceding section. The function I is strictly increasing and differentiable in both arguments. It satisfies $I(0; x) = 0$ for all x, while for any given $e' \geq 0$ the marginal cost $\frac{\partial I}{\partial e'}$ is increasing in x, approaching ∞ as $x \to \infty$.

The value function of a parent with education e and a child whose learning cost parameter is x is then

$$V(e|x) \equiv \max_{0 \leq e' \leq \bar{e}} \left[u(w(e) - I(e'; x)) + \delta W(e') \right] \tag{14}$$

where $W(e') \equiv \mathsf{E}_{\tilde{x}} V(e'|\tilde{x})$. Let the corresponding policy function be $e'(e; x)$. Given that wages are bounded above by $w(\bar{e})$, consumptions are also bounded above. Given this and the feature that u is unbounded below, consumptions can be bounded from below almost surely.[6] Hence the marginal utility of consumption is bounded almost surely, implying that $W'(0) \equiv \mathsf{E}_{\tilde{x}}[u'(w(0) - I(e'(0; \tilde{x}))]$ is bounded.

[6]Any policy where consumption approaches $-\infty$ with positive probability will be dominated by a policy where parents never invest.

We can therefore define $x^*(e)$ as the solution for x in the equation $\frac{\partial I(e';x)}{\partial e'}\big|_{e'=0} = \frac{\delta W'(0)}{u'(w(e))}$. Then the optimal policy function takes the form $e'(e;x) = 0$ if $x \geq x^*(e)$ and positive otherwise.[7] In other words, parents decide to acquire no education for their children if and only if their learning cost parameter is larger than a threshold $x^*(e)$. These 'non-investors' consume their entire earnings $w(e)$ – just like those parents with the same education e whose children have a learning cost parameter of $x = 0$. For those whose children have intermediate learning ability, parents spend a positive amount on education.

We thus have a similar non-monotone pattern of variation of parental consumption with their children's learning costs as in the two-occupation case. This ensures that a similar policy of educational subsidies funded by income taxes on all parents with the same education will reduce the riskiness of parental consumption, and thereby permit a Pareto improvement.

The essential argument is thus simple. Non-investing parents within any given occupation will by definition consume more than investing parents. The educational subsidy funded by the income tax on this occupation then redistributes consumption away from those consuming high amounts to those consuming less. Since these consumption variations arise from the 'ability lottery' of their children, the policy increases interim expected utilities of each occupation. The preceding analytical details were needed to ensure that there is a positive mass of investors and non-investors respectively, so as to allow a strict Pareto improvement.

4.4 Financial Bequests

There is however one important assumption underlying the above reasoning: that educational investments constitute the sole means by which parents transfer wealth to their children. In practice parents have other means as well, such as leaving them financial bequests or physical assets. The simple logic then breaks down: a parent that does not invest in his

[7]This follows since the value function is concave, owing to a direct argument.

child's education owing to low learning ability of the latter could provide financial bequests instead. It no longer follows that education non-investors invest less when we aggregate across different forms of intergenerational transfers.

We now consider the consequences of allowing parents to leave financial bequests besides investing in their children's education. To simplify matters, suppose that the rate of return $(1+r)$ on financial bequests is exogenously given, as in Becker and Tomes (1979) or Mookherjee and Ray (2010). This could correspond to a globalized capital market where the savings of any given country leave the interest rate unaffected. Even if the interest rate depends on the supply of savings, a 'neutralization' policy allows policy-makers to ensure that the after-tax interest rate is unchanged. For the same reason we here abstract from general equilibrium effects in the labor market and suppose that wages of different occupations are exogenously given.

Let us further simplify to the case of two occupations, skilled and unskilled, where the education cost of the former is denoted x and the latter equals zero. And suppose that parental altruism is paternalistic, where a parent with lifetime wealth W and education cost x chooses financial bequest $b \geq 0$ and education investment $I \in \{0, 1\}$ to maximize $u(W - b - Ix) + \delta Y(W')$ where Y is a strictly increasing and strictly concave function of the child's future wealth $W' = (1+r)b + Iw_1 + (1-I)w_0$.

This problem can be reformulated as follows. Let $C \equiv b + Ix$ denote the total parental investment expenditure on his child. An efficient way to allocate C across financial bequest and educational expenses is the following: $I = 0$ if either $C < x$, or $C \geq x$ and the rate of return on education is dominated by the return on financial assets: $\frac{w_1 - w_0}{x} < 1 + r$. Conversely, if the rate of return on education exceeds r and $C \geq x$, then $I = 1$, and $b = C - x$. Then the child ends up with wealth $W' \equiv R(C; x)$ given by

$$R(C;x) = \begin{cases} (1+r)C + w_0 & \text{if } C < x, \text{ or } C \geq x \text{ and } \frac{w_1 - w_0}{x} \leq 1 + r, \\ (1+r)C + w_1 - (1+r)x & \text{if } C > x \text{ and } \frac{w_1 - w_0}{x} > 1 + r. \end{cases} \tag{15}$$

It is illustrated in Figure 4.

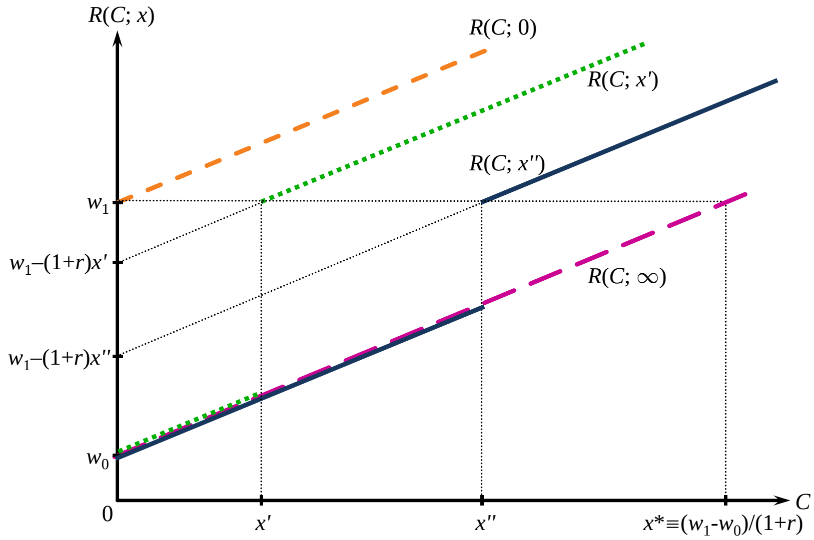

Figure 4: Child wealth as function of total investment expenditure C, given cost x

Define the *BT (Becker-Tomes) bequest* as the optimal bequest of a parent in the absence of any opportunity to invest in education, with a given flow earning w of the child when the parent leaves a zero bequest. This is the problem of choosing $C \geq 0$ to maximize $u(W - C) + \delta Y((1+r)C + w)$. Denote the BT bequest by $C^{BT}(W; w)$. It is easily checked that this is increasing in parental wealth W and decreasing in w.

Recall that a parent will invest in education only if the child has enough ability to ensure that $x \leq x^* \equiv \frac{w_1 - w_0}{1+r}$. Whenever $x > x^*$, there will be no investment in education, and the optimal bequest equals the BT bequest $C^{BT}(W; w_0)$. When $x < x^*$, the optimization problem entails a nonconvexity and the solution is more complicated. The dotted and solid lines in Figure 4, for instance, respectively represent the nonconvex sets of feasible (C, W')-combinations for parents with children whose education costs x' and x'' lie below x^*.

Nevertheless we can illustrate the solution for some extreme cases, corresponding to different parental wealths.

Case A. *W sufficiently large:* Suppose W is large enough that $C^{BT}(W; w_1 - (1+r)x) > x$

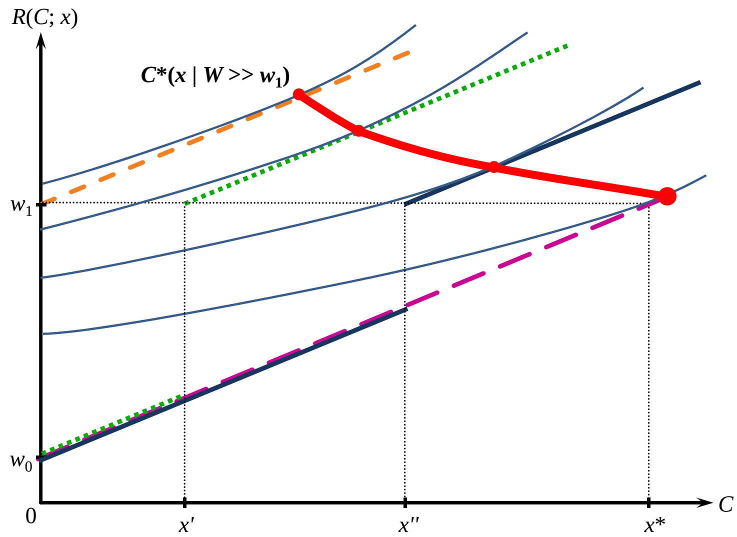

Figure 5: Investment expenditures of sufficiently wealthy parents (case A)

