Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance

By ROLAND BÉNABOU *

This paper examines how socioeconomic stratification and alternative systems of education finance affect inequality and growth. Agents interact through local public goods or externalities (school funding, neighborhood effects) and economy-wide linkages (complementary skills, knowledge spillovers). Sorting families into homogeneous communities often minimizes the costs of existing heterogeneity, but mixing reduces heterogeneity faster. Integration therefore tends to slow down growth in the short run yet raise it in the long run. A move to state funding of education presents society with a similar intertemporal trade-off. Local and global complementarities play major roles in determining the efficient social and educational structures. (*JEL* D31, I22, O40, J24)

This paper studies how economic stratification affects inequality and growth, and how alternative systems of education finance contribute to these dynamics. In the United States a person's income, education, ethnic background, and lifestyle can be predicted quite accurately from his zip code. Gated communities are multiplying while in other neighborhoods poverty is becoming entrenched. This high degree of socioeconomic segregation has combined with a system where primary and secondary education is largely financed from property taxes, resulting in considerable disparities in school districts' resources. A profound legal and political debate has ensued, pitting the proponents of equalization against the advocates of home rule. Nearly half the states are currently facing constitutional challenges to their education financing systems. Compounding the inequality in school funding, there has been an even greater divergence in the norms of behavior, role models and values to which the young from different social strata are exposed during their formative years.[1]

The production of goods and services thus brings together in factories and offices, workers on one hand, managers and professionals on the other, whose upbringing and levels of human capital are becoming increasingly disparate. This polarization has been put forth as a contributing factor, not only to the rise of income inequality over the last two decades, but also to the productivity slowdown. One reads for instance in the MIT Commission's Report on Industrial Productivity (Michael Dertouzos et al., 1989; pp. 84–85)

> American and foreign students differ not only in their average scores on standardized tests, but also in the dispersion of

* Department of Economics, New York University, 269 Mercer Street, New York, NY 10003, NBER and CEPR. I am grateful to Olivier Blanchard, Ricardo Caballero, Peter Diamond, Steven Durlauf, John Heaton, Richard Romano, Jacques Thisse, Jean Tirole, Yoram Weiss and especially Julio Rotemberg for helpful discussions. I also thank Preston McAfee, Paul Milgrom and three referees for many useful comments. Financial support from the National Science Foundation (SES92-09267) is gratefully acknowledged.

[1] Martin Weiss (1989) provides a fascinating description of the socioeconomic "clusters" used in commercial and political marketing, and Jonathan Kozol (1991) a stark account of local disparities in school resources. Using Census tract data for 1970, 1980, and 1990, Paul Jargowsky (1995) demonstrates that significant nationwide increases in economic segregation occurred during the seventies and especially the eighties. The evidence on how neighborhood and school composition affect the young's educational and socioeconomic outcomes is surveyed by Christopher Jencks and Susan Mayer (1990). The evidence on the role of purchased educational inputs is surveyed by Eric Hanushek (1986).

those scores around the mean. The Japanese aim at bringing all students to a high common level of competence, and they are largely successful; as a result ... new entrants to the Japanese work force are generally literate, numerate, and prepared to learn. In the U.S. work force, employers have discovered high rates of illiteracy and difficulty with basic mathematics and reading in workers with high school diplomas ... Only a tiny fraction of young Americans are technologically literate and have some knowledge of foreign societies.

This paper provides a framework in which these issues can be formally analyzed. I model the dynamics of income distribution, human capital and productivity growth in a class of economies with both local public goods or externalities, such as community-funded schools or neighborhood effects, and economy-wide linkages, such as complementarity in production or knowledge spillovers. Two main questions are then addressed. I first ask which form of social organization is more efficient: stratification of communities and schools by economic status, or integration? The second issue stems directly from the policy debate: should education should be funded privately, locally, or nationally? This turns out to be, in large part, a special case of the previous question.

Whether sorting or mixing is more conducive to growth depends on the interplay of two effects, which can give rise to an interesting intertemporal trade-off. The first effect measures how efficient each social structure is at *processing heterogeneity,* that is, at aggregating disparate levels of human capital into the production of goods and, ultimately, new knowledge. Stratification tends to minimize the drag on growth from any given amount of heterogeneity, especially when family background and community quality are complements in a child's education. The second effect is dynamic: an integrated society is better at *reducing heterogeneity.* It converges faster to a homogeneous outcome, or in the presence of shocks, to a less unequal distribution of skills. Since mixing delivers its payoff only gradually it may slow down growth in the short run, yet enhance it in the long run. Initially, rich families lose more than poor ones gain, but eventually all incomes rise, resulting in a Pareto improvement if agents have a low enough discount rate.

Naturally, it is not always the case that mixing is more efficient in the long run. For instance, stratification remains preferable if the degree of *complementarity* between individuals' levels of human capital is sufficiently stronger in *local* interactions than in *global* interactions. Intuitively, this means that disparities in knowledge at the community level (for example, in schooling) entail sufficiently greater losses than at the aggregate level (for example, in production). What this paper offers is thus a framework in which the costs and benefits of stratification can be spelled out and the critical parameters identified. In addition to a theory, it aims to provide guidance for future empirical work on spillovers, neighborhood and peer effects. I show for instance how the standard practice of not distinguishing between the mean of the logs and the log of the mean (the geometric and arithmetic averages) can bias both econometric estimates and policy conclusions.

One important case where the intertemporal trade-off does occur is education finance reform. Equalizing school budgets at the state level amounts to integrating the tax bases of segregated communities, or mixing the populations themselves when the only local interactions are fiscal. The distortions from constraining heterogeneous families to uniform funding initially reduce growth, but in the long run the gains from homogenization dominate. These arise because credit market imperfections do not allow poor communities and families to borrow from rich ones, even though their return to human capital investment is higher. Generally, a move to state or national financing (I make no distinction) involves more than a simple redistribution of educational monies; it also affects the total resources which parents and voters choose to invest in human capital. Whereas most of the literature treats voters as myopic, I solve for the time-consistent choices of dynastic families. The results tend to further strengthen the case for national funding, which allows voters to better internalize the intertemporal spillovers from mobility across communities and economy-wide complementarities in production.

This paper draws most directly on Bénabou (1993), Robert Tamura (1991), and Gerhard Glomm and B. Ravikumar (1992). Bénabou (1993) shows that individuals' incentives to segregate in response to local spillovers in education can have severe general equilibrium effects on productivity and welfare, even with perfect capital markets. But that paper's representative agent, steady-state model does not address issues of inequality or dynamics. Tamura shows how heterogeneity among agents who face decreasing returns in human capital accumulation but are linked by an economy-wide spillover slows down aggregate growth. This effect is only transitory, as his deterministic economy converges to a homogeneous outcome. Glomm and Ravikumar show that long-run growth is higher when education is privately purchased than when it is publicly funded, as a private system provides better incentives to invest in human capital. I show that just taking into account the randomness in children's ability, or any other shock to human wealth, will generally *reverse* this ranking. The paper thus makes clear the importance of analyzing stratification and education finance in a dynamic and stochastic framework. That emphasis is shared with Steven Durlauf (1996) and Suzanne Cooper (1992), whose papers are closest in spirit to this one. Durlauf demonstrates how endogenous community formation and local funding of education can generate path dependence in dynastic income, including poverty traps. Cooper incorporates redistribution into his model, by allowing communities linked through production externalities to vote on cross subsidies for education. The political economy of local school finance is examined by Raquel Fernandez and Richard Rogerson (1996), who compare the effects of different reforms on communities' composition and education spending. The political economy of state public education and some of its consequences for inequality and growth are studied by Glenn Loury (1981), Roberto Perotti (1993) and Gilles Saint-Paul and Thierry Verdier (1993).

Section I develops a general dynamic model with local and global interactions. Section II compares short- and long-run growth in a segregated and in an integrated economy. Section III analyzes individual welfare and discusses the model's implications for empirical work on spillovers. Section IV applies the general analysis to the comparison between nationally funded public education, locally funded public education, and private education. Section V concludes. All proofs are gathered in Appendix B.

I. Local and Global Interactions in Human Capital

To build up intuition for the general model I begin this section with some concrete examples. The first one centers on community funding of schools and imperfect substitutability in production; it will be extended later on to analyze education finance policy. The others, taken mostly from the literature, involve peer or neighborhood effects and economy-wide knowledge spillovers. Observing that these problems all share the same underlying structure (reduced form of the equilibrium dynamics), I then present the core model in Section I.C.

A. *A First Model: Education and Production*

There is a continuum of overlapping-generation families $i \in \Omega$, of unit measure. During each period adults work, consume, and spend time rearing their single child. At time zero, the adult member of dynasty i faces the following problem:

$$\text{maximize} \quad U_0^i \equiv E_0\left[\sum_{t=0}^{\infty} \rho^t \ln c_t^i\right],$$

subject to:

$$(1) \qquad c_t^i = (1 - \tau_t^i) y_t^i$$

$$(2) \qquad y_t^i = \nu_t^i w_t^i$$

$$(3) \qquad h_{t+1}^i = \kappa \xi_t^i ((1 - \nu_t^i) h_t^i)^\delta (E_t^i)^{1-\delta}$$

and h_0^i given. There are no financial assets, only human wealth. At time t, adult i has human capital h_t^i. He spends a fraction ν_t^i of his unit time endowment at work, earning the hourly wage w_t^i, and devotes the rest to helping his child learn. The term h_t^i in (3) could also be due in part to inherited ability; the unpredictable component of the child's innate talent is represented by the independently and identically distributed (i.i.d.) shock ξ_t^i, with $E[\xi_t^i] = 1$. The other input in the production of human capital is formal schooling, which is financed by taxing the labor income of local residents. Per-capita expenditures are therefore

$$(4) \qquad E_t^i = \tau_t^i Y_t^i \equiv \tau_t^i \int_0^\infty y \, dm_t^i(y)$$

where m_t^i is the distribution of income and Y_t^i its average, in the community Ω_t^i to which family i belongs at time t: city or suburb, state, and so on. In addition to primary and secondary education, E_t^i could also include other locally provided, skill-enhancing public goods, such as libraries and public safety.

Consider now the production sector. Final output is produced by competitive firms with the technology $Y_t = (\int_0^\infty (x_t^s)^{(\sigma-1)/\sigma} ds)^{\sigma/(\sigma-1)}$, where x_t^s denotes intermediate input s and $\sigma > 1$. I assume that a worker must specialize in a single input; given downward-sloping demand curves $p_t^s = (x_t^s/Y_t)^{-1/\sigma}$, each individual then chooses a different task, $s(i) = i$. A worker with human capital h_t^i, working $\nu_t^i \leq 1$ hours, produces $x_t^i = \nu_t^i h_t^i$ and earns $y_t^i = p_t^i x_t^i = (\nu_t^i h_t^i)^{(\sigma-1)/\sigma}(Y_t)^{1/\sigma}$. Equivalently, $y_t^i = h_t^i \partial Y_t / \partial h_t^i = \nu_t^i w_t^i$, where $w_t^i \equiv p_t^i h_t^i = \partial Y_t / \partial \nu_t^i$ is his hourly wage. I shall now state a result which is intuitive, given logarithmic utility, but whose proof is sufficiently complex that it will be deferred until Section IV (Proposition 6).

LEMMA: *The optimal fraction of time ν_t^i spent working and the tax rate τ_t^i unanimously chosen under any voting system are both time invariant, and independent of community composition.*

The equilibrium values of ν and τ will be derived later on, but for present purposes one can simply treat them as fixed parameters.[2]

With $\nu_t^i = \nu$ for all i, aggregate output simplifies to

$$(5) \qquad Y_t = \nu \left(\int_0^\infty h^{(\sigma-1)/\sigma} \, d\mu_t(h) \right)^{\sigma/(\sigma-1)} \equiv \nu H_t,$$

where μ_t denotes the distribution of human capital in the whole labor force Ω. Similarly, each worker's earnings depend positively on a simple *economy-wide* index of human capital:

$$(6) \qquad y_t^i = \nu w_t^i = \nu (H_t)^{1/\sigma} (h_t^i)^{(\sigma-1)/\sigma}.$$

This complementarity captures the idea that poorly educated, insufficiently skilled production or clerical workers will drag down the productivity of engineers, managers, doctors, and so on. Conversely, lagging advances in knowledge by scientists, engineers and other professionals will mean lagging wages for basic workers. That there is *some* degree of interdependence seems quite plausible; the model requires σ to be finite, but it can be very large.[3] This interdependence is also reflected in the per-capita income of each community:

$$(7) \qquad Y_t^i = \int_0^\infty y \, dm_t^i(y) = \nu (H_t)^{1/\sigma} \left(\int_0^\infty h^{(\sigma-1)/\sigma} \, d\mu_t^i(h) \right) \equiv \nu (H_t)^{1/\sigma} (L_t^i)^{(\sigma-1)/\sigma},$$

where μ_t^i denotes the distribution of human capital in community Ω_t^i. Finally, incorporating

[2] Log utility also leads to a constant investment rate in Tamura (1991) and to constant tax rates in Glomm and Ravikumar (1992). Note that Proposition 6, hence also the lemma, assume that the shocks ξ_t^i and initial conditions h_0^i are log-normally distributed.

[3] Note also that monopolistic competition in the labor market is essentially a shortcut which allows us to abstract from occupational choice. Similar economy-wide linkages arise in a neoclassical model with competitive markets for complementary labor inputs, such as those of workers and managers; see Bénabou (1993). Tamura (1992) offers yet another model of specialization, which leads to expressions very similar to (5) and (6).

