I. A BRIEF INTRODUCTION TO SOME THEORIES OF GROWTH AND DISTRIBUTION

Economists who went to graduate school in the 1970s and 80s may well be forgiven for viewing “inequality” as a peripheral topic. Those who think about it at all will almost certainly think of it as an ‘outcome’ – the result of some underlying distribution of assets; of the work and saving decisions of individuals; and of prices determined by a range of different markets: for land, labor, capital and goods. Up until early in this decade, the contemporary economics profession seemed to have little to say about the impact that inequality – or the distributions of income and wealth, more generally - might have on other variables, such as the overall efficiency of an economy proxied, for instance, by its growth rate.

This had not always been the case. A concern with the importance of distribution was central to the thinking of classical economists such as David Ricardo and Karl Marx. Atkinson (1997, p.298) cites Ricardo as arguing that Political Economy should be “an enquiry into the laws which determine the division of the produce of industry, amongst the classes who concur in its formation.” Later, one strand of nascent growth theory in the 1950s had distribution playing a crucial role: the capital-labor ratio in Kaldor (1956, 1957) was driven to its steady-state equilibrium value by the different savings rates of ‘capitalists’ and workers: if K/L rose above its equilibrium value, the wage-to-profits ratio would also rise. With savings out of wages assumed to be lower than out of profits, this led to a decline in the rate of capital accumulation, driving K/L back down towards equilibrium.

There was another strand of growth – or development – theory in the 1950s in which distribution played an important role. This was based on the path-breaking works of Arthur Lewis (1954) and Simon Kuznets (1955). Lewis’s model of growth “with unlimited supplies of labor” was fundamentally different from Kaldor’s (or Solow’s), in that it was driven by a movement of a factor of production (labor) from a low-productivity sector to a higher-productivity one. Kuznets’s seminal 1955 contribution owed much to the observation that, if inequality between these two sectors was rather more substantial than that within each sector, then inequality would first rise – as people moved across sectors - and then fall, as most of them found themselves in the new sector, or the economy reached a point where factor movement was equalizing returns across sectors. Whence the stylized Kuznets ‘inverted-U’ curve.

But ‘mainstream’ growth theory, as most of us came to learn it, evolved from another paper written in 1956, by Robert Solow. Unlike Kaldor’s, Solow’s model did not require a distributional mechanism to generate a stable growth path. Instead, it relied on a production
function $F(K, L)$ which displayed constant returns to scale and had diminishing returns to each factor. It was therefore concave with respect to the capital-labor ratio. Given the assumption of a constant and exogenous population growth rate, and certain conditions on $F(K, L)$, this sufficed for there to be a unique capital-labor ratio which was universally stable (in the sense that the system converged to it from any other ratio). Because it relied on a neoclassical production function, this model inaugurated what came to be known as neoclassical growth theory. It spawned an important literature in the 1960s, which sought to address two perceived shortcomings of Solow’s basic model. Arrow (1962), Uzawa (1965) and others sought to bring technical progress and the generation of new ideas into the picture, and explain their links to growth. Separately, Cass (1965), building on earlier work by Ramsey, replaced Solow’s exogenous, arbitrary savings rate with an inter-temporal consumption path chosen so as to optimize fully explicit inter-temporal preferences.

These were interesting and often complex theoretical issues. Since dealing with them did not require an explicit distribution of income or wealth – and introducing one would have complicated things further - all of these models were based on what came to be known as a ‘representative agent’. The problem of economic growth was studied as if society were monolithic, making inter-temporal choices as a single individual would. By the time Romer (1986) and Lucas (1988) came along, building on the models of the 1960s to generate set-ups where positive steady-state growth in per capita income was driven by ‘endogenous’ technical progress or by human capital accumulation, this was something of an established tradition. Growth models generally described economies made up of a single, representative agent. The importance of distribution, so crucial to the classical economists, to Kaldor and to Kuznets, seemed to have been lost somewhere along the way.