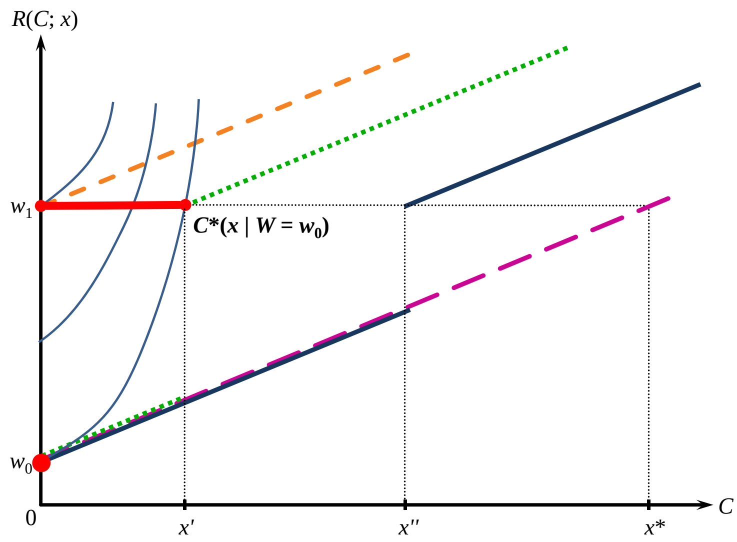

Figure 6: Investment expenditures of poor parents (case B)

for all $x \le x^*$.[8] In words, irrespective of where x lies below x^*, the parent will always supplement education investments with a financial bequest. See Figure 5.

Case B. *W sufficiently small:* Suppose $W = w_0$, $\delta(1 + r) \le 1$ and $Y \equiv u$. Then the BT bequest $C^{BT}(w_0; w) = 0$ for all $w \ge w_0$, and the parent will never make a financial bequest. If however the child learning cost x is sufficiently small, the parent will invest in education. The optimal choice of expenditure C^* is illustrated in Figure 6, where the low parental wealth is reflected by steep indifference curves.

The implied consumption patterns of sufficiently wealthy and poor households are illustrated in Figure 7. For parents with very small wealth W, investment decisions are exactly as in our simple model without any financial bequests, and 'non-investors' consume more than the 'investors'. The situation is very different, however, for sufficiently wealthy parents. Their parental consumption (conditional on wealth W) is strictly decreasing in x over $x \in [0, x^*]$, and constant thereafter. The 'non-investors' (those with $x > x^*$) now consume *less* than the 'investors', opposite to the pattern in the model without any financial bequests.

The argument that educational subsidies (financed by income or wealth taxes) lower consumption risk no longer applies to wealthy households falling under case A. They would instead raise risk. *So an opposite result holds here: an educational tax for parents with wealths falling in case A which funded a wealth subsidy (or income tax break) on the same set of households would reduce risk.* Starting with laissez faire, such a policy would be Pareto improving. It would, however, have opposite macroeconomic effects, as educational investments among such parents would fall. The resulting decline in skilled agents implies that the result about superiority of conditional transfers may not apply if the status quo policy is progressive, as this would worsen the government's fiscal balance.

On the other hand, our previous arguments would continue to apply for poor households in case B, who never make any financial bequests, and behave exactly as described in previous sections. For such poor households, therefore, our previous results remain unchanged:

[8] A sufficient condition for this is $C^{BT}(W; w_1) > x^*$.

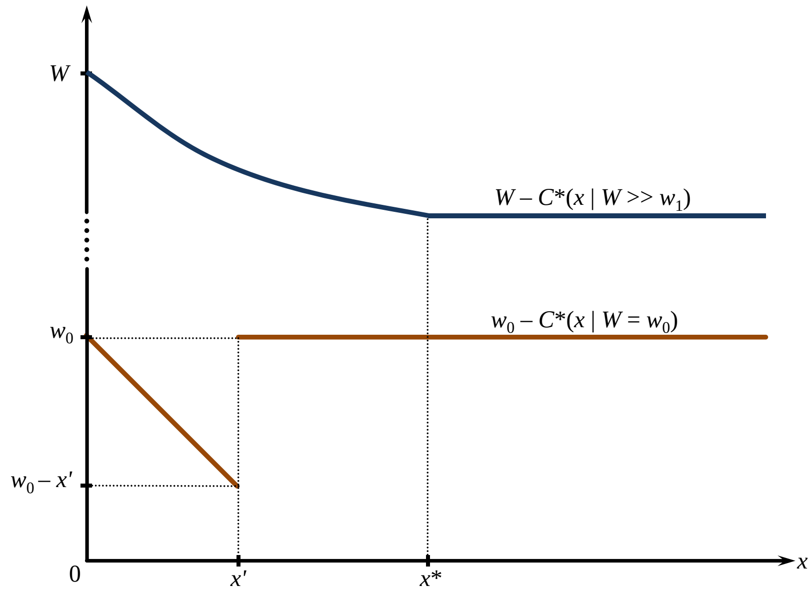

Figure 7: Consumption of sufficiently wealthy and poor parents

educational subsidies funded by income taxes would be Pareto improving as well as generate macro improvements.

For other classes of households, whether parents make financial bequests typically depends on the child's ability: they are made when the child is of sufficiently high ability, as well as when ability is low. For intermediate abilities, they make no financial bequests and make educational investments alone. The comparison of consumptions across 'investors' and 'non-investors' can go either way depending on the child's ability.

This suggests that arguments for educational subsidies should be limited to household wealth classes which make little or no financial bequests. The exact range of such households is an empirical matter. In the model of Abbott, Gallipoli, Meghir and Violante (2013) calibrated to fit the NLSY 1997 data, all parents in the bottom quartile of the wealth distribution make inter-vivos transfers (inclusive of imputed value of rent when children lived with parents) to their children (when the latter were between ages of 16–22) which were smaller than what the latter spent on educational tuitions. The same was true for most of the second quartile as well. On the other hand, many parents in the top quartile transferred more than education tuition costs, and this happened to be true for *all* parents in the top 5%. This suggests case A applies to the top 5% of the US population, while case B applies to the bottom third of the population.

Indeed, our results suggest that it may be optimal for the government to use mixed policies of the following form: educational taxes for the population in case A, and subsidies for those in case B. The effects on educational investments in these two classes could then offset each other, leaving aggregate education investments unaltered. The composition of the educated would however change: since marginal children in case B are likely to be of higher ability than those in case A, there would be a rise in the average returns to education which would augment the efficiency benefits from the risk effects. We conjecture that it is generally possible to construct such mixed policies which Pareto dominate a policy consisting of income-based transfers alone.

5 Relation to Literature

Sinn (1995, 1996) and Varian (1980) evaluate incentive and insurance effects of social insurance provided by a progressive fiscal policy in a setting with ex ante representative households and missing credit and/or insurance markets. Typically, ex ante efficiency entails an interventionist fiscal policy which trades off incentive and insurance effects. At the interim or ex post stage, however, unanimous agreement is generally unlikely. Agents with positive income shocks who are required to subsidize others with negative shocks will prefer laissez faire to the interventions, with the opposite true for those with negative shocks. In contrast, the cross-subsidization in our context occurs across parents with the same ex post income across different ability realizations of their children, generating agreement at the interim stage (and also ex post, with sufficient altruism). Our paper thus helps explain the substantially larger consensus typically observed across classes and political parties regarding the desirability of government subsidization of elementary schooling, compared with fiscal policies that redistribute from rich to poor. Moreover, we show that education subsidies can be designed to offset redistributive or adverse incentive effects.

Subsequent literature in public economics has examined implications of redistributive tax distortions for education subsidies.[9] Bovenberg and Jacobs (2005) have argued in a static model without any borrowing constraints or income risk that redistributive taxes and education subsidies are 'Siamese twins': the latter are needed to counter the effects of the former in dulling educational incentives. Alternatively, the presence of progressive income taxes implies a 'fiscal externality' associated with education which enables agents to earn higher incomes and thereby pay higher taxes. Educational subsidies are needed to ensure that agents internalize this externality. Jacobs, Schindler and Yang (2012) show the same result obtains when the model is extended to a context with uninsurable income risk. Unlike our paper, these arguments for educational subsidies arise from pre-existing income tax distortions, which disappear in the case of a laissez faire status quo. None of these models incorporate ability heterogeneity and missing credit markets, which create an efficiency role for educational subsidies in our model, even in the absence of any progressive income taxes.