(4) and (7) into (3) yields the reduced form:

$$(8)\quad h^i_{t+1} = \Theta\xi^i_t(h^i_t)^\delta(L^i_t)^{(1-\delta)(\sigma-1)/\sigma} \times (H_t)^{(1-\delta)/\sigma}$$

where $\Theta = \kappa(1-\nu)^\delta(\nu\tau)^{1-\delta}$. Equation (8) involves both a *local linkage* L^i_t, because public goods are funded by community income, and a *global linkage* $H_t = Y_t/\nu$, because all workers are complementary in production. These raise the following questions. Is it more efficient, in the sense of increased total output, for the population to *stratify* into homogeneous jurisdictions ($L^i_t = h^i_t$) or to *mix* in representative, integrated districts and towns ($L^i_t = H_t$)? Does the answer depend on whether one takes a short or a long-run perspective? Can integration be Pareto-improving without compensating transfers to richer families?

One one hand, parental background h^i_t and school or community quality L^i_t are complements, suggesting that assortative matching of families is efficient. On the other, the complementarity and symmetry of individual inputs in (5) imply that worker heterogeneity lowers productivity; this argues in favor of integration and its homogenizing effect. Which way the balance goes in this model is worked out in Section IV, which is specifically devoted to school finance. Indeed, when the only local externality is fiscal, integrating communities or integrating their tax bases through a shift to state funding is essentially equivalent. But the set of issues raised above clearly extends beyond pecuniary spillovers and the functional forms embodied in (8). It is therefore important to relax these assumptions and broaden the study to a wider class of models.

B. *Other Models Raising the Same Issues*

There are many potential channels through which local and economy-wide interactions in human capital can arise, leading to a structure similar to (8) and the same questions regarding stratification. Sociologists have long described, and economists recently modeled, several mechanisms through which a community's makeup affects the educational outcomes of its young people. These sources of "social capital," in the terminology of Glenn Loury (1977) and James Coleman (1990), include: peer effects in school between students with different backgrounds or abilities (Abhijit Banerjee and Timothy Besley, 1990); the role models and networking contacts provided by neighboring adults, whether beneficial or contributing to a "culture of poverty" (William Julius Wilson, 1987; Peter Streufert, 1991; James Montgomery, 1991); and crime or other activities which interfere with education. Formally, $h^i_{t+1} = F(h^i_t, L^i_t)$, where L^i_t is some socioeconomic characteristic of the relevant community: school, neighborhood, ethnic group (George Borjas, 1992), and so on. There is also substantial empirical support for such peer or local interactions. Jencks and Mayer (1990) survey most of the evidence, while Bénabou (1993) and Durlauf (1996) discuss more recent studies. But perhaps most eloquent is parents' constant preoccupation with the quality of their children's classmates, as in the contentious debate over tracking. The following model provides a starting point for thinking about this last issue. In each generation, the young choose studying effort to maximize

$$(9)\quad u^i_{t+1} \equiv \max_e\{u(h^i_{t+1}) - e \mid h^i_{t+1} = ef(\xi^i_t, h^i_t, L^i_t)\}.$$

In addition to his own talent and family background, a child's return to studying is affected by the quality of his classmates, measured by some index L^i_t specified below. When adult, he consumes all his income, equal to his human capital. With constant relative risk aversion $1/a$, optimal effort rises or falls with the other inputs, depending on whether the income or substitution effect dominates: $e^i_{t+1} = f(\xi^i_t, h^i_t, L^i_t)^{a-1}$. Maximized utility is solely a function of human wealth, $u^i_{t+1} = ((h^i_{t+1})^{(a-1)/a} - a)/(a-1)$, and intergenerational dynamics take the simple form $h^i_{t+1} = f(\xi^i_t, h^i_t, L^i_t)^a \equiv F(\xi^i_t, h^i_t, L^i_t)$. Tracking and untracking correspond to different compositions of the group over which L^i_t is computed. Let this index reflect both the innate abilities and the socioeconomic backgrounds of child

(i, t)'s set of classmates, denoted Ω_t^i:

$$(10) \qquad L_t^i = \int_{\Omega_t^i} (\xi_t^j)^b (h_t^j)^{1-b}\, dj.$$

When students are randomly matched, $\Omega_t^i = \Omega$ and $L_t^i = E[\xi^b]E_t[h^{1-b}] \equiv L_t$ for all i. Tracking, on the other hand, consists of sorting them according to their performance on some test: $\Omega_t^i = \{j \in \Omega \mid \pi_t^j = \pi_t^i\}$. For simplicity, let this test yield an exact measure of a student's value to his peers: $\pi_t^i = (\xi_t^i)^b (h_t^i)^{1-b}$, so $L_t^i = \pi_t^i$.[4] In the law of motion $h_{t+1}^i = F(\xi_t^i, h_t^i, L_t^i)$, the third input is now $L_t^i = (\xi_t^i)^b(h_t^i)^{1-b}$ instead of $L_t^i = L_t$ previously. Tracking magnifies the effect of talent and, most importantly, the persistence of human wealth disparities across families. These effects are more important, the larger the intertemporal elasticity of substitution a.

Turning now to the aggregate level, knowledge spillovers à la Robert Lucas Jr. (1988) may affect everyone's productivity: $y_t^i = \Theta(h_t^i)^a(H_t)^b$, or more generally $y_t^i = F(h_t^i, H_t)$, where H_t is some economy-wide index of human capital. As long as schooling and R&D require produced resources, h_{t+1}^i will then be affected by H_t. Alternatively, Tamura (1991) assumes that the aggregate level of knowledge directly affects the creation of new human capital: $y_t^i = h_t^i$ but $h_{t+1}^i = \Theta(h_t^i)^a(H_t)^b$.[5] This same reduced form also arises in Glomm and Ravikumar's (1992) model of public education. Indeed, aggregate resources matter if the accumulation of knowledge uses (directly or through its inputs from the production sector) any kind of nationally provided public good: funding for R&D, infrastructure, defense, public safety, and so on.

[4] More generally, test scores depend on some combination of own ability and parental education, $\pi_t^i = (\xi_t^i)^c(h_t^i)^{1-c}$. Integrating L_t^i over Ω_t^i when $b \neq c$ is quite complicated, but it can be done under assumptions of lognormality, with qualitatively similar results.

[5] Some of these knowledge spillovers could also be localized, if only because geographical distance limits frequency of interaction. Lucas (1988) points to cities as the very manifestation of local externalities. Edward Glaeser et al. (1992) find evidence of external increasing returns at the city level, and James Rauch (1993) of human-capital spillovers.

C. *A General Framework*

The previous examples all share the same reduced form. In addition to own ability and family background, a child's education is affected by the local and economy-wide distributions of human capital, through a variety of fiscal, technological, and social spillovers: $h_{t+1}^i = \mathcal{F}(\xi_t^i, h_t^i, \mu_t^i, \mu_t)$. The problem is to find functional forms which make these dynamics tractable, yet yield general insights into the economic forces at work. For instance, when schools or communities are integrated the rich lose and the poor benefit. The net loss or gain must be weighed against the social value, or cost, of a more homogeneous workforce in the next generation. This raises in turn the question: *how costly or beneficial is heterogeneity at the local and aggregate levels*? Do the stronger students in a classroom pull everyone up, or do the weaker ones drag everyone down? How effectively can highly educated managers make up for poorly literate workers? Intuitively, the lower (upper) tail of the distribution of human capital is more important in shaping the outcome, the greater the *complementarity* (substitutability) between individual contributions. Most issues of interest can therefore be captured with the following simple structure:

$$(11) \qquad h_{t+1}^i = \Theta\xi_t^i(h_t^i)^\alpha(L_t^i)^\beta(H_t)^\gamma$$

$$(12) \qquad L_t^i = \left(\int_0^\infty h^{(\varepsilon-1)/\varepsilon}\, d\mu_t^i(h)\right)^{\varepsilon/(\varepsilon-1)}$$

$$(13) \qquad H_t = \left(\int_0^\infty h^{(\sigma-1)/\sigma}\, d\mu_t(h)\right)^{\sigma/(\sigma-1)}$$

$$(14) \qquad \ln h_0^i \sim \mathcal{N}(m, \Delta^2) \qquad \text{and} \qquad \ln \xi_t^i \sim \mathcal{N}(-s^2/2, s^2).$$

The acquisition of human capital reflects the influence of family, community and economy-wide factors, with respective weights α, β, and γ. Individual shocks are normalized so that $E[\xi_t^i] = 1$. Human wealth is initially distributed lognormally across agents, making the losses from heterogeneity at each level exactly

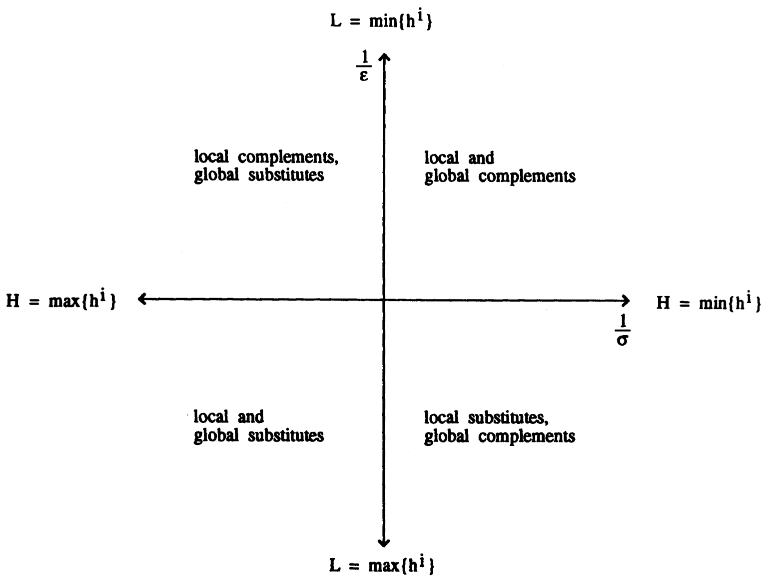

FIGURE 1. THE COSTS OF HETEROGENEITY: LOCAL AND GLOBAL DEGREES OF COMPLEMENTARITY $1/\varepsilon$ AND $1/\sigma$

proportional to the degree of complementarity: $H_0 = E[(h_0^i)^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)} = e^{-\Delta^2/2\sigma}E[h_0^i]$ and $L_0 = e^{-\Delta^2/2\varepsilon}E[h_0^i]$. All these properties will be shown to remain true over time. Of critical importance, naturally, are the overall *costs of heterogeneity* in economy-wide and community-specific interactions, $1/\sigma$ and $1/\varepsilon$. As shown on Figure 1, I allow them to take any values, positive or negative. As $1/\sigma$ decreases from $+\infty$ to $-\infty$, H_t spans the whole range of technologies from Leontieff or "weakest link," $H_t = \min\{h_t^i, i \in \Omega\}$, to "best shot," $H_t = \max\{h_t^i, i \in \Omega\}$, as in Kevin Murphy et al. (1991) where the best innovation is incorporated into the next generation's know-how. Similarly at the local level, I allow all cases from peer effects where "one bad apple spoils the bunch" to role models where the best individual sets the standard.[6]

[6] There is an interesting parallel here with Anthony Atkinson's (1970) index of inequality aversion in a social-welfare function. But the underlying economic motivation is quite different, allowing in particular for "inequality-loving" interactions.

While the exposition in the paper will focus on (11)–(14), the analysis extends beyond these functional and distributional assumptions. In the Appendix I incorporate multiple spillovers,

$$(15)\quad h_{t+1}^i = F(\xi_t^i, h_t^i, L_{1,t}^i, \dots, L_{K,t}^i, H_{1,t}, \dots, H_{N,t}),$$

and only require F to be homogeneous of arbitrary degree in $(h, L_1, \dots, L_K, H_1, \dots, H_N)$. The results show that L_t^i and H_t in (12)–(13) can legitimately be viewed as *composites* of all local and global interactions. Thus $1/\sigma$ reflects the relative importance of say, complementarity in the production of goods, which makes heterogeneity costly ($1/\sigma_1 > 0$), and substitutability in the creation of nonrival, nonexcludable ideas, where inequality is beneficial ($1/\sigma_2 < 0$). Similarly, $1/\varepsilon$ incorporates the contributions to education of school funding ($1/\varepsilon_1 = 1/\sigma_1$), role models (say, $1/\varepsilon_2 < 0$), and so forth. The general model also reveals how the interplay between parental background and community inputs contributes to the efficient social structure. Clearly, the complementarity inherent in the standard Cobb-Douglas specification (11) tends to favor segregation. Making F a constant elasticity of substitution function (CES) with elasticity λ, say, shows how the long-run effects of stratification depend on the relative values of the elasticities ε, σ and λ operating within and between inputs to education.

Finally, it is worth stressing that (15) is generally the *reduced form* of a fully specified, choice-theoretic model, and as such embodies all appropriate optimality and equilibrium conditions.[7] In particular, any effects of community composition on agents' decisions are already captured by the dependence of F on the L_k^i's; one can thus study the consequences of stratification without being subject to the Lucas critique.

