European Economic Review 36 (1992) 603–611. North-Holland

Distribution and growth in models of imperfect capital markets

Philippe Aghion

European Bank for Reconstruction and Development, London, UK
DELTA, Paris, France

Patrick Bolton

Laboratoire d'Econométrie, Ecole Polytechnique and Institut d'Etudes Européennes, Université Libre de Bruxelles, Brussels, Belgium

> A full analysis of the distribution of income should include both the inequality in income between different generations of the same family – what is usually called intergenerational 'social' mobility – and the inequality in income between different families in the same generation. [Becker-Tomes (1979, p. 1154)]

1. Introduction

This paper reviews recent attempts to explain the generation of 'social' mobility and inequality in a market economy. It also discusses the interaction of mobility and the distribution of wealth with the process of capital accumulation. Any attempt at explaining intergenerational mobility as well as the endogenous generation of inequality must include uncertainty in individual incomes. This basic insight has first been exploited by Gibrat (1931) and Champernowne (1953) to derive an endogenous distribution of individual income and wealth resembling those observed in the major developed market economies. Before turning to a discussion of the existing theories of distribution and mobility it is useful to briefly mention some fundamental stylised facts.

Consider first the distribution of income. It has been repeatedly established

that the distribution of the lowest incomes and the middle range of incomes approximates respectively a normal and lognormal distribution. As for the highest incomes, the best approximation is the Pareto distribution.[1] It is remarkable that this shape of the income distribution appears to be stable over time. The distribution of wealth has similar properties at least for the middle range of wealth levels, as individual wealth in this range is the result of the accumulation of an individual's lifetime savings; but for the highest wealth levels matters are more complicated since most of this wealth is inherited [see Atkinson (1971)].

The measurement of mobility is more problematic since one must now track the wealth levels of several generations; very often adequate statistical records of the evolution of wealth levels within a family across several generations are hard to find. We shall just mention one study based on U.K. data by Goldthorpe (1980) whose findings are representative of the findings of other studies (in the U.K. and elsewhere). Instead of looking at income (or wealth) directly Goldthorpe establishes a relation between the occupations of father and son within a family. In table 1, occupations are classified into seven categories ranging from unskilled jobs (entry VII) to senior managerial positions or equivalent occupations (entry I). Table 1 reveals a substantial degree of mobility across all occupations; it is highest within the middle range of activities but it is relatively low at the top and at the bottom. Moreover, Goldthorpe (1980) reports that mobility is positively correlated with growth and access to education.

Another important stylised fact worth mentioning is the well known relation between growth and the distribution of income first uncovered by Kuznets (1955). His statistical findings point to an inverse U-shaped relation between income inequality (measured by the variance of the logarithm of income) and GNP per head.[2]

The theoretical contributions we discuss below attempt to explain these stylised facts. Section 2 deals with the determination of the long run distribution of wealth and mobility. It stresses the importance of capital market imperfections as the source of the endogenous generation and persistence of wealth inequalities. Section 3 discusses the interdependence of the distribution of wealth and the degree of mobility with the process of development. For lack of space we shall only provide a selective review of the literature reflecting our personal interests.

[1]The Pareto distribution has the property that the mass of incomes above a given level decreases at a constant rate in that level. For a recent discussion of the statistical properties of existing income distributions see Phelps-Brown (1988).

[2]The relation between growth and income inequality uncovered by Kuznets is based on a cross section regression of individual countries' levels of GNP per head and the degree of inequality in their respective income distributions. The lowest and highest levels of GNP per head are associated with low inequality but middle levels are associated with higher inequality.

Table 1

Class of distribution of respondents by class of father at respondent's age 14.