Dynamic models of investment in physical and/or human capital which incorporate missing credit and insurance markets and agent heterogeneity have been studied in the literature on macroeconomics and fiscal policy (Loury (1981), Aiyagari (1994), Aiyagari, Greenwood and Sheshadri (2002), Bénabou (1996, 2002)). Loury (1981) provided a pioneering analysis of human capital investments by altruistic parents in an environment with ability shocks and no financial markets. Most of his analysis concerned the characterization of dynamic properties of competitive equilibria. He showed that redistributive policies could raise aggregate output and welfare, but did not explore the efficiency properties of

[9]A large part of the recent dynamic public finance literature (e.g., Golosov, Tsyvinski and Werning (2006), or Golosov, Troshkin and Tsyvinski (2016)) is unrelated insofar as it abstracts from human capital investments and assumes that skills follow an exogenous Markov process. Its focus is to extend the Mirrlees (1971) optimal income tax model to a dynamic setting and examine consequences for optimal taxation of labor and savings. Other strands of literature on public education address its political economy. For instance, Glomm and Ravikumar (1992) study an endogenous growth model and compare human capital accumulation under laissez faire to public schooling funded by a linear income tax, where the tax rate is determined by a political majority. The focus there is on macroeconomic implications (per capita income, inequality and resulting tradeoffs), rather than the scope for efficiency enhancing interventions.

laissez faire equilibria. Aiyagari, Greenwood and Sheshadri (2002) study a model of human capital investment where education entails fixed and variable resource costs, besides child care. Education takes the form of increasing efficiency units of homogenous labor acquired by the child, as a function of the child's ability realization, parental resource and child care expenses. Apart from incorporating child care, the model is more general than Bénabou's or ours by incorporating physical capital and financial bequests. But the main focus of their paper is different: to characterize first-best Pareto efficient allocations which can be decentralized with complete markets, and contrast these to laissez faire allocations that result when there are no credit or insurance markets. They do not consider the effects of fiscal policy.[10]

Efficiency properties of competitive equilibria with endogenous human capital and missing financial markets are studied in Bénabou (1996). His model is more general than ours regarding parental labor supply decisions. On the other hand, attention is restricted to particular functional forms for utility and production, and specific distributions are assumed for ability and productivity shocks. These are realized after investment choices are made, which removes the key heterogeneity in our model: all agents with a given income level take identical investment decisions and have identical consumptions. In this case there is no point in redistributing within the same income/occupational class. However, given that incomes follow an ergodic process in equilibrium, individual investments are complements in production and utility is strictly concave in consumption, every dynasty prefers a positive amount of inter-occupational redistribution from a long-term perspective. In the short run, tax-funded education subsidies reduce consumption for rich dynasties but the greater the discount factor, the more agents find that the long-term gains dominate. So analogous to our result concerning ex post Pareto improving policies that redistribute within occupations, Bénabou finds that collective financing of education becomes a Pareto improvement

[10]By contrast, Phelan (2006) or Farhi and Werning (2007) assume a fiscal planner who fully controls agents' savings and designs a dynamic mechanism to provide insurance against time-varying taste shocks. They study implications of divergence between planner and household discount rates for long run inequality in ex ante efficient allocations.

in a sufficiently patient society.[11] The main contrast with our paper is that we focus on redistribution among parents in the same occupation who have children with different abilities, which achieves interim Pareto improvements without any constraints on intensity of parental altruism.

Versions of these models have been calibrated to fit data of real economies in order to evaluate the welfare and macroeconomic effects of various fiscal policies in numerical simulations (Heathcote (2005), Bohacek and Kapicka (2008), Cespedes (2014), Berriel and Zilberman (2011), Abbott, Gallipoli, Meghir and Violante (2013), Findeisen and Sachs (2015, 2016)). These studies rely on specific functional forms for technology and preferences, and focus on aggregate measures of welfare. Apart from the need to understand the source of these welfare effects (e.g., evaluating attendant insurance effects), these papers leave open the question whether there may exist other policies which could have resulted in a Pareto improvement, or what the effects might be in economies with different preferences and technology. Our paper complements this literature by providing purely qualitative results concerning efficient fiscal policies which apply irrespective of the specific welfare function, technology or preferences.

Our model is related to those studied in the literature on occupational choice and development (Banerjee and Newman (1993), Galor and Zeira (1993), Ljungqvist (1993), Freeman (1996), Aghion and Bolton (1997), Bandopadhyay (1997), Lloyd-Ellis and Bernhardt (2000), Matsuyama (2000, 2006), Ghatak and Jiang (2002), Mookherjee and Ray (2002, 2003, 2008, 2010), Mookherjee and Napel (2007)). With few exceptions, this literature focuses on macroeconomic outcomes rather than normative consequences. Mookherjee and Ray (2003) examine efficiency of steady states in an OLG model which abstracts from ability heterogeneity, and show the existence of laissez faire steady state allocations that are

[11] Bénabou (2002) specializes the production side of his earlier model while considering aggregate efficiency properties of a richer set of redistribution schemes. In both models, the income distribution is log-normal in any period, i.e., it has unbounded support. Proposition 4 in Bénabou (1996) hence establishes a Pareto improvement asymptotically (for $\delta \to 1$), while sufficient patience could be identified with $\delta \geq \bar{\delta}$ for some $\bar{\delta} < 1$ in models with bounded incomes like ours.

Pareto efficient. Our paper therefore shows that this result is not robust to the presence of ability heterogeneity. It complements our previous work (Mookherjee and Napel (2007)) which focused on uniqueness and stability properties of a closely related model with ability heterogeneity.

D'Amato and Mookherjee (2013) investigate efficiency properties of equilibria in a closely related model with ability heterogeneity, where the labor market is additionally characterized by signaling (i.e., productivity depends on ability in addition to education). They examine effects of educational loans provided by the government, funded by bonds released to the public. They obtain a result similar to our first result, viz. competitive equilibria are Pareto dominated by such a loan program. This intervention works differently from ours by changing the composition of the educated in favor of children from low-income families who have higher abilities than children from high income families. Per capita education and output in the economy are unchanged. Such interventions require parents to take loans on behalf of their children, unlike the interventions studied in the current paper.

Our paper also contributes to debates concerning the design of anti-poverty programs (e.g., Mookherjee (2006), Mookherjee and Ray (2008), Ghatak (2015)): whether transfers to poor households should be uniform/cash/unconditional rather than in kind/conditional on investments in human capital of children. While there are general arguments based on the Pareto criterion in favor of the former in static contexts – as in the Mirrlees (1971) or Atkinson-Stiglitz (1976) models – this no longer applies in dynamic settings when effects on investments need to be incorporated. Our results provide a general argument for superiority of conditional transfers in such settings.

6 Concluding Observations

We have provided theoretical arguments for Pareto-superiority of fiscal policies involving educational subsidies funded by income taxes imposed on the same income/occupational class. These dominate laissez faire outcomes, as well as policies where transfers are not conditioned on education decisions. The results apply quite generally, provided parents do not supplement education investments with financial bequests. In the presence of financial bequests, laissez faire outcomes continue to be Pareto dominated by similar policies applied only to poor households that do not leave financial bequests. For wealthy household classes that always leave financial bequests, Pareto optimality requires an opposite policy involving educational taxes or fees which fund unconditional transfers within the same class.

The main contribution of the paper is to establish results on the inefficiency of laissez faire equilibria in economies with incomplete financial markets and idiosyncratic abilities that depend little on detailed assumptions concerning preferences or technology, or on the nature of social preferences for redistribution. They provide suggestions for policies based only on the Pareto criterion, that would generate no distributional conflict and create rather than destroy incentives to invest. The analysis helps provide a better understanding of the source of estimated welfare effects of educational subsidy policies in calibrated macro models.

The investigated welfare state is a rather minimal one, partly because the model is stark and the policy objective is confined to Pareto improvements. Some cross-occupational redistribution is involved, but the major component of the proposed intervention operates at an intra-occupational level. The same non-monotonic pattern of consumption utility in a given income/occupational class has been demonstrated to arise in several extensions of the baseline model. Further generalizations are desirable but left for future research. For instance, a child's future wage income and financial inheritance could be subject to random shocks. Analysis of the corresponding extension of the scenario considered in Section 4.4 would be complicated by gains from diversifying the risky payoffs to financial vs. educational investment. For the very poor households who do not leave financial bequests, the identified pattern of parental consumption however should prevail, suggesting that a scheme along the indicated lines could still raise interim welfare.

One question we did not address is the underlying source of missing markets for credit or insurance. Couldn't members of, say, the unskilled occupation – or profit-maximizing companies – organize a similar kind of scheme as does the government in our model? Why is public intervention needed? Mutual aid and benefit societies, fraternal lodges, trade unions and guilds have historically provided many private insurance services that have been taken over – and to some extent crowded out – by the welfare state (see Beito (2000)). Such societies usually have better social monitoring and enforcement possibilities than commercial companies. Still, collective education financing at more than a very localized scale seems to have been the exception rather than the rule. One can only speculate what the underlying reasons may have been — adverse selection (associated with opportunistic non-participation of parents who do not expect to benefit from it ex post), or the general equilibrium effects of such schemes (which lower the profitability of private insurance firms and households owing to the induced changes in the skill premium in the labor market) that are neutralized by the government in our construction.

References

Abbott B., Gallipoli G., Meghir C. and Violante G.L. (2013). "Education policy and intergenerational transfers in equilibrium," Discussion Paper No. 1887, Cowles Foundation.

Aghion P. and Bolton P. (1997). "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies* 64(2), 151-172.

Aiyagari S.R. (1994), "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics* 109(3), 659–684.

Aiyagari S.R., Greenwood J. and Seshadri A. (2002), "Efficient Investment in Children," *Journal of Economic Theory* 102(2), 290–321.