II. Stratification and Growth: The Short and the Long Run

Given the model described by (11)–(14), I shall compare human capital accumulation and welfare under two polar regimes. Under perfect segregation adults are sorted into completely homogeneous communities ($L_t^i = h_t^i$), so the local environment compounds family differences:

$$h_{t+1}^i = \Theta \xi_t^i (h_t^i)^{\alpha+\beta} (H_t)^{\gamma}. \tag{16}$$

Under perfect integration each community is a representative sample of the population at large ($L_t^i = \hat{L}_t$), so everyone shares in the same level of local externality or public good. Denoting all variables with a hat,

$$\hat{h}_{t+1}^i = \Theta \xi_t^i (\hat{h}_t^i)^{\alpha} (\hat{L}_t)^{\beta} (\hat{H}_t)^{\gamma}. \tag{17}$$

For simplicity, community composition will not be endogenized. This was done in Bénabou (1993) and Durlauf (1996), with segregation emerging as the equilibrium outcome. The same forces are at work here: (a) with $\partial^2 F / \partial h \partial L > 0$, better educated parents tend to outbid less educated ones for space in a better community; (b) in the absence of significant fixed costs, the rich have no desire to let in the poor and will vote for zoning or income requirements which help keep them out.[8] The other reason for treating group composition as exogenous is that the mixing and sorting regimes correspond to alternative policies: local versus state funding of schools (tax base integration), historically segregated schools districts versus busing or magnet schools, ability tracking versus comprehensive classrooms, low-income public housing projects versus subsidies to mixed income developments. Affirmative action in the workplace is yet another potential application.[9]

A. *Dynamics and Losses from Heterogeneity*

The dynamic path of the economy under each regime will now be derived. Conveniently, the distribution of human capital always remains lognormal. To see this, suppose that in a segregated economy $\ln h_t^i \sim \mathcal{N}(m_t, \Delta_t^2)$. Then $\ln H_t = m_t + \Delta_t^2(\sigma - 1)/2\sigma$, and (16) implies

$$\begin{cases} m_{t+1} = \theta - s^2/2 \\ \qquad\qquad + R m_t + \gamma\left(\dfrac{\sigma - 1}{\sigma}\right)\dfrac{\Delta_t^2}{2} \\ \Delta_{t+1}^2 = (\alpha + \beta)^2 \Delta_t^2 + s^2 \end{cases} \tag{18}$$

[7] In the model of education finance of Section I.A, the equilibrium time allocation, wages, and tax rate are all reflected in the coefficient Θ and the exponents α, β, γ of equation (8). Similarly, the transition equation derived for the tracking model of Section I.B captures students' optimal responses to their classroom environment. Finally, the intertemporal equilibrium models of Tamura (1991, 1992), Borjas (1992) and Glomm and Ravikumar (1992) also have (11)–(13) as a reduced form. The value of reduced form models of intergenerational transmission was persuasively demonstrated by Arthur Goldberger (1989).

[8] Alternatively, in Fernandez and Rogerson (1996) sorting occurs through different local tax rates. Bénabou (1996) provides a general analysis of the conditions on technologies, preferences and capital markets which lead to stratification.

[9] Note also that the accumulation factor Θ is kept constant across the two regimes. Equating growth rates when agents are identical better highlights the effects of heterogeneity within and across communities. If optimal investment decisions vary across regimes it is easy to allow for distinct Θ and $\hat{\Theta}$, as done in Section IV when comparing different systems of education finance. Similarly, it is immediate to accommodate the case where the exponent on the shocks ξ_t^i differs across regimes, as in the tracking model of Section I.B when $b \neq 0$.

where $\theta \equiv \ln \Theta$ and $R \equiv \alpha + \beta + \gamma$. Similarly in an integrated economy, if $\ln \hat{h}_t^i \sim \mathcal{N}(\hat{m}_t, \hat{\Delta}_t^2)$ then $\ln \hat{L}_t = \hat{m}_t + \hat{\Delta}_t^2(\varepsilon - 1)/2\varepsilon$ and

$$(19)\quad \begin{cases} \hat{m}_{t+1} = \theta - s^2/2 + R\hat{m}_t \\ \quad + \left(\gamma \dfrac{\sigma - 1}{\sigma} + \beta \dfrac{\varepsilon - 1}{\varepsilon}\right) \dfrac{\hat{\Delta}_t^2}{2} \\ \hat{\Delta}_{t+1}^2 = \alpha^2 \hat{\Delta}_t^2 + s^2. \end{cases}$$

Integrating communities changes the effective technology of accumulation. Comparing the two dynamical systems shows that this has a dual effect on the economy. First, it alters the impact of any *given amount* of heterogeneity on the growth rate; more on this below. Second, it *accelerates convergence* towards a (more) homogeneous society: $\hat{\Delta}_t^2 < \Delta_t^2$ in each period, and when $s^2 > 0$ this remains true in the limit:

$$(20)\quad \hat{\Delta}_\infty^2 \equiv \frac{s^2}{1 - \alpha^2} < \frac{s^2}{1 - (\alpha + \beta)^2} \equiv \Delta_\infty^2.$$

The interplay between these instantaneous and dynamic effects is what creates the potential for different rankings of the two economies in the short and long run. Whether or not this trade-off occurs, the combined impact of heterogeneity and social structure can be brought out most clearly by a simple change of variables. In tracking aggregate growth I shall focus on per-capita human wealth $A_t \equiv \int_0^\infty h \, d\mu_t(h)$, rather than $m_t = \ln A_t - \Delta_t^2/2$ or some other index such as H_t which is not invariant to mean-preserving spreads in the distribution of human capital.[10] Under stratification, (16) implies $A_{t+1} = \Theta(\int_0^\infty h^{\alpha+\beta} \, d\mu_t(h)) H_t^\gamma$, which (18) allows me to write as

$$(21)\quad \ln\left(\frac{A_{t+1}}{A_t}\right) = \theta + (R - 1)\ln A_t - \left((\alpha + \beta)(1 - \alpha - \beta) + \frac{\gamma}{\sigma}\right) \frac{\Delta_t^2}{2}.$$

The first two terms give the growth rate of a standard representative agent economy. When levels of knowledge are unequal, however, $\int_0^\infty h^{\alpha+\beta} \, d\mu_t(h) \neq A_t^{\alpha+\beta}$ and $H_t^\gamma \neq A_t^\gamma$ due to Jensen's inequality; these differences are reflected in the last term of (21). In fact, the drag on growth from heterogeneity is simply the product of the current variance Δ_t^2 with a constant term measuring the economy's *efficiency loss per unit of dispersion*:

$$(22)\quad \mathcal{L} \equiv (\alpha + \beta)(1 - \alpha - \beta) + \gamma/\sigma.$$

The intuition is straightforward: losses reflect the concavity of the function $h^{\alpha+\beta}$ and the complementarity $1/\sigma$ of agents' inputs in the economy-wide aggregate H, which has weight γ. Conversely, heterogeneity is a source of gains when communities face increasing returns or agents are substitutes. For the integrated economy, similar derivations lead to

$$(23)\quad \ln\left(\frac{\hat{A}_{t+1}}{\hat{A}_t}\right) = \theta + (R - 1)\ln \hat{A}_t - \left(\alpha(1 - \alpha) + \frac{\beta}{\varepsilon} + \frac{\gamma}{\sigma}\right) \frac{\hat{\Delta}_t^2}{2}$$

so that the reduction in growth per unit of variance is now

$$(24)\quad \hat{\mathcal{L}} \equiv \alpha(1 - \alpha) + \beta/\varepsilon + \gamma/\sigma.$$

The interpretation is the same, except that it is now family rather than community returns to scale which matter, and that the interaction of heterogeneous individuals at the local level results in an additional loss β/ε. While the terms in $\mathcal{L}$ and $\hat{\mathcal{L}}$ will often be described as losses, it should be kept in mind that the model allows for any configuration of $(\alpha, \beta, \gamma, \sigma, \varepsilon)$. It does not impose that inequality be bad for growth, even though this is a consistent finding of empirical studies (for example, Alberto Alesina

[10] It is immediate to go from the arithmetic average A_t to any other one: $H_t = A_t e^{-\Delta_t^2/2\sigma}$, $L_t = A_t e^{-\Delta_t^2/2\varepsilon}$, and so on, but only A_t is neutral toward segregation or integration: as $\Delta_t^2 > \hat{\Delta}_t^2$, $\hat{A}_t = A_t$ implies $\hat{H}_t > H_t$ for $1/\sigma > 0$, and vice-versa for $1/\sigma < 0$. Also, while H_t corresponds to output in the model of Section I.A, such is not the case in general. For instance, $Y_t = A_t \neq H_t$ in Tamura (1991) and similarly $Y_t = A_t \neq L_t$ in my tracking model; see Section I.B.

and Dani Rodrik, 1994; Torsten Persson and Guido Tabellini, 1994).

B. *The Short Run*

Having identified the instantaneous and dynamic effects of social structure on growth, I now examine the potential trade-off between the two. I first ask which economy grows faster, for a given distribution of human capital. In other words, suppose that at $t = 0$ previously integrated populations become segregated, say through the exodus of upper and middle class families from the cities to the suburbs. Will human capital at $t = 1$ be higher or lower?

PROPOSITION 1: *The mixed economy has higher growth in the short run, that is, for any given amount of heterogeneity, if and only if* $\mathcal{L} > \hat{\mathcal{L}}$, *or*

$$(25) \quad \phi \equiv \beta(1 - 2\alpha - \beta - 1/\varepsilon) > 0.$$

When $2\alpha + \beta < 1$ the education production function is less concave in the previous generation's human capital under integration than under segregation; this tends to make $\phi = \mathcal{L} - \hat{\mathcal{L}}$ positive.[11] Under constant returns, however, the opposite case seems more plausible, as $2\alpha + \beta > 1$ is equivalent to $\alpha > \gamma$. Thus if parental background matters more to a child's education than economy-wide factors, *sorting* tends to improve accumulation in the short run.[12] Local complementarity $1/\varepsilon > 0$ will strengthen this result, as mixing causes a "leveling down" of school or community quality to $L_0 < A_0$; the reverse holds when $1/\varepsilon < 0$. Moving beyond specific functional forms, three general forces determine whether the gains which integration confers on the poor are sufficient to offset the losses which it inflicts on the rich. To make them apparent, rewrite ϕ as:

$$(26) \quad \phi \equiv -2\alpha\beta + \beta(1 - \beta) - \beta/\varepsilon$$
$$= -2F_{12}/F - F_{22}/F - F_2/(F\varepsilon),$$

where all functions are evaluated at $(h, L, H) = (1, 1, 1)$. When $F_{12} > 0$ children from well educated families lose more from a marginal decline in community quality L than their poorer counterparts gain from a corresponding increase. This tends to make sorting efficient, as in Gary Becker's (1981) theory of marriages; the converse is true when parental and local inputs are substitutes, as allowed in the Appendix. When $F_{22} < 0$ a given improvement in L has a higher marginal product in a poor community, where local inputs are initially low, than in a rich one where they are already high; this argues in favor of integration. Finally, the term in $1/\varepsilon$ measures whether poorly educated individuals drag down the quality of a mixed community more or less than well educated ones pull it up.

C. *The Long Run*

Because of mixing's homogenizing effect, the drag on growth due to dispersion eventually becomes smaller in an integrated economy. Is this sufficient for the latter to make up its initial handicap and overtake the segregated economy? The systems of difference equations in (Δ_t^2, A_t) and $(\hat{\Delta}_t^2, \hat{A}_t)$ are solved in the Appendix, for any values of α, β, γ. I only report here the asymptotic results for $\alpha + \beta < 1$ and $R \leq 1$, thus abstracting from explosive growth or inequality. Two cases must be distinguished: a deterministic economy, where the effect of social structure on the growth rate is only transitory, and a stochastic one, where it can be permanent.

PROPOSITION 2: *(The effect of initial inequality.) Let* $s^2 = 0$. *Under constant returns,*

[11] Such is clearly the case if the local spillover operates through the arithmetic average ($1/\varepsilon = 0$), as in Glomm and Ravikumar (1992). This is why they find that growth in their model can be temporarily higher under public education than under private education if $2\alpha + \beta < 1$, that is, $\alpha(1 - \alpha) < (\alpha + \beta)(1 - \alpha - \beta)$. On the other hand if $\hat{L}_t$ is a geometric average ($1/\varepsilon = 1$) as in Borjas (1992), the mixed economy is more vulnerable to heterogeneity than the segregated one.

[12] The same is true for $\ln(\hat{H}_1/H_1) = \beta(\Delta^2/2)((1 - 2\alpha - \beta)(1 - 1/\sigma) + 1/\sigma - 1/\varepsilon)$, unless $1/\sigma$ is sufficiently large.