Father's class	Respondent's class (1972) (percentage by row)								
	I	II	III	IV	V	VI	VII	*N*	%
I	45.7 (45.2)	19.1 (18.9)	11.6 (11.5)	6.8 (7.7)	4.9 (4.8)	5.4 (5.4)	6.5 (6.5)	680 (688)	7.9 (7.3)
II	29.4 (29.1)	23.3 (23.1)	12.1 (11.9)	6.0 (7.0)	9.7 (9.6)	10.8 (10.6)	8.6 (8.7)	547 (554)	6.4 (5.9)
III	18.6 (18.4)	15.9 (15.7)	13.0 (12.8)	7.4 (7.8)	13.0 (12.8)	15.7 (15.6)	16.4 (16.9)	687 (694)	8.0 (7.3)
IV	14.0 (12.6)	14.4 (11.4)	9.1 (8.0)	21.1 (24.4)	9.9 (8.7)	15.1 (14.4)	16.3 (20.5)	886 (1,329)	10.3 (14.1)
V	14.4 (14.2)	13.7 (13.6)	10.2 (10.1)	7.7 (7.7)	15.9 (15.7)	21.4 (21.2)	16.8 (17.6)	1,072 (1,082)	12.5 (11.5)
VI	7.8 (7.8)	8.8 (8.8)	8.4 (8.3)	6.4 (6.6)	12.4 (12.3)	30.6 (30.4)	25.6 (25.9)	2,577 (2,594)	30.0 (27.5)
VII	7.1 (6.5)	8.5 (7.8)	8.8 (8.2)	5.7 (6.6)	12.9 (12.5)	24.8 (23.5)	32.2 (34.9)	2,126 (2,493)	24.8 (24.6)
N	1,230 (1,285)	1,050 (1,087)	827 (870)	687 (887)	1,026 (1,091)	1,883 (2,000)	1,872 (2,214)	8,575 (9,434)	
%	14.3 (13.6)	12.2 (11.5)	9.6 (9.2)	8.0 (9.4)	12.0 (11.6)	22.0 (21.2)	21.8 (23.5)		

Source: Goldthorpe (1980).

2. Endogenous risk bearing and the generation of inequality

To understand why wealth is distributed unequally one must find out how inequality emerges in the first place and why it tends to be reproduced from generation to generation. The simplest explanation of the emergence and persistence of inequality is to attribute differences in earnings and savings behaviour to inherent differences in productivity which may be genetically transmitted from parent to child. Genetic differences among individuals may explain inequality in earnings but an explanation based on genetic differences alone cannot account for the stylised facts that mobility is increasing with growth and the extension of education [see Goldthorpe (1980)], nor can it account for the so called Kuznets curve described above.

The theories we discuss here identify other reasons for the existence of inequality which are not based on genetic differences between individuals. In order to abstract from those latter considerations we shall take the view that individuals are essentially identical and that differences in their wealth are attributable to differences in luck possibly amplified by differences in inherited wealth. When individuals are assumed to be intrinsically identical it is no longer obvious that initial differences in wealth necessarily persist over time. In fact persistence of inequalities across generations is possible only if capital markets are imperfect.[3] A recent paper highlighting the role of capital market imperfections in the reproduction of inequalities is Galor and Zeira (1988). Their basic point is simple and can be stated as follows: in the absence of capital market imperfections every individual has access to the same investment opportunities so that future wealth inequalities can at most reflect initial inequalities; and when the proportion of income saved is non-increasing in initial wealth then all individuals must end up with the same wealth level. However, if borrowing involves a deadweight loss, then individuals with low inherited wealth no longer have access to the same investment opportunities as individuals with high initial wealth. Galor and Zeira show that this is sufficient for initial wealth inequalities to persist in steady state.

They do not explain where the initial wealth inequalities come from. This is all the more troubling that the shape of the steady state distribution in their model depends on the shape of the initial distribution. In particular if there is perfect equality initially then there is also perfect equality in steady

[3]In a thought-provoking paper Mincer (1958) explains that observed differences in income may simply reflect the profile of lifetime earnings in a world where individuals invest in human capital when they are young. This form of income inequality is compatible with perfect equality in the net present value of lifetime earnings if the capital market is perfect. If one does not correct for age one would see persistent inequality in such a world. Of course, what we have in mind here is persistence of inequality in the net present value of wealth which we argue in the text can only be explained through the existence of capital market imperfections. Note also that Mincer's theory cannot explain the stylised facts about mobility.