Atkinson A. and Stiglitz J. (1976),"The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics* 6(1–2), 55–75.

Bandyopadhyay D. (1997). “Distribution of Human Capital and Economic Growth,” Working Paper No. 173, Department of Economics, Auckland Business School.

Banerjee A. and Newman A. (1993). “Occupational Choice and the Process of Development,” *Journal of Political Economy* 101(2), 274–298.

Barro R. and Becker G. (1989) “Fertility Choice in a Model of Economic Growth,” *Econometrica* 57(2), 481–501.

Becker G. and Tomes N. (1979), “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility,” *Journal of Political Economy* 87(6), 1153–89.

Beito D.T. (2000), *From Mutual Aid to the Welfare State – Fraternal Societies and Social Service, 1890–1967*, Chapel Hill, NC: University of North Carolina Press.

Bénabou R. (1996), “Heterogeneity, Stratification, and Growth: Macroeconomic implications of Community Structure and School Finance,” *American Economic Review* 86(3), 584–609.

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

Berriel T.C. and Zilberman E. (2011) “Targeting the Poor: A Macroeconomic Analysis of Cash Transfer Programs,” Working Paper No. 726, Fundação Getulio Vargas/EPGE.

Bohacek R. and Kapicka M. (2008), “Optimal human capital policies,” *Journal of Monetary Economics* 55(1), 1–16.

Bovenberg A.L. and Jacobs B. (2005), “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics* 89(11–12), 2005–2035.

Cespedes N. (2014) “General Equilibrium Analysis of Conditional Cash Transfers,” Working Paper No. 2014-25, Peruvian Economic Association.

D'Amato M. and Mookherjee D. (2013), "Welfare Economics of Educational Policy When Signaling and Credit Constraints Co-exist," Working Paper, University of Salerno and Boston University.

Farhi E. and Werning I. (2007), "Inequality and Social Discounting," *Journal of Political Economy* 115(3), 365–402.

Findeisen S. and Sachs D. (2015), "Designing Efficient College and Tax Policies," Working Paper, University of Mannheim and University of Cologne.

Findeisen S. and Sachs D. (2016), "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Journal of Public Economics* 138, 1-21.

Freeman S. (1996), "Equilibrium Income Inequality among Identical Agents," *Journal of Political Economy* 104(5), 1047–1064.

Galor O. and Zeira J. (1993), "Income Distribution and Macroeconomics," *Review of Economic Studies* 60(1), 35–52.

Ghatak M. and N. Jiang (2002), "A Simple Model of Inequality, Occupational Choice and Development," *Journal of Development Economics* 69(1), 205–226.

Ghatak M. (2015), "Theories of Poverty Traps and Anti-Poverty Policies," *World Bank Economic Review* 29, S77–105.

Glomm G. and Ravikumar B. (1992), "Public versus Private Investment in Human Capital: Endogenous Growth and Income Inequality," *Journal of Political Economy* 100(4), 818–834.

Golosov M., Troshkin M. and Tsyvinski A. (2016) "Redistribution And Social Insurance," *American Economic Review* 106(2), 359–386.

Golosov M., Tsyvinski A. and Werning I. (2006), "New Dynamic Public Finance: A

User's Guide," *NBER Macroeconomics Annual* 21, 317–388.

Heathcote J. (2005), "Fiscal Policy with Heterogeneous Agents and Incomplete Markets," *Review of Economic Studies* 72(1), 161–188.

Jacobs J., Schindler D. and Yang H. (2012), "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics* 114(3), 908–931.

Ljungqvist L. (1993), "Economic Underdevelopment: The Case of Missing Market for Human Capital," *Journal of Development Economics* 40(2), 219–239.

Lloyd-Ellis H. and Bernhardt D. (2000), "Enterprise, Inequality and Economic Development," *Review of Economic Studies* 67(1), 147–168.

Loury G. (1981), "Intergenerational Transfers and the Distribution of Earnings," *Econometrica* 49(4), pp. 843–867.

Matsuyama K. (2000), "Endogenous Inequality," *Review of Economic Studies* 67(4), 743–759.

Matsuyama K. (2006), "The 2005 Lawrence Klein Lecture: Emergent Class Structure," *International Economic Review* 47(2), 327–260.

Mirrlees J. (1971), "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies* 38(2), 175–208.

Mookherjee D. (2006), "Poverty Persistence and Design of Anti-Poverty Policies," in: Banerjee A., Bénabou R. and Mookherjee D. (eds.), *Understanding Poverty*, pp. 231–241, New York, NY: Oxford University Press.

Mookherjee D. and Napel S. (2007), "Intergenerational Mobility and Macroeconomic History Dependence," *Journal of Economic Theory* 137(1), 49–78

Mookherjee D. and Ray D. (2002), "Is Equality Stable?" *American Economic Review*

(Papers and Proceedings) 92(2), 253–259.

Mookherjee D. and Ray D. (2003), "Persistent Inequality," *Review of Economic Studies* 70(2), 369–394.

Mookherjee D. and Ray D. (2008), "A Dynamic Incentive-Based Argument for Conditional Transfers," *Economic Record* 84(S1): S2–16.

Mookherjee D. and Ray D.(2010), "Inequality and Markets: Some Implications of Occupational Diversity," *American Economic Journal: Microeconomics* 2(4), 38-76.

Phelan C. (2006), "Opportunity and Social Mobility," *Review of Economic Studies* 73(2), 487–504.

Sinn H.W. (1995), "A Theory of the Welfare State," *Scandinavian Journal of Economics* 97(4), 495–526.

Sinn H.W. (1996), "Social Insurance, Incentives and Risk Taking," *International Tax and Public Finance* 3(3), 259–280.

Varian H.R. (1980), "Redistributive Taxation as Social Insurance," *Journal of Public Economics* 14(1), 49–68.

Appendix: Proofs

Proof of Lemma 1: Part (i) follows from the fact that $w_{1t} > w_{0t}$ if and only if $\lambda_t < \bar{\lambda}$, and $V_t(1,x) > V_t(0,x)$ for any x if and only if $w_{1t} > w_{0t}$. If (ii) is false and $\lambda_t \geq \bar{\lambda}$ at some date, we have $V_t(1,x) = V_t(0,x)$ for all x, implying that no parent with a child with $x > 0$ will want to invest in education at $t-1$, so $\lambda_t = 0 < \bar{\lambda}$ – a contradiction.

For (iii) note that (3) follows straightforwardly from the optimization problem faced by parents. And $x_{0t} < x_{1t}$ follows from (i) and (ii) above. To show the next claim in (iv), suppose it is not true. Then we can find a subsequence $\{x_{c,t_n}\}_{n=1,2,\ldots}$ along which x_{c,t_n} for some occupation c either tends to 0 or ∞. In the former case, (3) implies $[W_{1,t_n+1} - W_{0,t_n+1}]$ must converge to 0, which in turn requires λ_{t_n+1} to converge to $\bar{\lambda}$. Then x_{d,t_n} must tend to 0 for *both* occupations $d = 0,1$, and (2) implies λ_{t_n+1} converges to 0 – a contradiction. In the latter case $[W_{1,t_n+1} - W_{0,t_n+1}]$ must converge to ∞, implying x_{d,t_n} must tend to ∞ for both occupations $d = 0,1$ by virtue of (3). Equation (2) then implies λ_{t_n+1} approaches 1. This contradicts (ii) above. Since $\lambda_t \geq F(x_{0t})$ (owing to (2) and $x_{1t} > x_{0t}$), it follows that λ_t is uniformly bounded away from 0. Moreover, the argument which ruled out that sequence $\{x_{ct}\}_{t=1,2,\ldots}$ has a cluster point at 0 also ensures λ_t is bounded away from $\bar{\lambda}$. The bounds on consumption follow from the bounds on wages and on investment thresholds. ■

Proof of Theorem 2:

A useful preliminary result shows that any government budget surplus can be disposed of in an ex post Pareto improving manner while leaving investment incentives unchanged.