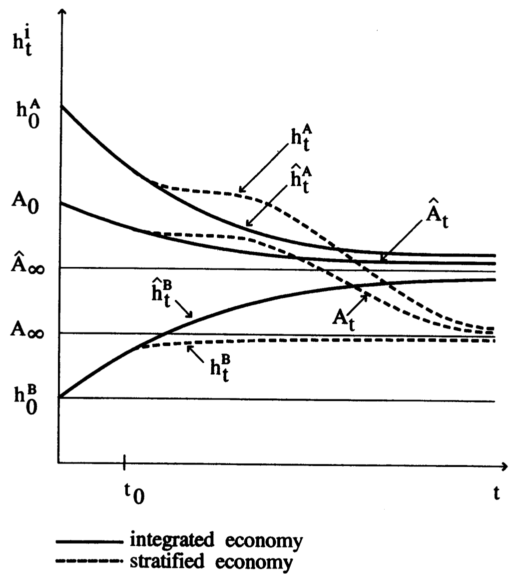

FIGURE 2. THE SHORT- AND LONG-RUN EFFECTS OF STRATIFICATION

the gap between the integrated and segregated economies converges to a finite limit: $\ln(\hat{A}_\infty/A_\infty) = \Phi\Delta^2/2$, *where*

$$(27)\quad \Phi \equiv \frac{(\alpha+\beta)(1-\alpha-\beta)+\gamma/\sigma}{1-(\alpha+\beta)^2} - \frac{\alpha(1-\alpha)+\beta/\varepsilon+\gamma/\sigma}{1-\alpha^2}.$$

More generally, for t large enough:

$$(28)\quad \ln\left(\frac{\hat{A}_t}{A_t}\right) \approx \left(\frac{\mathcal{L}}{R-(\alpha+\beta)^2} - \frac{\hat{\mathcal{L}}}{R-\alpha^2}\right)\frac{\Delta^2}{2}R^t.$$

The coefficient Φ is the long-run counterpart to ϕ, and embodies some of the main insights of the paper. The two numerators represent each economy's *instantaneous loss per unit of variance*, $\mathcal{L}$ and $\hat{\mathcal{L}}$. The two denominators reflect the different *speeds of convergence* towards a homogeneous society. Clearly, Φ decreases in β/ε and increases in γ/σ: there is a trade-off between incurring the costs of local heterogeneity and reducing at a faster rate the losses from aggregate heterogeneity. The other determinant factor is the difference in the concavity of the technologies facing a community and a family, adjusted by the appropriate convergence speeds. This makes Φ positive for $1/\varepsilon = 1/\sigma = 0$. With any $R \leq 1$ this remains true for the bracketed term in (28), which has the same general interpretation and properties as Φ.

Figure 2 illustrates the case $\phi < 0 < \Phi$ and $R = 1$, which is the most interesting. Since it is difficult to graph evolving distributions I focus on two "representative" dynasties, one rich and one poor: $i = A, B$.[13] All variables are detrended. Suppose that at time t_0 a previously integrated population becomes stratified ("flight to the suburbs"). Initially, the rich benefit by more than the poor lose: the distribution of income worsens while overall growth accelerates. Over time, growth slows down due to the fact that society remains more heterogeneous. Eventually, even the A's accumulation is dragged down, and all dynasties converge to a common level which is lower than if society had remained integrated. Plausible parameter values suggest that these effects can be quite large. For instance, let $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$, $1/\varepsilon = 1/\sigma = 0$ and $\Delta = 0.6$. The secession of the rich at first raises growth by $\phi\Delta^2/2 = 1.6$ percent, but eventually lowers the steady-state path of the economy by $\Phi\Delta^2/2 = 2$ percent. The initial boom is erased two generations later.

Stratification can easily have not just level effects, but permanent *growth-rate effects*. To see this, one need only follow the dictates of realism and take into account the randomness in children's talent or return to education, which induces social mobility and ensures a nondegenerate long-run income distribution.

[13] The same graph applies to an economy constituted of two homogeneous groups, A and B, with arbitrary proportions $(n, 1-n)$ and small initial dispersion $\Delta^2 \equiv n(1-n)\ln(h_0^A/h_0^B)$. See Appendix A and previous versions of this paper.

PROPOSITION 3: *(The effect of ongoing inequality.) If $s^2 > 0$ then for t large enough:*

$$(29) \quad \ln\left(\frac{\hat{A}_t}{A_t}\right) \approx \tfrac{1}{2}(\mathcal{L}\Delta_\infty^2 - \hat{\mathcal{L}}\hat{\Delta}_\infty^2)\left(\frac{1-R^t}{1-R}\right)$$

$$= \Phi\frac{s^2}{2}\left(\frac{1-R^t}{1-R}\right)$$

where Φ is the same as in (27). Under constant returns, $(1 - R^t)/(1 - R)$ becomes equal to t, and the integrated economy's long-run growth rate exceeds that of the segregated economy by $\Phi s^2/2$.

The more general message from this result is that by omitting idiosyncratic uncertainty one severely underestimates the long-term consequences of sorting arrangements like residential segregation, tracking, or decentralized education finance. Intuitively, recurrent sources of heterogeneity impact the economy one level higher than initial dispersion. When $R < 1$, s^2 affects long-run levels whereas the effect of Δ^2 vanishes asymptotically; Figure 2 still describes the expected trajectories of dynasties A and B. When $R = 1$, s^2 affects long-run growth in the very same way as Δ^2 affected long-run levels. In all cases the degree of stratification has a *permanent effect* on the economy. Different returns to scale essentially dampen or magnify the key trade-off between the costs of local and economy-wide heterogeneity, embodied in Φ.[14]

A general remark might be useful at this point: "long-run" results should not be viewed as relevant only to the distant future. In reality, many generations interact simultaneously. Children are influenced not just by the adults in their community, but also by older siblings and friends. Similarly, a manager will work with several cohorts of employees during his career. The overlapping-generations setup overstates the lags involved in the dynamics of human capital, and should therefore not be taken too literally.

D. *Local Complementarity, Global Complementarity, and Efficient Sorting*

Integrated communities and schools can induce a substantially better long-term outcome, even when causing aggregate losses in the short term. But of course this is not the only possible scenario. In particular, stratification remains preferable over long horizons if disparities in knowledge entail *sufficiently* greater losses at the community level (for example, in schooling) than at the aggregate level (for example, in production). Under constant returns, for instance:

$$(30) \quad \Phi \gtrless 0 \quad \text{as} \quad \frac{1}{\varepsilon} - \frac{1}{\sigma} \lessgtr \left(1 - \frac{1}{\sigma}\right) \times \left(\frac{1-\alpha}{1+\alpha+\beta}\right).$$

The boundary is illustrated on Figure 3. Together with its short-run counterpart $\phi = 0$ they divide the plane into four regions. When ϕ and Φ are both positive (respectively, negative) the optimal community structure is the same over all horizons: an induction argument shows that $\hat{A}_t > A_t$ (respectively, $\hat{A}_t < A_t$) for every $t \geq 1$. When $\phi\, \Phi < 0$ the paths of A_t and $\hat{A}_t$ cross, reflecting a genuine trade-off.[15] While there may be multiple intersections, the present value $\Sigma_{t=1}^{\infty}\, \rho^{t-1}\ln(\hat{A}_t/A_t)$ remains monotonic in ρ. The proper way to think about the *medium run* is therefore to consider the locus where this present value equals 0; as the discount factor rises from 0 to 1, this line pivots from the $\phi = 0$ to the $\Phi = 0$ boundary.

The triangular area between the two axes and the $\Phi = 0$ locus is especially noteworthy. Since $1/\sigma \leq 0 \leq 1/\varepsilon$, heterogeneity creates

[14] While Proposition 3 focuses on the arithmetic or per-capita average of human capital A_t, it readily extends to any other CES index with elasticity λ: one simply adds $(\Delta_\infty^2 - \hat{\Delta}_\infty^2)/2\lambda$ to the right-hand side of (29), which amounts to replacing Φ by $\Phi + [(1 - R)/\lambda][1/(1 - (\alpha + \beta)^2) - 1/(1 - \alpha^2)]$. When $R = 1$ this clearly makes no difference.

[15] The case $\phi > 0 > \Phi$ requires that heterogeneity be beneficial in local interactions, but even more so in global interactions: $\mathcal{L} < \hat{\mathcal{L}} < 0$; see (27). By contrast, $\phi < 0 < \Phi$ can occur with $\mathcal{L}$ and $\hat{\mathcal{L}}$ taking any sign.

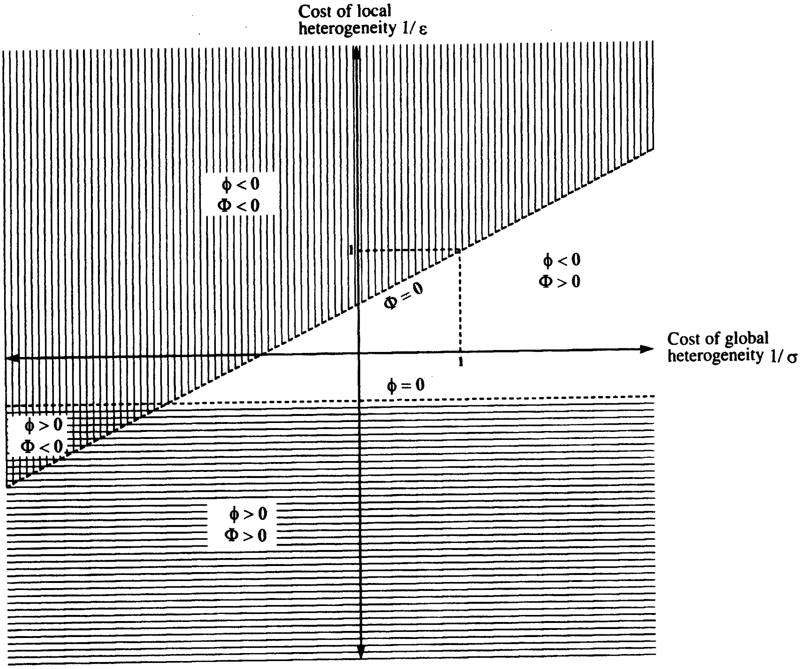

FIGURE 3. THE EFFICIENT SOCIAL STRUCTURE OVER SHORT AND LONG HORIZONS

negative spillovers in local interactions but positive ones at the aggregate level. Nonetheless, integrated communities lead to a superior long-run outcome, due to the effects of decreasing returns in accumulation which were mentioned earlier. To overcome this effect, $1/\varepsilon$ must be sufficiently larger than $1/\sigma$. A related result is that when heterogeneity entails similar costs at the local and economy-wide levels, $1/\varepsilon = 1/\sigma$, then $\Phi > 0$ provided $1 > 1/\sigma$. This is a plausible assumption: it means that parental background, community quality and economy-wide inputs are poorer substitutes in a child's education than individuals with different skills in the production of output or know-how.[16]

Figure 3 is perhaps best summarized by a simple story, which relates to the quotation given in the introduction: it will be efficient for the professional and working classes to live and be educated separately, only if good managers can make up for poorly qualified workers in the production process much more easily than students from favorable backgrounds can offset the effect of weaker schoolmates in peer interactions.

III. Welfare Analysis and Empirical Implications

A. *Discounting and Welfare*

As shown on Figure 2, each dynasty's distribution of human capital eventually converges—or asymptotes, with endogenous growth—to the economy's cross-sectional distribution. The latter always has a lower dispersion under integration than under segregation, and when $\Phi > 0$ it also has a higher mean. This suggests that integration could be *Pareto improving*, without need for compensating transfers, if people value the future sufficiently. More generally, one wants to assess

[16] When $F(h, L, H)$ is a CES with elasticity λ, the relevant comparison is $1/\lambda > 1/\sigma$; see (A6) in the Appendix, which also provides the analogue of (30) when $R < 1$.

how community structure affects a family's welfare, given its initial wealth h_0^i and discount factor $\rho < 1$. To derive simple, closed form expressions, I assume logarithmic utility in human capital.[17]

PROPOSITION 4: *Let U_0^i and $\hat{U}_0^i$ denote family i's expected present value of log-human capital under stratification and integration respectively. Then*

$$(31)\quad \hat{U}_0^i - U_0^i$$

$$= \frac{1}{2}\left(\Delta^2 + \frac{\rho s^2}{1-\rho}\right)\left(\frac{\rho}{1-\rho R}\right)$$

$$\times \left(\frac{\mathcal{L}}{1-\rho(\alpha+\beta)^2} - \frac{\hat{\mathcal{L}}}{1-\rho\alpha^2}\right)$$

$$+ \frac{1}{2}\left(\Delta^2 + \frac{\rho s^2}{1-\rho}\right)$$

$$\times \left(\frac{1}{1-\rho(\alpha+\beta)^2} - \frac{1}{1-\rho\alpha^2}\right)$$

$$+ \frac{\rho\beta(m - \ln(h_0^i))}{(1-\rho\alpha)(1-\rho(\alpha+\beta))}.$$

When $\Phi > 0$, the fraction of families who are better off under integration tends to one with the discount factor ρ. This may still be true when $\Phi < 0$.

The degree to which society is stratified affects a family's welfare through three channels: *efficiency, insurance, and redistribution*. First, integration lets all dynasties share in the aggregate—or complete markets—gains or losses computed earlier: the first component of (31) is simply the present value $\Sigma_{t=0}^{\infty} \rho^t \ln(\hat{A}_t/A_t)$. Note how the expression involving $\mathcal{L}$ and $\hat{\mathcal{L}}$ tends to ϕ when $\rho \to 0$ and to Φ when $\rho \to 1$. Equalizing the local input L_t^i across all children partially replaces the missing loans and insurance markets, by bringing closer to equality families' marginal rates of substitution (growth rates of human wealth) across dates and states. Thus in the second component of (31), equal to $\Sigma_{t=0}^{\infty} \rho^t(\Delta_t^2 - \hat{\Delta}_t^2)/2 > 0$, the term in s^2 is the insurance value of reducing the persistence of shocks by equalizing L_t^i; the term in Δ^2 is the gain, averaged over all agents (or computed ex ante), of smoothing out the impact of initial conditions on the consumption path. Finally, the difference between agent i's ex-ante and ex-post (knowing h_0^i) valuations of this smoothing is captured by the last component of (31): ignoring efficiency and insurance effects, dynasties which start above the mean lose from integration, while those which start below the mean gain.