state. To explain why perfect equality cannot persist one must again appeal to capital market imperfections which prevent individuals from insuring themselves perfectly against future income uncertainty. Thus, in his seminal contribution Champernowne (1953) shows how the lognormal distribution emerges as the unique invariant distribution when individuals cannot insure against income risk taking the form of idiosyncratic shocks increasing or lowering wealth by an amount equiproportional to initial wealth. It is important to emphasize that this long run distribution is independent of the initial wealth distribution. (Champernowne also shows that under an alternative set of assumptions about the distribution of idiosyncratic shocks the Pareto form can be obtained as a unique invariant distribution). Note that Champernowne's theory also accounts for persistent intergenerational mobility so that – according to Becker and Tomes' standards – his work is the first successful attempt at providing a 'full analysis of the distribution of income'. There are, however, important limitations in Champernowne's model. Individual behaviour is completely mechanic and the assumed form of capital market incompleteness (that is the absence of any insurance possibilities) is not discussed. The more recent contributions, following in the footsteps of Champernowne, deal with both of these shortcomings.

Some of the microfoundations missing in Champernowne have first been introduced by Loury (1981). He allows for optimising behaviour by agents in letting savings and investments (in human capital) be determined by intertemporal utility maximisation over consumption and bequests. He postulates an extreme form of capital market incompleteness: both insurance markets and credit markets are missing. As a result both mobility and inequality in wealth emerge; moreover these inequalities persist over time; in fact, Loury establishes the existence and uniqueness of an invariant wealth distribution. While Champernowne was able to rely on a law of large numbers theorem to prove the same result (because of the assumed independence of individual income shocks) Loury could not use the same methods since intertemporal maximisation, in his model, introduces correlation between income variables of parents and children within the same family. (Correlation arises as a result of investment in human capital which has the effect of shifting the earnings distribution; because of the absence of credit markets the amount of investment in human capital is directly related to inherited wealth so that earnings are correlated across generations of the same family). To prove existence and uniqueness of an invariant distribution Loury relies instead on the contraction mapping theorem, which incidentally also establishes convergence to the unique invariant distribution starting from any initial wealth distribution.[4] The contraction mapping property of

[4]Several other papers introduce maximising behaviour into models of wealth inequality, most notably Becker and Tomes (1979) and Eckstein, Eichenbaum and Peled (1985); but these papers do not establish the existence and uniqueness of an invariant distribution.

the income distribution arises from the assumed diminishing returns to human capital investments together with the strict concavity of individual utility functions (over consumption and bequests).

Loury has thus shown that one can reconcile Champernowne's approach to the determination of an endogenous distribution of wealth with individual optimizing behaviour. It remains to incorporate into the theory the possibilities for borrowing and lending or insurance observed in practice. This has first been attempted in a path-breaking paper by Banerjee and Newman (1991). The main innovation of their paper is to derive capital market imperfections from the well known incentive problems in financial contracting. When one takes account of incentive problems it is easy to see that perfect insurance is feasible but not desirable and that borrowing may be rationed.

Banerjee and Newman focus on the classic moral hazard problem of (unobservable) effort supply by risk averse agents as formalized in Mirrlees (1975) and Holmstrom (1979). Their basic set up comprises a continuum of agents who may differ only in their initial wealth. This wealth can either be invested in the capital market at an exogenously fixed riskless rate of return, or in the agent's own risky venture (all individual projects are of the same fixed size and the returns on those projects are independently, identically distributed). The expected return on individual projects is higher the more effort the individual supplies. At the first best level of effort the expected return on an individual project is higher than the market rate of return.[5] Since agents are risk averse they prefer to share the risks of their projects with other investors by issuing equity on the capital market. But the more shares they issue the smaller is the fraction of returns on the individual project they obtain and hence the lower is their incentive to supply effort. Thus, they cannot diversify away all the risk on their individual project for then they would not credibily supply the enough effort. The extent to which individuals can diversify risk depends on their initial wealth holdings.