Lemma 2 *Given any sequence of non-negative budgetary surpluses $\{R_t\}_{t=0,1,\ldots}$ resulting from a fiscal policy $\{\tau_{ct}, e_{ct}\}_{c;t}$ and an associated DCE $\{\lambda_t, w_{ct}, x_{ct}, \mathcal{U}_{ct}\}_{c;\,t}$, suppose that the surplus is strictly positive at some date. Then there exists another fiscal policy $\{\tau'_{ct}, e'_{ct}\}_{c;\,t}$ with $\tau'_{ct} > \tau_{ct}, e'_{ct} > e_{ct}$ for all $c = 0,1$ and $t = 0,1,\ldots$ with an associated DCE with the same skill ratios, wages and thresholds $\{\lambda_t, w_{ct}, x_{ct}\}_{c;\,t}$ which ex post Pareto dominates the* *original DCE, i.e., with $\mathcal{U}'_{ct} > \mathcal{U}_{ct}$ for all c, t.*

Proof of Lemma 2: Let the original DCE involve wages $\{w_{ct}\}_{t=0,1,2,...}$ and investment thresholds $\{x_{ct}\}_{t=0,1,2,...}$ in occupation c. For any period t and positive budgetary amount $R_{ct} \leq R_t$ to be disposed of to households in occupation c in t, select $\Delta\tau_{ct}(R_{ct}) \geq 0, \Delta e_{ct}(R_{ct}) \geq 0$ as defined by the unique solution to:

$$\begin{aligned} R_{ct} &= \alpha_{ct}[\Delta\tau_{ct} + F(x_{ct})\Delta e_{ct}] \\ u(w_{ct} + \tau_{ct}) - u(w_{ct} + \tau_{ct} + e_{ct} - x_{ct}) &= u(w_{ct} + \tau_{ct} + \Delta\tau_{ct}) \\ &\quad - u(w_{ct} + \tau_{ct} + \Delta\tau_{ct} + e_{ct} + \Delta e_{ct} - x_{ct}) \end{aligned} \tag{16}$$

where α_{ct} equals λ_t if $c = 1$ and $1 - \lambda_t$ otherwise. This results in a change in interim consumption utility of a household in occupation c in period t by

$$\begin{aligned} \Delta\mathcal{U}_{ct}(R_{ct}) &= \big[u(w_{ct} + \tau_{ct} + \Delta\tau_{ct}) - u(w_{ct} + \tau_{ct})\big](1 - F(x_{ct})) \\ &\quad + \int_0^{x_{ct}} \big\{u(w_{ct} + \tau_{ct} + \Delta\tau_{ct} + e_{ct} + \Delta e_{ct} - x) - u(w_{ct} + \tau_{ct} + e_{ct} - x)\big\} dF(x) \end{aligned}$$

provided the investment threshold remains x_{ct}.

$\Delta\tau_{ct}(R_{ct}), \Delta e_{ct}(R_{ct})$ and $\Delta\mathcal{U}_{ct}(R_{ct})$ are continuous, strictly increasing functions, taking the value 0 at $R_{ct} = 0$. By the Intermediate Value Theorem, for any $R_t > 0$ there exist R_{0t} and R_{1t} such that $R_{0t} + R_{1t} = R_t$ and $\Delta\mathcal{U}_{1t}(R_{1t}) = \Delta\mathcal{U}_{0t}(R_{0t})$. This ensures that $\mathcal{U}_{1t} - \mathcal{U}_{0t}$ is unchanged.

Because the definition of $\Delta\tau_{ct}$ and Δe_{ct} in (16) keeps investment sacrifices constant for threshold types x_{1t}, x_{0t}, the same investment strategies remain optimal for households in period t if they expect an unchanged welfare difference $W_{1,t+1} - W_{0,t+1}$. The sequence $\{W_{1t} - W_{0t}\}_{t=0,1,2,...}$ remains unchanged given that there is no change to the sequence of consumption utility differences $\{\mathcal{U}_{1t} - \mathcal{U}_{0t}\}_{t=0,1,2,...}$. The policy is constructed precisely to assure this, where preservation of the original investment thresholds also preserves skill ratios $\{\lambda_t\}_{t=1,2,...}$ and associated pre-tax wages $\{w_{1t}, w_{0t}\}_{t=1,2,...}$. The government budget is then balanced, while transfers to all households have increased. ■

The proof of Theorem 2 proceeds in five steps.

Step 1: Conditions (a)–(d) imply the status quo fiscal policy and DCE satisfy the following properties:

(i) λ_t is uniformly bounded away from 0 and 1;

(ii) x_{ct} is uniformly bounded above, and uniformly bounded away from zero;

(iii) consumptions of all agents are uniformly bounded.

To see this note that the bounds on income transfers and on marginal products g_c over $[0, 1]$ imply that post-tax incomes are uniformly bounded. These imply existence of: a uniform upper bound on consumption (since consumption is bounded above by post-tax income); a uniform upper bound on W_{ct} (given the upper bound on consumption); a uniform lower bound on W_{ct} (from the option of always consuming all post-tax income); and, consequently, a uniform upper bound on $\Delta W_t = W_{1t} - W_{0t}$.

The latter also is a uniform upper bound on the utility sacrifice of investing parents. Combined with the uniform bounds on post-tax incomes, we infer that investment thresholds x_{ct} are uniformly bounded above, which in turn implies equilibrium consumption is uniformly bounded from below. So (iii) holds. Condition (b) implies post-tax income differences between the skilled and unskilled occupation are bounded away from zero. Hence ΔW_t is uniformly bounded away from zero, implying the same for investment thresholds. This establishes (ii).

The first part of (ii) implies that λ_t is uniformly bounded away from 1 because the distribution of education costs has full support on $\mathbb{R}_+$. Moreover, full support and uniform positive lower bound on investment thresholds ensure that λ_t is uniformly bounded away from 0. So (i) holds.

Step 2: For arbitrary $\epsilon > 0$, construct the following policy change. Denote the status quo DCE by a * superscript. Choose any occupation c and select alternative fiscal policy $\tau'_{ct}(\epsilon) =$

$\tau_{ct} - \epsilon F(x^*_{ct}), e'_{ct}(\epsilon) = \epsilon(1-\mu_t)$ for this occupation, while leaving that for the other occupation $d \neq c$ unchanged, where

$$\mu_t \equiv (1 - F(x^*_{ct})) \left[1 - \frac{F(x^*_{ct}) u'(w^*_{ct} + \tau_{ct})}{\int_0^{x^*_{ct}} u'(w^*_{ct} + \tau_{ct} - x) dF(x)} \right] \tag{17}$$

independently of ϵ. It is evident that $\mu_t \in (0,1)$ for all t. By Step 1 and the concavity of u, it is uniformly bounded away from 0 and 1 respectively.

For either occupation $i \in \{c, d\}$, define post-reform post-tax wages: $w^\epsilon_{it} = w^*_{ct} + \tau_{ct} - \epsilon F(x^*_{ct})$ if $i = c$, and $w^*_{dt} + \tau_{dt}$ otherwise, and education subsidies: $e^\epsilon_{it} = \epsilon(1 - \mu_t)$ if $i = c$ and 0 otherwise. Let the corresponding investment thresholds be denoted x^ϵ_{it} and dynastic utilities be denoted W^ϵ_{it}. In the following we refer to the effect of a policy reform where the value of ϵ is raised slightly above 0.

Claim: Assuming that the policy change leaves after-tax wages unchanged for each occupation, it generates: (i) a positive first order increase in investment thresholds for parents in occupation c at every t; (ii) a zero first order effect on thresholds for parents in occupation d at every $t \geq 1$, and (iii) a zero first order effect on the dynastic utilities at every $t = 0, 1, 2, \ldots$. Specifically, $\frac{\partial x^\epsilon_{ct}(0)}{\partial \epsilon}$ is positive and uniformly (with respect to t) bounded away from zero, while $\frac{\partial x^\epsilon_{dt}(\epsilon)}{\partial \epsilon}$ and $\frac{\partial W^\epsilon_{it}(\epsilon)}{\partial \epsilon}, i = c, d$ converge uniformly (with respect to t) to zero as $\epsilon \downarrow 0$.

To prove this, we proceed as follows. For arbitrary thresholds $x_{it}, i \in \{c, d\}$ define the C^2 function

$$\mathcal{U}_{it}(x_{it}, \epsilon) \equiv u\big(w^\epsilon_{it}\big)[1 - F(x_{it})] + \int_0^{x_{it}} u\big(w^\epsilon_{it} + e^\epsilon_{it} - x\big) dF(x). \tag{18}$$

And define for $i \in \{c, d\}$:

$$W^\epsilon_{it} = \mathsf{E}_x V^\epsilon_t(i, x) \tag{19}$$

where

$$V^\epsilon_t(i, x) \equiv \max_{I \in \{0,1\}} \Big[u\big(w^\epsilon_{it} + I\{e^\epsilon_{it} - x\}\big) + \delta \mathsf{E}_x V^\epsilon_{t+1}(I, x) \Big]. \tag{20}$$

We restrict $\epsilon \leq \bar{\epsilon}$ for some bound $\bar{\epsilon} < \infty$. Given any such bound, it is evident that $w^\epsilon_{it}, e^\epsilon_{it}$ and therefore consumptions are uniformly bounded above (i.e., for all $i \in \{c, d\}$, all t and all $\epsilon \leq \bar{\epsilon}$). (In what follows, statements concerning uniform bounds will be taken to mean this.) Hence W_{it}^{ϵ} is also uniformly bounded above.

Next, note that W_{it}^{ϵ} is also uniformly bounded below, since a parent at any t and in any occupation i always has the option of not investing in education for its child and consuming its (uniformly bounded) post-tax income. So (given the intertemporal consistency of dynastic utility) the present value dynastic payoff associated with the parent as well as all subsequent descendants never investing forms a lower bound to W_{it}^{ϵ}.

Define $\Delta W_t^{\epsilon} \equiv W_{1t}^{\epsilon} - W_{0t}^{\epsilon}$, which constitutes the post-reform return to any dynasty to investing at $t-1$. The preceding arguments imply that this return ΔW_t^{ϵ} is uniformly bounded above. So parents' sacrifices associated with investment must be uniformly bounded above, implying that post-reform consumptions will be uniformly bounded from below.