As ρ increases towards 1 this redistributive effect becomes dominated by the other two.[18] That sum is generally positive because efficiency gains ($\Phi > 0$) are somewhat more likely than losses, as explained earlier, and because the insurance effect is always beneficial. Hence the second part of Proposition 4. In practice, the distribution of human wealth at any point in time has finite support. Therefore integration is indeed Pareto improving if individuals care enough about future generations.

B. *Implications for Empirical Research on Spillovers*

The *elasticities of substitution* in local and global interactions were shown to play a key role in the dynamics and welfare properties of a heterogeneous economy. To demonstrate concretely how important it is for empirical

[17] If income and consumption depend not only on own human capital but also on some index of aggregate productivity H_t, as in (6), this will understate the relative benefit of integration with respect to segregation when $1/\sigma > 0$. The bias is reversed when $1/\sigma < 0$. In either case, going from (31) to the expected present value of income or consumption is straightforward. Note finally that I shall require $\max\{\rho R, \rho(\alpha+\beta)^2\} < 1$ to ensure finite utility, but need not restrict R or $\alpha + \beta$ to be less than 1.

[18] Except in the case $s^2 = 0$ and $R < 1$, where the effect of integration is transitory, like that of initial conditions $\ln(h_0^i) - m$. Absent compensating transfers, the fraction of families with a net gain then remains bounded even as $\rho \to 1$.

work on spillovers to estimate these parameters—rather than constrain them, as usually done—I use Borjas' (1992) study of "ethnic capital." Using longitudinal data, he estimates the model:

$$\text{(32)} \quad \ln h_{t+1}^i = \alpha \ln h_t^i + \beta \ln L_t^i + \text{control variables} + u_t^i,$$

where h_{t+1}^i and h_t^i are a son's and his father's levels of human capital, measured by their hourly wages; L_t^i is "ethnic capital," defined as the *geometric* average of human wealth among adults in the father's ethnic group: $\ln L_t^i \equiv \int_0^\infty \ln h \, d\mu_t^i(h)$; and u_t^i is a random shock. Borjas estimates both α and β to be between 0.25 and 0.30 and statistically significant.[19] These results are very interesting in and of themselves, adding to the body of evidence that group interactions influence the acquisition of skills.[20]

One might hope that they could also shed light on the momentous issue of *ethnic segregation*, if only from a narrowly economic point of view: will a country be more productive if poorer minorities or immigrants share neighborhoods, schools, and other public goods with the richer majority population, or if they remain isolated in ethnic "ghettos"? Unfortunately, my results show that using a geometric average constraints the answer a priori: with $1/\varepsilon = 1$, $\phi < 0$ and $\Phi < 0$. Therefore total income is always higher under segregation. But suppose now that ethnic capital actually operates through the *arithmetic* average: $\ln L_t^i \equiv \ln(\int_0^\infty h \, d\mu_t^i(h))$. With $1/\varepsilon = 0$, $\phi \gtrless 0$ and $\Phi > 0$.[21] Integration is now more efficient, at least in the long run! The common practice of not distinguishing between the mean of the logs and the log of the mean can thus be very misleading. The point is more general, and extends beyond the issue of integration. Unless the particular aggregator which they impose happens to be the true one, standard "spillover" equations such as (32) are misspecified, and the presence of heterogeneity within each group can bias the estimates. One solution is to specify L_t^i as a CES index and estimate $1/\varepsilon$ by nonlinear methods. More simply, one could include in the linear regression not only the mean, but also the variance $(\Delta_t^i)^2$ of log-human capital in the individual's group. The ratio of their coefficients, times two, will provide an estimate of $(\varepsilon - 1)/\varepsilon$.

IV. Education Finance

Our general analysis has implications for the current debate on school funding. In over half the states, the inequities arising from the traditional reliance on property-tax revenues have resulted in constitutional challenges. While a variety of "remedies" have been proposed, those adopted in practice represent a significant move toward the equalization of expenditures per pupil—a situation which is the norm in most other countries. The first and best-known case is the 1976 ruling of the California Supreme Court (*Serrano* v. *Priest II*) declaring unconstitutional any formula which results in a positive correlation between districts' school expenditures and taxable wealth. Another instructive example is that of New Jersey. Following a 1973 Supreme Court decision, the state enacted an income tax to finance increased aid to poor school districts. But as communities were allowed to supplement these funds with property taxes, disparities in expenditures continued to grow. This lead to a new Supreme Court ruling in 1990 (*Abott* v. *Burke*) mandating "substantially equal" expenditures between the poorest and richest districts (*New York Times*, 1990). Most recently, Michigan has moved to replace its property-tax financing by an increase in the sales tax.

I therefore use the model developed in Section I.A, which involves only pecuniary spillovers, to compare three systems. Under national or *state funding*, all school budgets are equalized: $E_t^i = \tau_{t,N} Y_t$, and the tax rate $\tau_{t,N}$ is determined by a national vote. Under *local*

[19] Using years of education instead of log wages leads to similar results. The discussion below applies equally to that specification.

[20] One suspects, and Borjas' (1995) later work indeed tends to indicate, that "ethnic capital" really arises from neighborhood effects combined with ethnic segregation.

[21] Here and above I used (27) with $\gamma = 0$, as Borjas does not allow for aggregate spillovers. The inequalities remain unchanged when $|\gamma/\sigma| > 0$, unless it is large. The same results follow from (28) if u_t^i only represents measurement error.

funding, expenditures reflect community resources and locally chosen tax rates: $E_t^i = \tau_{t,L}^i Y_t^i$. I shall assume that income classes are perfectly segregated, reflecting the wide disparities in tax base observed in practice; thus $Y_t^i = y_t^i$. Finally, I also consider *private education*, as in Glomm and Ravikumar (1992); E_t^i is then replaced in (4) by $e_t^i = \tau_{t,P}^i y_t^i$, where $\tau_{t,P}^i$ represents the fraction of his income which adult i invests in his child's education.

In comparing these alternatives, it is useful to distinguish two questions. The first is whether a given amount of educational resources is best allocated equally or unequally across different families. The second is how the total money and time invested in human capital vary with the financing system, through the incentives facing parents and voters. I therefore first compare the three regimes assuming the same, constant investment rates τ and ν. I then endogenize τ_t^i and ν_t^i for each education system. Finally, I put together reallocation and aggregate investment effects.

A. *Reallocating Educational Funds: Stratification Effects*

Going back to the model of Section I.A, note that for given τ and ν the three systems of education finance differ only by their tax base: family, community, or state. With homogenous jurisdictions, moreover, local and private funding lead to the same outcome, $E_t^i = \tau Y_t^i = \tau y_t^i = e_t^i$, and both are equivalent to *segregation* in the reduced form (8): $L_t^i = h_t^i$. Conversely, state funding transforms the local fiscal spillover into an aggregate one, $E_t^i = \tau Y_t$, and is therefore equivalent to *integration: $L_t^i = H_t$*. When agents are identical all three systems result in the same growth rate $\Theta = \kappa(1-\nu)^{\delta}(\nu\tau)^{1-\delta}$, where δ and $1-\delta$ are the weights of parental background and school expenditures in a child's education. This "level playing field" will better highlight the effects of a pure *redistribution of educational funds* across segregated communities or individual families. Recall now that (8) is a special case of (11)–(14), with the following restrictions: (a) the local and global spillovers L_t^i and H_t are defined by the same elasticity $\varepsilon = \sigma > 1$, inherited from the production sector; (b) the weights of parental, local and economy-wide inputs to human capital are $\alpha = \delta$, $\beta = (1-\delta)(\sigma-1)/\sigma$ and $\gamma = (1-\delta)/\sigma$; they sum to $R = 1$. Moreover, output is $Y_t = \nu H_t$. Therefore, replacing in (25) and (27) gives Proposition 5.

PROPOSITION 5: *For given rates of resource and time investment in education, τ and ν:*

(1) In the short run, state funding of education leads to less human capital accumulation than local or private funding, since $\phi = -\delta(1-\delta)(1-1/\sigma^2) < 0$. Output responds similarly if and only if $1/\sigma < \delta/(1-\delta)$, as $\ln(\hat{Y}_1/Y_1) = \phi'\Delta^2/2$ with

$$\phi' \equiv -(1-\delta)\left(\frac{\delta\sigma-(1-\delta)}{\sigma}\right)\left(\frac{\sigma-1}{\sigma}\right)^2.$$

(2) In the long run, state funding is more efficient than local or private funding, since:

$$\Phi = \left(\frac{\sigma}{2\sigma+\delta-1}\right)\left(\frac{1-\delta}{1+\delta}\right)\left(\frac{\sigma-1}{\sigma}\right)^2 > 0.$$

When initial endowments are the only source of inequality ($s^2 = 0$), state funding raises the long-run levels of human capital A_t and output Y_t by $\Phi\Delta^2/2$. When children's ability or returns to education are uncertain ($s^2 > 0$), it raises the long-run growth rate of human capital and output by $\Phi s^2/2$.

The interpretation of this intertemporal trade-off is simple: focusing on output growth, $\phi'\Delta^2/2$ reflects the *distortions* from constraining everyone to the same school budget, and $\Phi s^2/2$ the long-run benefits from *homogenization*. The higher is the intergenerational discount factor, the more families would vote for a national system; this remains true when τ and ν are made endogenous in the next section.[22] The key assumption in Proposition 5 is

[22] It will also be confirmed that the tax rate chosen under local financing is independent of community composition, as claimed in the Lemma of Section I. This makes Proposition 5 applicable without any change to the comparison between integrated and segregated communities, *given* local funding.

indeed not the invariance of τ and ν across regimes. It is, as in Loury (1981), the absence of credit markets where poor communities or families could borrow from rich ones to finance education. State funding amounts to a partial redistribution of human wealth, with a later payback to society in the form of a higher H_t. Since each dynasty faces decreasing returns $(\alpha + \beta = 1 - (1 - \delta)/\sigma)$, it seems intuitive that this would increase efficiency. But, perhaps surprisingly, this is only true in the long run: early on, human capital accumulation is actually reduced, as rich families lose more than poor ones gain. This translates into lower output if production complementarities are not too strong. Also unexpected is the result that a national scheme increases the steady-state growth rate by a positive amount even when dynasties face returns arbitrarily close to one: as workers become perfect substitutes, $\lim_{\sigma \to \infty}(\Phi) = \frac{1}{2}(1 - \delta)/(1 + \delta) > 0$.

B. *The Supply of Educational Funds: Incentive Effects*

I shall now endogenize the allocation of time ν_t^i and the tax or savings rate τ_t^i chosen under each system of education finance. In contrast to many models where "myopic" voters care only about their consumption and their child's education, I solve for the subgame perfect equilibrium between dynastic agents.

PROPOSITION 6: *(1) Parents devote the same, constant fraction of their time to educating their children whether education is funded privately, locally or nationally:* $1 - \nu = \rho\delta$.

(2) Under each system, the tax or savings rate is constant over time. Define $\tau_* < \tau < \tau^*$ *as:*

$$\tau_* \equiv \frac{\rho(1 - \delta)(1 - 1/\sigma)}{1 - \rho\delta + \rho(1 - \delta)(1 - 1/\sigma)} < \tau$$

$$\equiv \frac{\rho(1 - \delta)(1 - 1/\sigma)}{1 - \rho\delta} < \tau^* \equiv \frac{\rho(1 - \delta)}{1 - \rho\delta}.$$

(a) *Under private funding, adults contribute a fraction* $\tau_P = \tau$ *of their income to their child's education.*

(b) *Under local funding, the tax rate unanimously chosen is* $\tau_L = \tau_*$ *if there is full intergenerational mobility across communities. It is* $\tau_L = \tau$ *if there is no mobility and voters internalize the effects of current expenditures on future community resources.*

(c) *Under state funding, the tax rate unanimously chosen is* $\tau_N = \tau^*$ *if voters internalize the effects of current expenditures on future productivity and aggregate resources. It is* $\tau_N = \tau_*$ *if they fail to do so.*

Voters in a public system choose the tax rate τ_* if they only take into account the *private* marginal value of human capital. The inequality $\tau_* < \tau$ reflects the fact that this value is higher in a private system, where an additional unit allows the adult not only to consume more but also to purchase more education for his offspring. This is similar to Glomm and Ravikumar's (1992) result that the young take less leisure and study harder if education is privately purchased. In my model, time is allocated between production and home instruction, both of which allow parents to better educate their offspring. Differences in the return to human capital are then reflected in τ rather than ν.

Rational voters, however, should take into account the benefits which their child derives from other members of his cohort's being more educated. If a community remains composed of the same families over time, a higher current tax rate bestows on each child not only more schooling, but also access to a larger tax base Y_{t+1}^i from which to finance future education; thus $\tau_L = \tau$, as with private funding. More likely, children move away and form new associations in each period. This reshuffling, which necessarily occurs when idiosyncratic shocks combine with stratification, leads to $\tau_L = \tau_*$. A national system allows voters to internalize the tax spillovers from mobility across communities, as well as those from production complementarities: $y_{t+1}^i = (h_{t+1}^i)^{(\sigma-1)/\sigma}(H_{t+1})^{1/\sigma}$. Self-interested but economically sophisticated voters will then choose $\tau_N = \tau^*$, which reflects the *social* return to education.