Banerjee and Newman specify a utility function for agents such that individuals with higher inherited wealth must hold a larger fraction of shares in their own projects to meet the incentive constraints for effort supply (intuitively this property is verified if agents have a diminishing marginal utility of income, so that to compensate for a given marginal cost of effort they must obtain a larger fraction of marginal returns the higher their initial wealth). When initial wealth levels are very high the fraction of shares an individual must hold is so large that the risk exposure is not worth the investment in the individual project. This feature of their model is central for the derivation of a unique invariant distribution. It prevents the distribution

[5]The first best level of effort is given by the level equating marginal expected returns with the marginal cost of effort.

of wealth from spreading out indefinitely and it implies that, on average, the wealth of the poor increases faster than the wealth of the rich. This property of their model allows Banerjee and Newman to establish the existence of a unique ergodic distribution by appealing to convergence theorems for stationary Markov processes [see Doob (1953)].[6]

The model of Banerjee and Newman thus provides a complete theory of the distribution of wealth (in a simulation study it appears that their limit distribution resembles those observed in practice). It also develops a compelling explanation for why individual income risk is not fully diversified away. However, the particular form of capital market imperfection they focus on seems unrealistic. In practice investors are more demanding towards would-be entrepreneurs (in terms of collateral) the poorer they are, while in Banerjee and Newman the opposite is true. It is also odd that the poor's wealth grows faster on average than the wealth of the rich. In practice poor individuals are typically credit rationed, which prevents them from accumulating wealth as fast as the rich. Another shortcoming of Banerjee and Newman is the assumption of an exogenously determined market rate of return. Because of this assumption their model is not completely closed. The next section briefly describes a model of ours attempting to remedy those two deficiencies. It turns out that in the process we are able to establish a relation between capital accumulation and the distribution of wealth which resembles the above mentioned Kuznets curve.

3. Wealth distribution and growth

The general question of how the distribution of wealth affects growth in GNP and how in turn the process of accumulation impacts on the distribution of wealth is wide and has received a great deal of attention.[7] The literature on this issue, however, does not attempt to endogenously derive the distribution of wealth, nor does it analyse how the process of capital accumulation affects mobility (thus, the fact mentioned above that growth is associated with greater mobility has not yet found an adequate

[6]These theorems are based on the same property as the contraction mapping theorem (namely that low wealth levels tend to be mapped into higher wealth levels on average while high wealth levels tend to be mapped into lower future wealth levels) but apply to a larger class of transformations. The transformation in Banerjee and Newman does not satisfy the contraction mapping property but does satisfy sufficient conditions guaranteeing (weak) convergence of the stationary Markov process of wealth distributions.

[7]Most notably this issue has been a dominant theme of research in Cambridge in the 1950s with major contributions by Kaldor, Robinson, Goodwin, Pasinetti and others. A common thread in this literature is the separation of the economy into two classes, workers and capitalists, with differing propensities to save and to analyse how growth affects and in turn is affected by the distribution of output between these two classes. This literature does not attempt to provide microfoundations for the behaviour of the two classes, nor does it account for mobility (see Mervin King's introduction).

explanation). Aghion and Bolton (1991) is a first attempt at addressing these questions.

The basic set-up is that of Banerjee and Newman, with the following differences: agents are now risk neutral and the source of capital market imperfection is a moral hazard problem with limited wealth constraint as in Sappington (1983). Therefore individual agents now prefer self financing of their projects and they share the project's returns with other investors only if their initial wealth is insufficient to cover the set up costs of the project. The more they need to borrow to invest in their own project the less incentives they have to supply effort since they must now share the marginal returns from effort supply with other investors (because of the limited wealth constraint 'safe debt' is not a feasible financial contract so that marginal returns to effort supply must be shared with outside investors). In other words, the level of effort supplied in equilibrium is a decreasing function of the amount borrowed. As a result, three classes of agents emerge endogenously in this model: the very wealthy who have enough funds to invest not only in their own project but also in the projects of others; the middle class composed of agents investing only in their own project (to cover the set up costs of their project they must complement their initial wealth with loans); and finally the poor who do not invest in their own project (either because of credit rationing or because the size of repayments for these agents is so large that it is not worth investing in the own project). The equilibrium terms at which the middle class can borrow is determined by equalising the aggregate demand for funds (emanating from this class) with aggregate supply (by the very wealthy and the poor).