In what follows, we shall say that a family of real-valued functions $\{y_t^{\epsilon}; \epsilon \in [0, \bar{\epsilon}), t \geq 0\}$ *satisfies the Cauchy (C) property* if given any $\eta > 0$, there exists $\zeta > 0$ such that $\sup_t |y_t^{\epsilon_1} - y_t^{\epsilon_2}| < \eta$ whenever $\epsilon_1, \epsilon_2 < \zeta$.

Since $|w_{it}^{\epsilon_1} - w_{it}^{\epsilon_2}| \leq (\epsilon_1 + \epsilon_2) F(x_{it}^*) \leq 2\bar{\epsilon}$, it follows that w_{it}^{ϵ} satisfies the C-property. By a similar argument, e_{it}^{ϵ} also satisfies the C-property.

We claim that W_{it}^{ϵ} satisfies the C-property. To prove this, let us define consumptions conditional on investment I and ability draw x as follows: $c_{it}^{\epsilon}(I, x) \equiv w_{it}^{\epsilon} + I[e_{it}^{\epsilon} - x]$. Then $|c_{it}^{\epsilon_1}(I, x) - c_{it}^{\epsilon_2}(I, x)| = \left|w_{it}^{\epsilon_1} - w_{it}^{\epsilon_2} + I[e_{it}^{\epsilon_1} - e_{it}^{\epsilon_2}]\right| < 2\eta$ for all i, t if $\epsilon_1, \epsilon_2 < \zeta$. Since consumptions are uniformly bounded, marginal utility is uniformly bounded above and away from zero. Hence for any given η^*, there exists $\bar{\epsilon}$ such that $|u(c_{it}^{\epsilon_1}(I, x)) - u(c_{it}^{\epsilon_2}(I, x))| < \eta^*$ for all I, t, x, i if $\epsilon_1, \epsilon_2 < \bar{\epsilon}$. This implies that $|V_{it}^{\epsilon_1}(i, x) - V_{it}^{\epsilon_2}(i, x)| < \frac{\eta^*}{1-\delta}$ if $\epsilon_1, \epsilon_2 < \bar{\epsilon}$ for all i, x. Hence W_{it}^{ϵ} satisfies the C-property.

Next, we claim for suitable choice of $\bar{\epsilon}$, ΔW_t^{ϵ} is uniformly bounded away from zero. The argument used in Step 1 established existence of $b > 0$ such that $\Delta W_t^0 > b$ for all t. The claim in the previous paragraph implies that ΔW_t^{ϵ} satisfies the C-property. Hence there exists $\zeta > 0$ such that $|\Delta W_t^\epsilon - \Delta W_t^0| < \frac{b}{2}$ for all t if $\epsilon < \zeta$. This implies $\Delta W_t^\epsilon > \frac{b}{2}$ for all t if $\epsilon < \zeta$.

Now investment threshold x_{it}^ϵ solves:

$$u(w_{it}^\epsilon) - u(w_{it}^\epsilon + e_{it}^\epsilon - x_{it}^\epsilon) = \delta \Delta W_{t+1}^\epsilon \tag{21}$$

which is a well-defined positive real number owing to the Implicit Function Theorem. Next, observe that x_{it}^ϵ satisfies the C-property, since $w_{it}^\epsilon, e_{it}^\epsilon, \Delta W_t^\epsilon$ all satisfy the C-property, and since marginal utility is uniformly bounded away from zero (since consumption is uniformly bounded above). Since x_{it}^0 is uniformly bounded away from zero, the same must be true for x_{it}^ϵ.

The preceding results imply that $x_{it}^\epsilon, \Delta W_t^\epsilon, w_{it}^\epsilon, e_{it}^\epsilon$ converge uniformly to $x_{it}^*, \Delta W_t^0, w_{it}^*, e_{it}$ respectively as ϵ goes to zero. Hence $\mathcal{U}_{it}(x_{it}^\epsilon, \epsilon)$ converges uniformly to $\mathcal{U}_{it}(x_{it}^*, 0)$, i.e., for every $\eta > 0$ there exists a $\zeta > 0$ such that $|\mathcal{U}_{it}(x_{it}^\epsilon, \epsilon) - \mathcal{U}_{it}(x_{it}^*, 0)| < \eta$ if $\epsilon < \zeta$.

Next, note that for any y

$$\frac{\partial \mathcal{U}_{ct}(y,0)}{\partial \epsilon} = -[1 - F(y)] u'(w_{ct}^\epsilon) F(x_{ct}^*) + [1 - \mu_t - F(x_{ct}^*)] \int_0^y u'(w_{it}^\epsilon + e_{it}^\epsilon - x) dF(x) \tag{22}$$

while $\frac{\partial \mathcal{U}_{dt}(y,0)}{\partial \epsilon} = 0$. So the preceding results and the definition of μ_t imply that $\frac{\partial \mathcal{U}_{it}(x_{it}^\epsilon, \epsilon)}{\partial \epsilon}$ converges uniformly to 0.

Moreover

$$\Delta W_t^\epsilon \equiv \sum_{k=0}^{\infty} \nu_k^\epsilon \Big[\mathcal{U}_{1,t+k}(x_{1,t+k}^\epsilon, \epsilon) - \mathcal{U}_{0,t+k}(x_{0,t+k}^\epsilon, \epsilon) \Big] \tag{23}$$

where $\nu_0^\epsilon \equiv 1$ and $\nu_k^\epsilon \equiv \delta^k \Pi_{l=0}^{k-1} [F(x_{1,t+l}^\epsilon) - F(x_{0,t+l}^\epsilon)]$. Since $\nu_k^\epsilon \leq \delta^k < \delta < 1$, the preceding results allow the order of summation and differentiation operators to be interchanged in the following expression:

$$\frac{\partial \Delta W_t^\epsilon}{\partial \epsilon} = \sum_{k=0}^{\infty} \nu_k^\epsilon \Big[\frac{\partial \mathcal{U}_{1,t+k}(x_{1,t+k}^\epsilon, \epsilon)}{\partial \epsilon} - \frac{\partial \mathcal{U}_{0,t+k}(x_{0,t+k}^\epsilon, \epsilon)}{\partial \epsilon} \Big] \tag{24}$$

where we use the Envelope Theorem to ignore the effects of changes in ϵ on the optimally chosen (interior) investment thresholds. It follows that $\frac{\partial \Delta W_t^\epsilon}{\partial \epsilon}$ converges uniformly to 0 as ϵ goes to 0. Moreover, by a similar argument, since $\frac{\partial \mathcal{U}_{it}(x_{it}^\epsilon, \epsilon)}{\partial \epsilon}$ converges uniformly to 0, the same is true of $\frac{\partial \mathcal{W}_{it}^\epsilon}{\partial \epsilon}$.

As we vary ϵ from 0, we claim that the threshold x_{ct} undergoes a first order increase, while the first order change in x_{dt} is zero. To see this, differentiating (21) for $i = d$ yields $\frac{\partial x_{dt}^\epsilon(0)}{\partial \epsilon} = 0$ because $\frac{\partial \Delta W_t^\epsilon(0)}{\partial \epsilon} = 0$. Moreover, from the uniform convergence of $\frac{\partial \Delta W_t^\epsilon}{\partial \epsilon}$, we can conclude that $\frac{\partial x_{dt}^\epsilon(\epsilon)}{\partial \epsilon}$ converges uniformly to 0 as ϵ goes to 0. In contrast, for $i = c$ we obtain

$$\frac{\partial x_{ct}^\epsilon(0)}{\partial \epsilon} = F(x_{ct}^*)\frac{u'(w_{ct}^* + \tau_{ct})}{u'(w_{ct}^* + \tau_{ct} - x_{ct}^*)} + (1 - \mu_t - F(x_{ct}^*)). \tag{25}$$

The concavity of u implies

$$\int_0^{x_{ct}^*} u'(w_{ct}^* + \tau_{ct} - x)dF(x) < F(x_{ct}^*)u'(w_{ct}^* + \tau_{ct} - x_{ct}^*). \tag{26}$$

Hence recalling the definition (17) of μ_t,

$$\mu_t < (1 - F(x_{ct}^*))\Big[1 - \frac{u'(w_{ct}^* + \tau_{ct})}{u'(w_{ct}^* + \tau_{ct} - x_{ct}^*)}\Big], \tag{27}$$

and substituting this into (25) we obtain

$$\frac{\partial x_{ct}^\epsilon(0)}{\partial \epsilon} > \frac{u'(w_{ct}^* + \tau_{ct})}{u'(w_{ct}^* + \tau_{ct} - x_{ct}^*)} \tag{28}$$

which is uniformly bounded away from 0. This concludes the proof of the Claim.