COROLLARY: *With rational, forward-looking voters, the representative agent economy's growth rate* $\Theta = \kappa(1 - \nu)^{\delta}(\nu\tau)^{1-\delta}$ *is higher* *under a state system than a private one:* $\Theta_N > \Theta_P$. *It is always lowest under local funding:* $\Theta_L \leq \min\{\Theta_P, \Theta_N\}$, *for any degree of public sophistication.*

Factors other than voter myopia could also affect these rankings. First, underinvestment by families or communities could be addressed by taxing consumption and subsidizing education, without going to *uniform* funding. I shall not develop this variant of the model here, but it is intuitive that a national subsidy scheme would also allow sophisticated voters to internalize the social value of human capital and thus equalize Θ_P and Θ_L with Θ_N, just as assumed in the previous section. Second, decentralized education systems may have desirable efficiency properties which also raise Θ_P or Θ_L relative to Θ_N: better monitoring of teachers and administrators, competition between schools or municipalities, and so on. Third, mandating equal school budgets could trigger a flight to private schools by wealthier families, leading in turn to decreased political support for public funding. It has been argued that this is what happened in California following the 1976 *Serrano* ruling: enrollment in private schools increased while public spending per student declined, relative to the national average (Lawrence Picus, 1991; Thomas Downes and David Shoeman, 1993). A model where public and private schooling exist side by side is beyond the scope of this paper, but a first approximation of this effect would be to decrease τ_N relative to τ_L, hence once again, Θ_N relative to Θ_L.

These last two arguments are not without problems. It is hard to argue that Japan or the many European countries which have a state funded, centralized education system are instructing their young less efficiently than the United States. International comparisons of student proficiency and education costs actually suggest the opposite. Neither have these countries seen any large-scale defection of well-to-do families towards private schools. In any case, it is easy to incorporate potential gains from subsidization, local control or competition into the analysis. Even though the model has $\Theta_N > \Theta_P > \Theta_L$ as the most natural case, the next section will allow for different rankings.

C. *Education Finance and Growth with Income Inequality*

The following proposition captures the essential elements of the debate over education finance, namely the relative importance of *incentive* and *reallocation* effects, and the relevant *time horizon*.

PROPOSITION 7: *Let* Θ *denote the trend growth rate reflecting the incentive effects of either private or local school funding in a representative agent economy, while* $\hat{\Theta}$ *similarly corresponds to state funding. Let* $\phi' \gtrless 0$ *and* $\Phi > 0$ *be the expenditure-equalization effects derived in Proposition 5. State finance of education may reduce output in the short run if* $\Theta - \hat{\Theta} > \phi' \Delta^2/2$, *but it leads to a higher growth rate in the long run if* $\Phi s^2/2 > \Theta - \hat{\Theta}$.

No restrictions are imposed here on the accumulation factors in the three regimes. If Θ_P, Θ_L, and Θ_N are the equilibrium values with rational voters derived in Proposition 6, long-run growth is *always* highest under state funding and lowest under local funding. If $\Theta_P \geq \Theta_N$ or $\Theta_L \geq \Theta_N$ due to voter myopia or any of the other factors discussed earlier, one still has the following corollary.

COROLLARY: *State funding maximizes the long-term growth rate, provided there is enough variation in innate ability or other idiosyncratic shocks.*

These propositions make clear the importance of analyzing education finance within a framework which is explicitly dynamic and stochastic. For instance, they reveal that Glomm and Ravikumar's (1992) finding that private education is superior in the long run arose not only from a particular model of time allocation, as seen earlier, but also from the omission of all uncertainty.[23]

[23] Their model's reduced form is $h^i_{t+1} = \kappa(h^i_t)^\alpha(e^i_t)^\beta$, with $e^i_t = \tau_P h^i_t$ under private education, $e^i_t = \tau_N \int_0^\infty h\, d\mu_t(h)$ under national education, and $\tau_P = \tau_N$ but $\kappa_P > \kappa_N$. This is a special case of (11)–(14), with no global interaction ($\gamma = 0$), perfect local substitutability ($1/\varepsilon = 0$) and *no shocks* ($s^2 = 0$). Note that for all $\alpha + \beta < 1$, $\Phi = 1/(1 + \alpha) - 1/(1 + \alpha + \beta) > 0$.

What about the robustness of this paper's results? Both the initial deadweight loss in output $\phi' \Delta^2/2$ and the long-run growth gain from homogenization $\Phi s^2/2$ embody the assumption of logarithmic utility, which makes education expenditures proportional to income (of the family, community or state) in each regime. This property made it possible to solve a forward-looking politico-economic model where heterogeneous families face different returns to educational investment. But one may ask whether it is empirically plausible, and how crucial it is to the conclusions. Theodore Bergstrom et al. (1982) provide the most extensive study of the demand for school expenditures. Their micro-based estimates of income elasticity are between 0.64 and 0.83. They also survey previous estimates, which range from 0.46 to 1.35, and conclude that a value of about $^2/_3$ seems the most plausible. A unit elasticity is thus not a bad approximation; in fact, that hypothesis cannot be rejected from Bergstrom et al.'s results. Moreover, long-run time-series data show no tendency for the share of income devoted to primary and secondary education to fall as the economy grows. In the United States the combined public and private share has remained stable from 1960 to 1990, at around 4 percent of GNP.

Suppose nonetheless that the true elasticity is some $\eta < 1$, so that school expenditures in equation (4) become $E_t^i = \tau(Y_t^i)^\eta$, for some constant τ.[24] The reduced form (8) remains the same, except that the weights of the local and global spillovers are both multiplied by η: $\beta = \eta(1 - \delta)(\sigma - 1)/\sigma$ and $\gamma = \eta(1 - \delta)/\sigma$. Applying the general results from Section II with $1/\varepsilon = 1/\sigma$, it is easy to show how this affects the path of output under each financing system: one can write $\ln \hat{Y}_1/Y_1 = \psi(\eta)\Delta^2/2$ and, for t large enough, $\ln \hat{Y}_t/Y_t \approx \Psi(\eta)(s^2/2)(1 - R^t)/(1 - R)$, with the following properties. First, $\psi(1)$ and $\Psi(1)$ are of course the values $\phi' \gtrless 0$ and $\Phi > 0$ computed in Proposition 5. Second, $\psi(\eta) \geq \eta\psi(1)$ and $\Psi(\eta) \geq \eta\Psi(1)$: the short-run costs of equalizing school budgets are reduced by at least a factor of η, while the long-run gains (normalized for returns to scale) decrease by no more than a factor of η.[25] The *main conclusion* is thus *unchanged*: the cumulated growth benefits from state funding rise over time to a positive limit, although this intertemporal profile is now a little flatter.

Naturally, certain issues remain beyond the scope of this study. One is how the outside option of private schooling affects the scope for a politically feasible redistribution of educational funds. Another is the argument that while one may need to supplement the education budgets of poor communities, it is inefficient to limit investment by rich ones, whether through direct state financing or "revenue limits" on local school spending. In Fernandez and Rogerson (1996), for instance, replacing an expenditure cap by a tax-financed floor can be Pareto-improving. My model does not deal with these more complex funding schemes, but it does suggest an important caveat: these arguments take the distribution of income as given. Replacing equalization by a floor would certainly alleviate the distortions from state intervention, but it would also slow down homogenization, thereby increasing permanent inequality and the associated costs incurred in production and human capital accumulation. Finally, while segregation and

[24] Unlike the previous one, this demand function is not derived from an optimizing model. This could be done using myopic preferences similar to those in the tracking model of Section I.B, but that is not the purpose of this discussion.

[25] Specifically, $\psi(\eta) = \eta\beta(1 - 2\alpha - \eta\beta)(\sigma - 1)/\sigma$ and $\Psi(\eta) = \eta\beta(\sigma - 1)/[\sigma(1 + \alpha)(1 + \alpha + \eta\beta)]$, where $\alpha = \delta$, $\beta = (1 - \delta)(\sigma - 1)/\sigma$ and $\gamma = (1 - \delta)/\sigma$ are the values corresponding to $\eta = 1$. Note that now $R < 1$, as would be the case if constant returns had not been imposed in (3), for simplicity. Although endogenous growth can be preserved via an economy-wide spillover (see below) one must now distinguish between the long-run behavior of A_t and that of $Y_t = \nu H_t$. Output is clearly the appropriate choice here, but using A_t would lead to fairly similar results: $\phi(\eta) \geq \eta\phi(1)$ and $\Phi(\eta) > 0$ provided $(\sigma - 1)/(\eta - 1)$ is large enough. To maintain $R = 1$ (or any other value) it suffices to augment (4) with an aggregate spillover of the form: $h_{t+1}^i = \kappa\xi_t^i(h_t^i)^\delta(\tau(Y_t^i)^\eta)^{1-\delta}(Y_t)^{(1-\delta)(1-\eta)}$. Proposition 8 and equation (A7) in Appendix A confirm that this alters neither $\psi(\eta)$ nor $\Psi(\eta)$. Adjusting the weight on h_t^i instead would only strengthen the results, by reducing $\psi(\eta)$ and increasing $\Psi(\eta)$.

decentralized funding yield a simple form of path-dependence when $\alpha + \beta = 1$ (or $1/\sigma = 0$), the homothetic framework developed here cannot generate the richer dynamics of a model like Durlauf's (1996): greater intergenerational persistence at low levels of income, poverty traps and nonergodicity of the income distribution.

V. Conclusion

The model developed in this paper is very simple. The accumulation of human capital reflects family, community and economy-wide inputs. The degrees of complementarity or substitutability in local and economy-wide interactions capture the direct costs or benefits of heterogeneity at each level. These three equations encompass the reduced forms of many previous models.

This framework allowed me to study several important issues. I analyzed how economic stratification affects growth, inequality and welfare over different horizons. I showed in particular that integration may slow down growth in the short run but nonetheless increase it in the long run, due to its gradual homogenizing effect. I also compared the performance of privately, locally and nationally funded education systems. Redistributing funds to equalize spending across students leads to the same intertemporal trade-off between initial distortions and long-term efficiency gains as integration. Incorporating the different investment incentives faced by parents and voters under each regime tends to raise the performance (over any horizon) of a national system relative to a private one, and to lower that of community-based funding.

The model could be extended in several directions. For instance, the coexistence of local school finance in the United States with national finance in Europe and Japan represents a puzzle: how can countries whose citizens have similar preferences, technologies and political rights make such radically different societal choices? This question is pursued in Bénabou (1995). The general idea of a dynamic interplay between local complementarities, clustering, and global interactions should also be applicable to a variety of other problems such as labor markets, learning, or economic geography.

Appendix A: More General Education Technologies

In this section I show how to extend the paper's analysis to the more general specification:

(i) There are K local and N globals spillovers, each of which is captured by a CES index: $h_{t+1}^i = F(\xi, h, L_1, \dots, L_K, H_1, \dots, H_N)$.
(ii) F is homogeneous of some degree R in $(h, L_1, \dots, L_K, H_1, \dots, H_N)$.
(iii) The dispersion of initial conditions and the variance of shocks are relatively small: $\mathrm{Var}[\ln h_0^i]$ and $\mathrm{Var}[\ln \xi_t^i] \ll 1$, and both random variables have bounded support.

1. *The Costs of Heterogeneity.*—Recall that for a lognormally distributed random variable the effects of heterogeneity on CES aggregates (or uncentered moments) depend explicitly on the elasticity of complementarity: if $\ln \tilde{h} \sim \mathcal{N}(E[\tilde{h}] - \Delta^2/2, \Delta^2)$ then $H \equiv (\int_0^\infty h^{(\sigma-1)/\sigma} d\mu(h))^{\sigma/(\sigma-1)} = E[\tilde{h}]e^{-\Delta^2/2\sigma}$. The underlying intuition is more general. For any positive random variable $\tilde{h}$, Hölder's inequality implies that H decreases in $1/\sigma$. Moreover if $\tilde{h}/E[\tilde{h}]$ has variance $\Delta^2 \ll 1$ and bounded support, $H \approx E[\tilde{h}]e^{-\Delta^2/2\sigma}$ remains true as a second-order approximation. This will allow us to use Taylor expansions.