How does capital accumulation affect mobility and the distribution of wealth in this set-up? Note first that as more capital is accumulated there are more funds available in the economy to finance individual projects; as a result mobility naturally increases with growth. Second, capital accumulation tends to shift gradually equilibrium lending terms in favor of borrowers since there are more and more funds available to lend to a smaller and smaller pool of borrowers. This basic effect can give rise to a Kuznets curve since in the early phases of development the lending terms are favourable to the lenders so that their wealth grows relatively faster.[8] In later stages the lending terms become more favourable to borrowers so that the wealth of the middle class can catch up with that of the rich (note also that as the equilibrium cost of capital decreases a larger fraction of the poor can invest in their own individual projects). In other words, initial phases of growth increase inequalities while later stages are accompanied by a reduction in inequalities. The gradual erosion of lending terms eventually puts an upper

[8]Note also that less favourable lending terms for the borrowers has a negative effect on their effort supply; this is another reason for which inequalities in wealth between borrowers and lenders tend to increase in early stages of development.

bound on the accumulation of an individual's wealth. Thus if capital accumulation is sufficiently rapid the distribution of wealth eventually converges to a unique invariant steady state distribution with compact support.

Finally, it remains to determine how the distribution of wealth affects in turn the rate of growth in the aggregate capital stock. Roughly speaking, one can distinguish between two phases in this set up; in early phases of development (when, say, no individual has enough inherited wealth to cover the set up costs of the individual project) an increase in inequalities may accelerate growth to the extent that fewer people need to borrow to invest in their own projects. In later phases, however, a redistribution of wealth from the very wealthy to the middle class enhances growth for the same reason. It is worth emphasizing that the motive for redistribution here is based entirely on incentive considerations, since agents are risk neutral. This is in contrast to the literature on optimal income taxation, where insurance or equity considerations are the motive for redistribution and where the extent of redistribution is limited by incentive constraints [see for example Mirrlees (1971)].

References

Aghion, P. and P. Bolton, 1991, A trickle-down theory of growth and development with debt-overhang, Mimeo.

Atkinson, A., 1971, The distribution of wealth and the individual life-cycle, Oxford Economic Papers 23, 239–254.

Banerjee, A. and A. Newman, 1991, Risk-bearing and the theory of income distribution, Review of Economic Studies 58, 211–235.

Becker, G. and N. Tomes, 1979, An equilibrium theory of the distribution of income and intergenerational mobility, Journal of Political Economy 6, 1153–1189.

Champernowne, D., 1953, A model of income distribution, Economic Journal 63, 318–351.

Doob, J., 1953, Stochastic processes (Wiley, New York).

Eckstein, Z., M. Eichenbaum and D. Peled, The distribution of wealth and welfare in the presence of incomplete annuity markets, Quarterly Journal of Economics 403, 789–806.

Galor, O. and J. Zeira, 1988, Income distribution and macroeconomics, Mimeo. (Hebrew University of Jerusalem).

Gibrat, R., 1931, Les inégalités économiques (Paris).

Goldthorpe, J., 1980, Social mobility and class structure in modern Britain (Oxford).

Holmstrom, B., 1979, Moral hazard and observability, Bell Journal of Economics 10, 79–91.

Kuznets, S., 1955, Economic growth and income inequality, American Economic Review 45, 1–28, Reprinted in his Economic growth and structure (London, 1966) 257–287.

Loury, G., 1981, Intergenerational transfers and the distribution of earnings, Econometrica 49, 843–867.

Mincer, J., 1958, Investment in human capital and personal income distribution, Journal of Political Economy 66, 281–302.

Mirrlees, J., 1971, An exploration in the theory of optimum income taxation, Review of Economic Studies 38, 175–208.

Mirrlees, J., 1975, The theory of moral hazard and unobservable behavior, Mimeo. (Nuffield College, Oxford).

Phelps-Brown, H., 1988, Egalitarism and the generation of inequality (Clarendon Press, Oxford).

Sappington, D., 1983, Limited liability contracts between principal and agent, Journal of Economic Theory 29, 1–21.