Step 3: In order to ensure that after-tax wages remain at their original levels, we introduce a wage neutralization policy at each t. First, for any $\epsilon \geq 0$ and $t \geq 0$, recursively define the skill ratio that would be induced in period $t+1$ by the investment thresholds $x_{ct}^\epsilon(\epsilon), x_{dt}^\epsilon(\epsilon)$

$$\lambda_{t+1}(\epsilon) = F(x_{1t}^\epsilon(\epsilon))\lambda_t(\epsilon) + F(x_{0t}^\epsilon(\epsilon))(1 - \lambda_t(\epsilon)) \tag{29}$$

with $\lambda_0(\epsilon) = \lambda_0$ given. Note that (28) combined with $x_{1t}^\epsilon(\epsilon) > x_{0t}^\epsilon(\epsilon)$ at all t implies that $\lambda'_{t+1}(0)$ is positive (and uniformly bounded away from 0).

Now switch to the following modified policy $\big(\tilde{\tau}_{ct}(\epsilon), \tilde{\tau}_{dt}(\epsilon), \tilde{e}_{ct}(\epsilon), \tilde{e}_{dt}(\epsilon)\big)$ for each period $t \geq 1$

$$\begin{aligned}
\tilde{\tau}_{ct}(\epsilon) &= w^*_{ct} - w_{ct}(\epsilon) + \tau'_{ct}(\epsilon) \equiv w^*_{ct} - w_{ct}(\epsilon) + \tau_{ct} - \epsilon F^*_{ct} && (30)\\
\tilde{e}_{ct}(\epsilon) &= e'_{ct}(\epsilon) \equiv \epsilon(1-\mu_t) && (31)\\
\tilde{\tau}_{dt}(\epsilon) &= w^*_{dt} - w_{dt}(\epsilon) + \tau'_{dt} \equiv w^*_{dt} - w_{dt}(\epsilon) + \tau_{dt} && (32)\\
\tilde{e}_{dt}(\epsilon) &= e'_{dt}(\epsilon) \equiv 0 && (33)
\end{aligned}$$

where $w_{ot}(\epsilon) = g_o(\lambda_t(\epsilon))$, $o = c, d$.

This modified policy induces a DCE with skill ratios $\{\lambda_t(\epsilon)\}_{t=1,2,...}$, investment thresholds $\{x^{\epsilon}_{ct}(\epsilon), x^{\epsilon}_{dt}(\epsilon)\}_{t=0,1,2...}$ and the interim utilities $\{\mathcal{U}_{ct}(x^{\epsilon}_{ct}(\epsilon),\epsilon), \mathcal{U}_{dt}(x^{\epsilon}_{dt}(\epsilon),\epsilon)\}_{t=0,1,2,...}$ which were constructed in Step 2 under the assumption of unchanged after-tax wages in each occupation at each date. Given investment thresholds $x^{\epsilon}_{ct}(\epsilon), x^{\epsilon}_{dt}(\epsilon)$ the resulting skill ratio is $\lambda_{t+1}(\epsilon)$ and hence pre-tax wages are $g_c(\lambda_{t+1}(\epsilon)), g_d(\lambda_{t+1}(\epsilon))$. The transfers defined by (30)–(33), therefore, ensure that the household's optimization problem in each period corresponds to the one under original wages $\{w^*_{1t}, w^*_{0t}\}_{t=0,1,2,...}$ and the policy $\{\tau'_c(\epsilon), \tau'_d(\epsilon), e'_c(\epsilon), e'_d(\epsilon)\}_{t=0,1,2,...}$.

Step 4: We next check that there is a first order improvement in government revenues at every t. Supposing that $c = 1, d = 0$ (an analogous argument works for the opposite case), the budget surplus for $t \geq 0$ is

$$\begin{aligned}
B_t(\epsilon) &= w_{0t} - w^*_{0t} - \tau_{0t}\\
&\quad + \lambda_t\big[(\tau_{0t} - \tau_{1t}) + (w^*_{0t} - w_{0t}) - (w^*_{1t} - w_{1t}) - \epsilon\{F_{1t}(1-\mu) - F^*_{1t}\}\big] \qquad (34)
\end{aligned}$$

where we abbreviate $F(x^{\epsilon}_{it})$ and w^{ϵ}_{it} by F_{it} and w_{it}, respectively. Indicating the corresponding derivatives w.r.t. ϵ by F'_{it} and w'_{it}, we then have

$$\begin{aligned}
B'_t(\epsilon) &= w'_{0t} + \lambda'_t\big[(\tau_{0t} - \tau_{1t}) + (w^*_{0t} - w_{0t}) - (w^*_{1t} - w_{1t}) - \epsilon\{F_{1t}(1-\mu_t) - F^*_{1t}\}\big]\\
&\quad - \lambda_t\big[w'_{0t} - w'_{1t} + \{F_{1t}(1-\mu_t) - F^*_{1t}\} + \epsilon F'_{1t}(1-\mu_t)\big]. \qquad (35)
\end{aligned}$$

We can use Euler's theorem to obtain $\lambda_t w'_{1t} + (1-\lambda_t)w'_{0t} \equiv 0$, i.e., the corresponding terms in (35) cancel. The progressivity assumption $\tau_{0t} \geq \tau_{1t}$ in condition (a) and $\lambda'_t \geq 0$ for all

ϵ sufficiently close to zero imply we may drop the corresponding term to bound (35) from below:

$$\begin{aligned} B_t'(\epsilon) &= \lambda_t F_{1t}\mu_t + \lambda_t'\big[(w_{0t}^* - w_{0t}) - (w_{1t}^* - w_{1t}) - \epsilon\{F_{1t} - F_{1t}^*\} + \epsilon F_{1t}\mu_t\big] \\ &\quad -\lambda_t\big[\{F_{1t} - F_{1t}^*\} + \epsilon F_{1t}'(1-\mu_t)\big]. \end{aligned} \tag{36}$$

The first term, $\lambda_t F_{1t}\mu_t$, is positive and uniformly bounded away from zero. We want to show that the rest of the RHS can be brought arbitrarily close to zero by choosing $\underline{\epsilon} > 0$ sufficiently small, whence we could conclude that $B_t'(\epsilon) > 0$ for all t and all $\epsilon \in [0, \underline{\epsilon})$.

We claim that λ_t' is uniformly bounded. Apply $\lambda_{t+1}' = F_{1t}'\lambda_t + F_{0t}'(1-\lambda_t) + (F_{1t} - F_{0t})\lambda_t'$ recursively to obtain

$$\lambda_{t+1}' = F_{1t}'\lambda_t + F_{0t}'(1-\lambda_t) + \sum_{k=1}^{t}(F_{1t} - F_{0t})\cdot\ldots\cdot(F_{1,t-k+1} - F_{0,t-k+1})[F_{1,t-k}'\lambda_{t-k} + F_{0,t-k}'(1-\lambda_{t-k})]. \tag{37}$$

Now F_{1t}' and F_{0t}' are uniformly bounded since the density of the ability distribution is bounded, and preceding arguments imply that $\frac{\partial x_{it}^\epsilon}{\partial \epsilon}$ is uniformly bounded. Hence any convex combination of F_{1t}' and F_{0t}' is uniformly bounded. So there exists $m > 0$ such that $[F_{1,t-k}'\lambda_{t-k} + F_{0,t-k}'(1-\lambda_{t-k})] < m$ for all $t \geq k$ if $\epsilon \in [0, \underline{\epsilon})$. There also exists $\alpha \in (0,1)$ such that $(F_{1,t-k+1} - F_{0,t-k+1}) < \alpha$ for all $t \geq k$ and all $\epsilon \in [0, \bar{\epsilon}]$. Hence $\lambda_{t+1}' < \frac{m}{1-\alpha}$, i.e., is uniformly bounded.

Choosing $\underline{\epsilon} > 0$ small enough therefore allows to make $\lambda_t'[(w_{0t}^* - w_{0t}) - (w_{1t}^* - w_{1t})]$ arbitrarily small: the skill change $\lambda_t^* - \lambda_t(\epsilon)$ can be made arbitrarily small; this extends to wage changes.

Given that $\lambda_t' F_{1t}\mu_t$ and $F_{1t}'(1-\mu_t)$ are uniformly bounded, products of ϵ and these terms in (36) can be made arbitrarily small for $\epsilon \in [0, \underline{\epsilon})$ through an appropriate choice of $\underline{\epsilon} > 0$. Because $\lambda_t F_{1t}\mu_t$ is uniformly bounded away from zero, this in summary means we can find $\underline{\epsilon} > 0$ such that

$$B_t'(\epsilon) \geq \frac{1}{2}\cdot\lambda_t F_{1t}\mu_t \tag{38}$$

for all t and $\epsilon \in [0, \underline{\epsilon})$. As $\lambda_t F_{1t}\mu_t$ is uniformly bounded away from zero, a small policy reform starting from $\epsilon = 0$ therefore generates a strictly positive budget surplus.