2. *Homogeneous Education Production Functions.*—Let $F(\xi, h, L, H)$ be any $\mathscr{C}_2$ function, homogeneous of degree R in (h, L, H). Let $\{\xi_t^i\}_{i\in\Omega}$ and $\{h_t^i\}_{i\in\Omega}$ both satisfy assumption (iii) above, denoting $s^2 \equiv \mathrm{Var}[\ln \xi] \approx \mathrm{Var}[\xi]/E[\xi]^2 \ll 1$ and $\Delta_t^2) \equiv \mathrm{Var}[\ln h_t^i] \approx \mathrm{Var}[h_t^i]/A_t^2 \ll 1$, where $A_t \equiv E[h_t^i]$. For h_t^i, $t \geq 1$ (which unlike ξ_t^i is neither exogenous nor i.i.d.) this is an induction hypothesis, to be checked at $t+1$. Finally, let $\Theta \equiv F(E[\xi], 1, 1, 1)$ and without loss of generality set $(\partial \ln F/\partial \ln \xi)(E[\xi], 1, 1, 1) = 1$. Under integration, we have $\hat{h}_{t+1}^i = F(\xi_t^i, \hat{h}_t^i, \hat{L}_t, \hat{H}_t) = F(\xi_t^i, \hat{h}_t^i, \hat{A}_t(1 - \hat{\Delta}_t^2/2\varepsilon) + o(\hat{\Delta}_t^2), \hat{A}_t(1 - \hat{\Delta}_t^2/2\sigma) + o(\hat{\Delta}_t^2))$. Expanding to the second order around

$(E[\xi], \hat{A}_t, \hat{A}_t\hat{A}_t)$, then integrating over $i \in \Omega$ yields:

$$\hat{A}_{t+1} = F\cdot(1 + \mathrm{Var}(\xi)F_{\xi\xi}/2F$$
$$+ [\hat{A}_t^2 F_{hh}/F - \hat{A}_t F_L/\varepsilon F - \hat{A}_t F_H/\sigma F]\hat{\Delta}_t^2/2)$$
$$+ o(\hat{\Delta}_t^2) + o(s^2).$$

Using homogeneity to evaluate F and all derivatives at $(E[\xi], 1, 1, 1)$—this will be denoted by a "bar"—and then taking logs, yields:

$$\text{(A1)} \quad \ln \hat{A}_{t+1} = \theta + R \ln \hat{A}_t - \hat{\mathcal{L}}\hat{\Delta}_t^2/2 + o(\hat{\Delta}_t^2) + o(s^2),$$

$$\text{(A2)} \quad \hat{\mathcal{L}} \equiv (-\bar{F}_{hh} + \bar{F}_L/\varepsilon + \bar{F}_H/\sigma)/\bar{F},$$

and $\theta \equiv \ln \Theta + (\bar{F}_{\xi\xi}/\bar{F})\mathrm{Var}(\xi)/2$; note that $F_{\xi\xi} = 0$ when ξ_t^i enters multiplicatively. Similarly, under segregation $h_{t+1}^i = F(\xi_t^i, h_t^i, h_t^i, H_t) = F(\xi_t^i, h_t^i, h_t^i, A_t(1 - \Delta_t^2/2\sigma) + o(\Delta_t^2) + o(s^2))$ leads to

$$A_{t+1} = F\cdot(1 + \mathrm{Var}(\xi)F_{\xi\xi}/2F$$
$$+ [A_t^2(F_{hh} + F_{LL} + 2F_{hL})/F - A_t F_H/\sigma F]\Delta_t^2/2)$$
$$+ o(\Delta_t^2) + o(s^2) \Rightarrow$$

$$\text{(A3)} \quad \ln A_{t+1} = \theta + R \ln A_t - \mathcal{L}\Delta_t^2/2 + o(\Delta_t^2) + o(s^2),$$

$$\text{(A4)} \quad \mathcal{L} \equiv (-\bar{F}_{hh} - \bar{F}_{LL} - 2\bar{F}_{hL} + \bar{F}_H/\sigma)/\bar{F}.$$

Note that $\mathcal{L}$ and $\hat{\mathcal{L}}$ are both *constant*, as are $\alpha \equiv hF_h/F = \bar{F}_h$, $\beta \equiv LF_L/F = \bar{F}_L$ and $\gamma \equiv HF_H/F = \bar{F}_H$. Moreover, the laws of motion for variances are, to a second-order approximation: $\hat{\Delta}_{t+1}^2 = \alpha^2\hat{\Delta}_t^2 + s^2 + o(\hat{\Delta}_t^2) + o(s^2)$ and $\Delta_{t+1}^2 = (\alpha + \beta)^2\Delta_t^2 + s^2 + o(\Delta_t^2) + o(s^2)$. The induction hypotheses $\hat{\Delta}_t^2 \ll 1$, $\Delta_t^2 \ll 1$ (uniformly in t) are thus satisfied if $\alpha + \beta \leq 1$, with strict inequality if $s^2 > 0$. The rest of the analysis proceeds as in the text and in the proofs of Propositions 2 and 3, given in Appendix B.

PROPOSITION 8: *Let $\alpha + \beta \leq 1$, with strict inequality when $s^2 > 0$. The short- and long-run gaps between the integrated and segregated economies are given by the same expressions as in Propositions 1 to 3, with $\hat{\mathcal{L}}$ and $\mathcal{L}$ now defined by (A2) and (A4). Thus $\phi = \mathcal{L} - \hat{\mathcal{L}} = -(2\bar{F}_{hL} + \bar{F}_{LL} + \bar{F}_L/\varepsilon)/\bar{F}$, and $\Phi \equiv \mathcal{L}/(1 - (\alpha + \beta)^2) - \hat{\mathcal{L}}/(1 - \alpha^2)$.*

The case where F is a CES index is of particular interest; for any $\alpha' + \beta' + \gamma' = 1$, $R \geq 0$ and $\lambda \gtreqless 0$, let

$$\text{(A5)} \quad h_{t+1}^i = \Theta\xi_t^i[\alpha'(h_t^i)^{(\lambda-1)/\lambda} + \beta'(L_t^i)^{(\lambda-1)/\lambda} + \gamma'(H_t)^{(\lambda-1)/\lambda}]^{R\lambda/(\lambda-1)}.$$

When $\lambda \to 1$ we obtain (11), with $\alpha \equiv R\alpha'$, $\beta \equiv R\beta'$ and $\gamma \equiv R\gamma'$. Proposition 8 implies $\mathcal{L} = (\alpha' + \beta')(1 - \alpha' - \beta')/\lambda + \gamma'/\sigma + (1 - R)(\alpha' + \beta')^2$ and $\hat{\mathcal{L}} = \alpha'(1 - \alpha')/\lambda + \beta'/\varepsilon + \gamma'/\sigma + (1 - R)(\alpha')^2$, so Φ can be rewritten as

$$\text{(A6)} \quad \Phi = \frac{\beta}{1-\alpha^2}\left[\frac{1}{\lambda} - \frac{1}{\varepsilon} + \left(\frac{1}{\sigma} - \frac{1}{\lambda}\right) \times \frac{\gamma(2\alpha+\beta)}{1-(\alpha+\beta)^2} + \left(\frac{1-R}{R}\right) \times \left(1 - R - \frac{1}{\lambda}\right)\frac{(2\alpha+\beta)}{1-(\alpha+\beta)^2}\right].$$

When $1/\lambda = 1/\varepsilon = 1/\sigma$ and $R = 1$, social structure has no steady-state effect on total human capital, as $\Phi = 0$. Aggregation causes similar losses whether it occurs at the level of a child's education (within $F(h, L, H)$), of a community (within L) or of the whole economy (within H). Departing from this neutral point by decreasing local complementarity $1/\varepsilon$ lowers the cost of reducing heterogeneity through integration, while increasing global complementarity $1/\sigma$ raises the corresponding benefit (if $\gamma > 0$). That benefit also goes up with complementarity $1/\lambda$ between the three inputs; this explains in particular why $\Phi > 0$

at the origin on Figure 3, where $1/\lambda = 1$ and $1/\varepsilon = 1/\sigma = 0$. The effects of R are generally ambiguous, because it affects (α, β, γ) and interacts with λ in the nonmonotonic function $(1-R)(1-R-1/\lambda)/R$.

3. *Multiple Spillovers.*— Denote by ε_k and σ_n respectively the elasticities of substitution of the L_k^i's and H_n's entering $h' = F(\xi, h, L_1, \ldots, L_K, H_1, \ldots, H_N)$, and by $\beta_k \equiv \partial \ln F/\partial \ln L_k$, $\gamma_n \equiv \partial \ln F/\partial \ln H_n$ their weights (partial elasticities) in F. Finally, let $\beta \equiv \Sigma_{k=1}^K \beta_k$ and $\gamma \equiv \Sigma_{n=1}^N \gamma_n$. Derivations similar to those which lead to (A2) and (A4) easily yield that this model is equivalent to $h' = \tilde{F}(\xi, h, L, H) \equiv F(\xi, h, L, \ldots, L, H, \ldots, H)$, where L and H are CES indices with elasticities given by

$$\text{(A7)} \qquad \beta/\varepsilon \equiv \sum_{k=1}^{K} \beta_k/\varepsilon_k,$$

$$\gamma/\sigma \equiv \sum_{n=1}^{N} \gamma_n/\sigma_n.$$

APPENDIX B: PROOFS OF THE PROPOSITIONS

PROOF OF PROPOSITIONS 2 AND 3:

The variances which solve (18) and (19) are $\Delta_t^2 = \Delta_\infty^2 + (\alpha+\beta)^{2t}(\Delta^2 - \Delta_\infty^2)$ and $\hat{\Delta}_t^2 = \hat{\Delta}_\infty^2 + \alpha^{2t}(\Delta^2 - \hat{\Delta}_\infty^2)$, where $\Delta_\infty^2 \equiv s^2/(1-(\alpha+\beta)^2)$ and $\hat{\Delta}_\infty^2 \equiv s^2/(1-\alpha^2)$. Note that this is true *even when* $\alpha + \beta > 1$ or $\alpha > 1$, in which case the fixed points do not represent asymptotic variances. It even holds for $\alpha + \beta = 1$ or $\alpha = 1$, using l'Hospital's rule. Next, the solutions to (21)–(23) are

$$\text{(B1)} \quad \ln A_t = R^t \ln A_0 + \left(\theta - \mathcal{L}\frac{\Delta_\infty^2}{2}\right) \times \left(\frac{1-R^t}{1-R}\right) - \frac{\mathcal{L}}{2}(\Delta^2 - \Delta_\infty^2) \times \left(\frac{R^t - (\alpha+\beta)^{2t}}{R-(\alpha+\beta)^2}\right), \quad \text{and}$$

$$\text{(B2)} \quad \ln \hat{A}_t = R^t \ln A_0 + \left(\theta - \hat{\mathcal{L}}\frac{\hat{\Delta}_\infty^2}{2}\right) \times \left(\frac{1-R^t}{1-R}\right) - \frac{\hat{\mathcal{L}}}{2}(\Delta^2 - \hat{\Delta}_\infty^2) \times \left(\frac{R^t - \alpha^{2t}}{R-\alpha^2}\right).$$

Suppose first that $(\alpha+\beta)^2 < R$, which includes $\alpha + \beta < 1$ as a special case. If $R \leq 1$, taking limits as $t \to \infty$ immediately yields Propositions 2 and 3, for $s^2 = 0$ and $s^2 > 0$ respectively. For $R > 1$, one gets

$$\text{(B3)} \quad \ln(\hat{A}_t/A_t) \approx \frac{1}{2}\left(\frac{\mathcal{L}}{R-(\alpha+\beta)^2} - \frac{\hat{\mathcal{L}}}{R-\alpha^2}\right) \times \left(\Delta^2 + \frac{s^2}{R-1}\right) R^t.$$

Note that the coefficient is the same as in equation (28), so that the usual interpretations apply. Suppose next that $(\alpha+\beta)^2 > R$ (requiring $\alpha + \beta > 1$); then

$$\text{(B4)} \quad \ln(\hat{A}_t/A_t) \approx \frac{\mathcal{L}}{2}\,\frac{\Delta^2 + s^2/((\alpha+\beta)^2 - 1)}{(\alpha+\beta)^2 - R} \times (\alpha+\beta)^{2t}.$$

Thus only $\mathcal{L} = (\alpha+\beta)(1-\alpha-\beta) + \gamma/\sigma$ matters, as the effects of exploding heterogeneity under segregation ultimately swamp their counterparts under integration. Finally, consider the borderline case $(\alpha+\beta)^2 = R$. If $\alpha + \beta > 1$, the term $(\alpha+\beta)^{2t}$ in (B4) is simply replaced by $t(\alpha+\beta)^{2(t-1)}$. If $\alpha + \beta = 1$, then $\gamma = 0$, hence $\mathcal{L} = 0$. Therefore $\ln(\hat{A}_t/A_t) \approx -t\hat{\mathcal{L}}\hat{\Delta}_\infty^2/2$.

PROOF OF PROPOSITION 6:

1. *Private Education.*—To simplify the exposition, I first assume that all agents except i

choose invariant participation and savings rates: $\nu_t^j = \bar{\nu}$ and $\tau_t^j = \bar{\tau}$. Later on I show that these restrictions are not binding. The Bellman equation for agent i is thus

$$\text{(B5)}\quad W_t(h_t^i) = \max_{(\nu,\tau)} \{ \ln((1-\tau) \times (\nu h_t^i)^{(\sigma-1)/\sigma}(Y_t)^{1/\sigma}) + \rho E W_{t+1}(\kappa \xi_t^i((1-\nu)h_t^i)^\delta \times (\tau(\nu h_t^i)^{(\sigma-1)/\sigma}(Y_t)^{1/\sigma})^{1-\delta}) \}.$$

The first order conditions are

$$\text{(B6)}\quad \frac{1}{\nu_t^i}\left(\frac{\sigma-1}{\sigma}\right) = \rho\left(\frac{\delta}{1-\nu_t^i} - \frac{1-\delta}{\nu_t^i}\left(\frac{\sigma-1}{\sigma}\right)\right) \times E\left[h_{t+1}^i \frac{\partial W_{t+1}}{\partial h}(h_{t+1}^i)\right]$$

$$\text{(B7)}\quad \frac{1}{1-\tau_t^i} = \frac{\rho(1-\delta)}{\tau_t^i} \times E\left[h_{t+1}^i \frac{\partial W_{t+1}}{\partial h}(h_{t+1}^i)\right].$$

Since the state of the economy is characterized by the Markov process (A_t, Δ_t^2), let us guess the form of the value function to be $W_t(h) = a\ln h + b\ln A_t - c\Delta_t^2 + d$. This leads to $\nu_t^i = \nu$, $\tau_t^i = \tau$, with

$$\text{(B8)}\quad \nu = \frac{1+\rho(1-\delta)a}{1+\rho(1-\delta)a+\rho\delta a\sigma/(\sigma-1)},$$

$$\tau = \frac{\rho(1-\delta)a}{1+\rho(1-\delta)a}.$$

Equilibrium then requires that $\bar{\nu} = \nu$ and $\bar{\tau} = \tau$. Replacing in (B5) and using the recursion equations (18) and (21) for A_{t+1} and Δ_{t+1}^2 identifies the constants. In particular

$$\text{(B9)}\quad a = \frac{(\sigma-1)/\sigma}{1-\rho\delta-\rho(1-\delta)(\sigma-1)/\sigma},$$

$$a+b = \frac{1}{1-\rho},$$

$$c = \frac{b}{2\sigma^2}\,\frac{\rho\mathcal{L}+(1-\rho)/\sigma}{1-\rho(\alpha+\beta)^2},$$

where $\alpha = \delta$, $\beta = (1-\delta)(\sigma-1)/\sigma$, $\gamma = (1-\delta)/\sigma$ and $\mathcal{L}$, $\hat{\mathcal{L}}$ are given by (22)–(24). Replacing a in (B8) yields the desired results. In this equilibrium, all agents choose the same constant values ν and τ. This is in fact the only solution, at least in the following sense. Consider any finite horizon ($T < \infty$) version of the model, without the restrictions $\nu_t^j = \bar{\nu}$, $\tau_t^j = \bar{\tau}$. A backwards induction similar to the derivations above shows that in each period: (a) $\nu_t^i = \nu_t$ and $\tau_t^i = \tau_t$, ensuring $Y_t = \nu_t H_t$; (b) $W_t(h) = a_t \ln h + b_t \ln A_t - c_t\Delta_t^2 + d_t$, where the (a_t, b_t, c_t, d_t)'s satisfy linear difference equations whose fixed points are (a, b, c, d) calculated above. The finite game thus has a unique Markov perfect equilibrium, and it involves symmetric strategies. As $T \to \infty$, it tends to the equilibrium derived above.

2. *National Public Education.*—Let us start again with some useful simplifications. First, let $\nu_t^j = \bar{\nu}$ for all $j \neq i$, so that the Markov process $(\hat{A}_t, \hat{\Delta}_t^2)$ fully describes the state of economy. This restriction is relaxed later on. Note from (23) and (19) that the tax rate $\tau_{t,N}$ implemented at t influences $\hat{A}_{t+1}$ through $\Theta_t = \kappa(1-\bar{\nu})^\delta(\bar{\nu}\tau_{t,N})^{1-\delta}$, but *does not affect* $\hat{\Delta}_t^2$. Second, let agent i choose his preferred rate τ_t^i as if he expected to be the decisive voter (or a dictator) not just at t but in all future periods. Because this leads to the same preferred tax rate for all agents, any other voting game (for example, agent i chooses τ_t^i as if he were decisive at t but expected the median voter to prevail at $t' > t$) will lead to the same outcome. Agent i's

Bellman equation and first order conditions are therefore

$$\text{(B10)}\quad V(h_t^i, \hat{A}_t, \hat{\Delta}_t^2) = \max_{(\nu,\tau)}\{\ln((1-\tau) \times (\nu h_t^i)^{(\sigma-1)/\sigma}(\hat{Y}_t)^{1/\sigma}) + \rho E V(\kappa \xi_t^i((1-\nu^i)h_t^i)^{\delta} \times (\tau \hat{Y}_t)^{1-\delta}, \hat{A}_{t+1}, \hat{\Delta}_{t+1}^2)\}$$

$$\text{(B11)}\quad \frac{1}{\nu_t^i}\left(\frac{\sigma-1}{\sigma}\right) = \frac{\rho\delta}{1-\nu_t^i} \times E\left[h_{t+1}^i \frac{\partial V}{\partial h}(h_{t+1}^i, \hat{A}_{t+1}, \hat{\Delta}_{t+1}^2)\right]$$

$$\text{(B12)}\quad \frac{1}{1-\tau_t^i} = \frac{\rho(1-\delta)}{\tau_t^i} \times E\left[h_{t+1}^i \frac{\partial V}{\partial h}(h_{t+1}^i, \hat{A}_{t+1}, \hat{\Delta}_{t+1}^2)\right] + \rho \frac{\partial \hat{A}_{t+1}}{\partial \tau} \times E\left[h_{t+1}^i \frac{\partial V}{\partial A}(h_{t+1}^i, \hat{A}_{t+1}, \hat{\Delta}_{t+1}^2)\right].$$

Equation (B12), which determines τ_t^{i*}, corresponds to a voter who internalizes the effect of taxes on $\hat{A}_{t+1}$, hence on $\hat{H}_{t+1}$ and $\hat{Y}_{t+1}$. If he does not, the rate corresponding rate τ_t^{i*} is obtained by dropping the last term. Again, I guess the form of the value function: $V(h, \hat{A}, \hat{\Delta}^2) = \hat{a} \ln h + \hat{b} \ln \hat{A} - \hat{c}\hat{\Delta}^2 + \hat{d}$. Then (B11)–(B12) imply $\nu_t^i = \hat{\nu}$, $\tau_t^{i*} = \tau_*$ and $\tau_t^{i*} = \tau^*$, where

$$\text{(B13)}\quad \hat{\nu} = \frac{1}{1 + \rho\delta\hat{a}\sigma/(\sigma-1)}$$

$$\text{(B14)}\quad \tau_* = \frac{\rho(1-\delta)\hat{a}}{1+\rho(1-\delta)\hat{a}} < \frac{\rho(1-\delta)(\hat{a}+\hat{b})}{1+\rho(1-\delta)(\hat{a}+\hat{b})} = \tau^*.$$

Whereas τ_* is based on the *private* marginal value of human capital $\hat{a}$, τ^* reflects the full *social* marginal value $\hat{a} + \hat{b}$. Replacing in (B10) and using the recursion equations for $(\hat{A}_{t+1}, \hat{\Delta}_{t+1}^2)$ leads to

$$\text{(B15)}\quad \hat{a} = \frac{(\sigma-1)/\sigma}{1-\rho\delta},$$

$$\hat{a} + \hat{b} = \frac{1}{1-\rho},$$

$$\hat{c} = \frac{\hat{b}}{2\sigma^2}\left(\frac{\rho\hat{\mathcal{L}} + (1-\rho)/\sigma}{1-\rho\alpha^2}\right).$$

Replacing in (B13) and (B14) leads to the desired results. Note that $\hat{a} < a < \hat{a} + \hat{b}$, implying that $\tau_* < \tau < \tau^*$. To exclude other equilibria, consider again the finite-horizon game, without the restriction $\nu_t^i = \bar{\nu}$. An agent's value function and (Markov) strategy depend on $(h_t^i, \hat{\mu}_t)$, where $\hat{\mu}_t$ is the distribution of human capital. Whether voters internalize the effect of $\tau_{t,N}$ on $\hat{\mu}_{t+1}$ or not, backwards induction easily shows that in each period: (a) $\nu_t^i = \hat{\nu}_t$, ensuring in particular that $\hat{\mu}_t$ remains lognormal and that $\hat{Y}_t = \hat{\nu}_t \hat{H}_t$; (b) $\tau_t^i = \hat{\tau}_t$, that is, there is *unanimity* over the sequence of tax rates; (c) $V_t(h, \hat{\mu}_t) = \hat{a}_t \ln h + \hat{b}_t \ln \hat{A}_t - \hat{c}_t\hat{\Delta}_t^2 + \hat{d}_t$, where the $(\hat{a}_t, \hat{b}_t, \hat{c}_t, \hat{d}_t)$'s satisfy linear difference equations whose fixed points are $(\hat{a}, \hat{b}, \hat{c}, \hat{d})$ calculated above. Letting the horizon tends to infinity concludes the proof.

3. *Local Public Education.*—Voters in small communities never internalize economy-wide complementarities in production. If they also ignore the intertemporal fiscal spillover (a higher tax rate at t raises the tax base at $t+1$), whether because children move away or due to or myopia, they choose $\tau_t^i = \tau_*$. If they fully internalize it, which requires perfectly stable communities, they operate like a family under private funding, and set $\tau_t^i = \tau$. These two numbers also provide lower and upper bounds in situations of partial mobility or limited forward-looking.

REFERENCES

Alesina, Alberto and Rodrik, Dani. "Distributive Politics and Economic Growth." *Quarterly*

Journal of Economics, May 1994, *109*(2), pp. 465–90.

Atkinson, Anthony. "On the Measurement of Inequality." *Journal of Economic Theory*, September 1970, *2*(3), pp. 244–63.

Banerjee, Abhijit and Besley, Timothy. "Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior." Working Paper No. 68, Princeton University, December 1990.

Becker, Gary. *A treatise on the family*. Cambridge, MA: Harvard University Press, 1981.

Bénabou, Roland. "Workings of a City: Location, Education, and Production." *Quarterly Journal of Economics*, August 1993, *108*(3), pp. 619–52.

______. "Unequal Societies." Mimeo, New York University, April 1995.

______. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies*, April 1996, *63*(2), pp. 237–64.

Bergstrom, Theodore; Rubinfeld, Daniel and Shapiro, Perry. "Micro-Based Estimates of Demand Functions for Local School Expenditures." *Econometrica*, September 1982, *50*(5), pp. 1183–205.

Borjas, George. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics*, February 1992, *107*(1), pp. 123–50.

______. "Ethnicity, Neighborhoods, and Human Capital Externalities." *American Economic Review*, June 1995, *85*(3), pp. 365–90.

Coleman, James. *Foundations of social theory*. Cambridge, MA: Harvard University Press, 1990.

Cooper, Suzanne. "A Positive Theory of Income Redistribution." Mimeo, Stanford University, July 1992.

Dertouzos, Michael; Lester, Richard and Solow, Robert. *Made in America: Regaining the productive edge*, Report of the MIT Commission on Productivity. Cambridge, MA: MIT Press, 1989.

Downes, Thomas and Shoeman, David. "School Financing Reform and Private Enrollment: Evidence from California." Mimeo, Northwestern University, June 1993.

Durlauf, Steven. "A Theory of Persistent Income Inequality." *Journal of Economic Growth*, March 1996, *1*(1), pp. 75–94.

Fernandez, Raquel and Rogerson, Roger. "Income Distribution, Communities and the Quality of Public Education." *Quarterly Journal of Economics*, February 1996, *111*(1), pp. 135–64.

Glaeser, Edward; Kallal, Hedi; Scheinkman, Jose and Shleifer, Andrei. "Growth in Cities." *Journal of Political Economy*, December 1992, *100*(6), pp. 1126–52.

Glomm, Gerhard and Ravikumar, B. "Public vs. Private Investment in Human Capital: Endogenous Growth and Income Inequality." *Journal of Political Economy*, August 1992, *100*(4), pp. 818–34.

Goldberger, Arthur. "Economic and Mechanical Models of Intergenerational Transmission." *American Economic Review*, June 1989, *79*(3), pp. 504–13.

Hanushek, Eric. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, September 1986, *24*(3), pp. 282–94.

Jargowsky, Paul. "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." Institute for Research on Poverty Discussion Paper No. 1056–95, January 1995.

Jencks, Christopher and Mayer, Susan. "The Social Consequences of Growing Up in a Poor Neighborhood," in Laurence Lynn and Michael McGeary, eds., *Inner city poverty in the United States*. Washington, DC: National Academy Press, 1990, pp. 111–86.

Kozol, Jonathan. *Savage inequalities: Children in America's schools*. New York: Crown Publishers, 1991.

Loury, Glenn. "A Dynamic Theory of Racial Income Differences," in Phyllis Wallace and Anette Le Mond, eds., *Women, minorities and employment discrimination*. Lexington, MA: Lexington Books, 1977, pp. 153–88.

______. "Intergenerational Transfers and the Distribution of Earnings." *Econometrica*, July 1981, *49*(4), pp. 843–67.

Lucas, Robert Jr. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, July 1988, *22*(1), pp. 3–42.

Montgomery, James. "Social Networks and Persistent Inequality in the Labor Market: Toward an Economic Analysis." *American*

Economic Review, December 1991, *81*(5), pp. 1408–18.

Murphy, Kevin; Shleifer, Andrei and Vishny, Robert. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics*, May 1991, *106*(2), pp. 503–30.

***New York Times*.** "The New Math of Rich and Poor." 10 June 1990, p. E5.

Perotti, Roberto. "Political Equilibrium, Income Distribution, and Growth." *Review of Economic Studies*, October 1993, *60*(4), pp. 755–76.

Persson, Torsten and Tabellini, Guido. "Is Inequality Harmful for Growth? Theory and Evidence." *American Economic Review*, June 1994, *84*(3), pp. 600–21.

Picus, Lawrence. "Cadillacs or Chevrolets: The Evolution of State Control over School Finance in California." *Journal of Education Finance*, Summer 1991, *17*, pp. 33–59.

Rauch, James. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics*, November 1993, *34*(3), pp. 1002–37.

Saint-Paul, Gilles and Verdier, Thierry. "Education, Democracy and Growth." *Journal of Development Economics*, December 1993, *42*(2), pp. 399–408.

Streufert, Peter. "The Effect of Underclass Social Isolation on Schooling Choice." Working Paper DP 954-91, University of Wisconsin-Madison, August 1991.

Tamura, Robert. "Income Convergence in an Endogenous Growth Model." *Journal of Political Economy*, June 1991, *99*(3), pp. 522–40.

______. "Efficient Equilibrium Convergence: Heterogeneity and Growth." *Journal of Economic Theory*, December 1992, *58*(2), pp. 355–76.

Weiss, Martin. *The Clustering of America*. New York: Harper and Row, 1989.

Wilson, William Julius. *The Truly Disadvantaged*. Chicago: University of Chicago Press, 1987.