Step 5: Finally, apply Lemma 2 in order to dispose of the resulting budget surplus in an interim Pareto-improving way. $\square$

Proof of Theorem 3

We denote the respective investment thresholds, skill ratios, wages, etc. in the DCELF associated with a given discount factor δ by x_{ct}^{δ}, λ_t^{δ}, w_{ct}^{δ}, etc. Recall that the induced skill ratio λ_t^{δ} must be strictly smaller than $\bar{\lambda}$ for every δ and t (cf. Lemma 1). From this follows that if we fix an arbitrary $\delta' > 0$ there exist $\underline{x}$ and $\bar{x}$ such that $0 < \underline{x} \leq x_{0t}^{\delta} \leq x_{1t}^{\delta} \leq \bar{x} < \infty$ for all t and all $\delta \in (\delta', 1)$. To see this, suppose otherwise, i.e., that there exists a sequence $\{t_n, \delta_n\}_{n=1,2,\ldots}$ such that (i) $x_{c,t_n}^{\delta_n} \to 0$ or (ii) $x_{c,t_n}^{\delta_n} \to \infty$. In case (i), vanishing investment by occupation c in period t_n requires that the benefit of having a skilled child in $t_n + 1$ vanishes. Then no parent in occupation $d \neq c$ would have an incentive to invest in t_n either, implying $\lambda_{t_n+1} \approx 0$. The consequent gap between skilled and unskilled equilibrium wages in period $t_n + 1$ and $\delta_n > \delta' > 0$ would then, however, induce a non-vanishing benefit of one's child to be skilled in $t_n + 1$ – a contradiction. In case (ii), benefits of having a skilled child in $t_n + 1$ would need to grow without bound. This implies that parents in occupation $d \neq c$ will also find it optimal to invest in t_n for arbitrarily large x. But $x_{0,t_n}^{\delta}, x_{1,t_n}^{\delta} \to \infty$ implies $\lambda_{t_n+1}^{\delta} \to 1$, in contradiction to $\lambda_t^{\delta} < \bar{\lambda}$ for all t.

We next establish that there exist $\delta' > 0$ and $\underline{b}$ such that

$$F(x_{1t}^{\delta})\lambda_t^{\delta}\mu_t^{\delta} \geq \underline{b} > 0 \text{ for all } t \text{ and } \delta \in (\delta', 1) \tag{39}$$

with

$$\mu_t^{\delta} \equiv (1 - F(x_{ct}^{\delta}))\left[1 - \frac{F(x_{ct}^{\delta})u'(w_{ct}^{\delta})}{\int_0^{x_{ct}^{\delta}} u'(w_{ct}^{\delta} - x)dF(x)}\right] \tag{40}$$

for a given $c \in \{0, 1\}$. To see (39) note, first, that $F(x_{1t}^{\delta}) \geq F(\underline{x}) > 0$ because F is strictly increasing. Second, $\lambda_t^{\delta} < \bar{\lambda}$ implies $w_{1t}^{\delta} > w_{0t}^{\delta}$ and $x_{1t}^{\delta} > x_{0t}^{\delta}$. From this follows $\lambda_t^{\delta} \geq F(x_{0t}^{\delta}) \geq F(\underline{x}) > 0$. Finally, note that $(1 - F(x_{ct}^{\delta})) \geq (1 - F(\bar{x})) > 0$ in (40). Now suppose that

$$\frac{u'(w_{ct_n}^{\delta_n})}{\frac{1}{F(x_{ct_n}^{\delta_n})}\int_0^{x_{ct_n}^{\delta_n}} u'(w_{ct_n}^{\delta_n}-x)dF(x)} \to 1$$

for some sequence $\{t_n, \delta_n\}_{n=1,2,\ldots}$. This would require $x_{ct_n}^{\delta_n} \to 0$, in contradiction to $x_{ct}^{\delta} \geq \underline{x} > 0$ for all t and $\delta \in (\delta', 1)$. Hence μ_t^δ is bounded away from zero.

Equation (39) implies that, by choosing $\epsilon \in (0, \bar{\epsilon})$ for a small $\bar{\epsilon} > 0$, a strictly positive budget surplus $B_t^\delta(\epsilon)$ is created for all t and $\delta \in (\delta', 1)$ in the first steps of the proof of Theorem 2. Since the full policy intervention is budget balancing, it must raise consumption in period t for at least one occupation $d \in \{0, 1\}$ by a non-vanishing amount. This must increase the respective interim consumption utility $\mathcal{U}_{dt}^\delta$ at a rate which is bounded away from zero, recalling that $g_1(0)$ is an upper bound to consumption in any laissez faire equilibrium and so every agent's marginal utility of consumption is bounded below by $u'(g_1(0)) > 0$. The final Step 5 of the proof of the Theorem 2 makes sure that differences between $\mathcal{U}_{1t}^\delta$ and $\mathcal{U}_{0t}^\delta$ in the status quo are preserved. Therefore, also interim consumption utility in occupation $c \neq d$ must locally increase in ϵ at a rate that is bounded away from zero. Let $\underline{\nu} > 0$ denote the corresponding uniform lower bound on marginal improvements in expected utility associated with period t consumption.

Increases in period t's average consumption utility of the skilled and unskilled arise mainly because parents who educated their children already in the laissez-faire benchmark consume their net subsidy. For them, the consumption and the dynastic components of utility go up. The same applies to non-investors in occupation d if the first steps of the scheme are targeted at only one occupation c (because of transfers everyone receives in Step 5). However, non-investors in the targeted occupation are net contributors and have to reduce consumption. We will establish that the policy constitutes an ex post Pareto improvement by showing that, for δ high enough, their gain in dynastic utility outweighs the loss in consumption utility. Note here that newly investing parents in the targeted occupation suffer an even bigger drop in consumption – despite being net beneficiaries – than their non-investing peers. But they could have stayed non-investors and reveal to be yet better off by investing.

The key reason why also net contributors to the scheme are better off for high δ is that the rate at which $\epsilon > 0$ decreases these non-investing parents' consumption utility is bounded above. Namely, there exists $L < \infty$ such that

$$\frac{\partial}{\partial \epsilon}\Big[u(w_{ct}^{\delta}) - u\big(w_{ct}^{\delta} - \epsilon F(x_{ct}^{\delta}) + S_{ct}^{\delta}(\epsilon)\big)\Big]\bigg|_{\epsilon=0} = \left(F(x_{ct}^{\delta}) - \frac{\partial S_{ct}^{\delta}(0)}{\partial \epsilon}\right) \cdot u'(w_{ct}^{\delta}) \tag{41}$$

$$\leq F(\bar{x}) \cdot u'(g_0(0)) < L \tag{42}$$

where $S_{ct}^{\delta}(\epsilon)$ denotes the (increasing) budget surplus which is allocated to occupation c in Step 5 of the proof of Theorem 2.

In contrast, the rate at which $\epsilon > 0$ increases the non-investing parents' dynastic utility $\delta\, W_{0,t+1}^{\delta}$ is unbounded. Namely, for every $M < \infty$ there exists $\underline{\delta} \in (0,1)$ such that for all $\delta \in (\underline{\delta}, 1)$:

$$\frac{\partial}{\partial \epsilon}\Big\{\delta\Big[W_{0,t+1}^{\delta}(\epsilon) - W_{0,t+1}^{\delta}\Big]\Big\}\bigg|_{\epsilon=0} > M \tag{43}$$

where $\mathcal{W}_{0,t+1}^{\delta} = \mathcal{W}_{0,t+1}^{\delta}(0)$ refers to interim welfare in the original DCELF. To see this, consider

$$\frac{\partial}{\partial \epsilon}\Big[W_{0,t+1}^{\delta}(\epsilon) - W_{0,t+1}^{\delta}\Big] = \frac{\partial}{\partial \epsilon}\left[\sum_{k=0}^{\infty} \delta^k \mathcal{U}_{0t}^{\delta}(\epsilon) + \delta \sum_{k=0}^{\infty} \delta^k F(x_{0,t+k}^{\delta}(\epsilon))\Delta W_{t+1+k}^{\delta}(\epsilon)\right] \tag{44}$$

and note that the derivative at $\epsilon = 0$ of the second summand in the brackets is zero (cf. Step 2 in Proof of Theorem 2). The corresponding derivative of the first summand is

$$\sum_{k=0}^{\infty} \delta^k \frac{\partial \mathcal{U}_{0t}^{\delta}}{\partial \epsilon}\bigg|_{\epsilon=0} \geq \sum_{k=0}^{\infty} \delta^k \underline{\nu} = \frac{\underline{\nu}}{1-\delta}, \tag{45}$$

and so the left-hand side of (43) grows without bound as $\delta \to 1$.

Combining (41) and (43), we can conclude that the total welfare change of non-investing parents satisfies

$$\frac{\partial}{\partial \epsilon}\left\{\Big[u(w_{ct}^{\delta} - \epsilon F(x_{ct}^{\delta}) + S_{ct}^{\delta}(\epsilon) + \delta\, W_{0,t+1}^{\delta}(\epsilon)\Big] - \Big[u(w_{ct}^{\delta}) + \delta\, W_{0,t+1}^{\delta}\Big]\right\}\Bigg|_{\epsilon=0} \geq \psi \tag{46}$$

for all t and $\delta \in (\underline{\delta}, 1)$ for some $\psi > 0$. We can therefore choose $\bar{\epsilon} > 0$ such that each individual's ex post welfare change is positive for any $\epsilon \in (0, \bar{\epsilon})$ for every $\delta \in (\underline{\delta}, 1)$. $\square$