Despite the reliance of growth theory on ‘representative’ agents, societies are patently not homogeneous, whether in incomes, wealth, or any other dimension. In fact, when we say that incomes in a certain population grew by x%, we are in fact referring to the growth rate of the mean of the distribution of incomes across that population. ‘Aggregate’ growth statistics, however useful, refer only to the evolution of one moment of the distribution, ignoring any changes in the rest of the distribution. Which is why, now that computing power allows us, more and more people investigate growth as a process of distributional dynamics, focusing on changes in the entire distribution, rather than only on its mean.

This is not to say that the rate of change in the mean is not important. It is, in fact, terribly important. It is merely to say that, by looking at growth as a process of distributional dynamics, we can look both at the mean and at a number of other phenomena, such as inequality and inter-distributional mobility, which may also be of interest. Figure 1 shows the density functions for the income distribution in Chile, from 1987 to 1994. It is clear that mean income was growing. But it is also clear that other things were happening to the shape of the distribution which, as it turned out, mattered to the welfare of poor people, for instance.
Thinking of ‘economic growth’ as a process of distributional dynamics is one thing we hope you will find natural when leaving this web site. But social scientists and economists are not interested only in describing what happens. They are mainly interested in linkages, causes, explanations. Are there any empirical regularities relating, say, the dispersion of a distribution and its growth rate? And if so, which way does the causation flow? Does growth affect the level of inequality? Does initial inequality affect the rate of growth? And if any of these is true, why may that be the case? The next sections address these questions.

II. THE EFFECT OF GROWTH ON DISTRIBUTION

Until recently, the most well established view on linkages between growth and income distribution was the Kuznets hypothesis. As indicated above, this postulated that growth (from the low income levels associated with predominantly agrarian societies) would first lead to an increase, and then to a decrease in income inequality. This was captured diagrammatically in the “Kuznets curve”, depicted on Figure 2.

Figure 2: A Stylized Kuznets Curve
The empirical backing for this came from Kuznets’s investigation of a time-series of inequality indicators for England, Germany and the United States. In the 1950s, these were basically the only countries for which a sufficiently long time series was available and, by that time, inequality was indeed falling in all three countries, after having risen earlier. As mentioned in Section I, the economic mechanism thought to underlie this phenomenon was the transfer of labor from low-productivity (and low inequality) sectors, to high-productivity (and medium inequality) sectors. The result would hold provided that inequality between the sectors was substantially greater than the inequality within them. Given the data available at the time, it was impossible to reject this rather sensible hypothesis.

So sensible did it seem, in fact, that the Kuznets curve became one of the stylized facts of the study of income distribution for nearly four decades. It was only recently that tests of the hypothesis based on much larger data sets (both across countries and over time for individual countries) have consistently refuted it. These studies were largely made possible by the compilation, in 1995-6, of the Deininger-Squire (1996a) international inequality database (http://www.worldbank.org/growth/dddeisqu.htm), which contains 682 ‘high-quality’ observations (of Gini coefficients and quintile shares) for 108 countries. Based on their own analysis, Deininger and Squire (1998), conclude that “our data provide little support for an inverted-U relationship between levels of income and inequality, when tested on a country-by-country basis, with no support for the existence of a Kuznets curve in about 90% of the countries investigated.” (Deininger and Squire 1998, p.573). On the basis of the much more plentiful information available to today’s empirical researchers, then, there seems to be no support for the Kuznets hypothesis.

There was also, however, a slightly modified ‘dynamic version’ of the hypothesis, which postulated that fast growth episodes tended to lead to higher inequality, regardless of the initial level of income. This is what is implied, for instance, by the popular observation, in a number of countries, that growth leads to ‘the rich getting richer, while the poor get poorer’. Deininger and Squire also investigated that possibility, by considering growth episodes defined by the availability of distributional data that spans at least one decade. They concluded that: “there appears to be little systematic relationship between growth and changes in aggregate inequality” (1996b, p. 587). Periods of growth were almost equally as often associated with increases in inequality as they were with declines.

More recently, however, it seems that economic reform in the transition economies of Eastern Europe and Central Asia (ECA) may have changed the nature of that empirical result. Looking at a sample of 64 changes in mean income and inequality, or ‘spells’, between 1981 and 1994, Ravallion and Chen (1997) (http://www.worldbank.org/research/journals/wber/revmay97/what.htm) find a significant negative correlation between economic growth and changes in inequality. In other words, the sample suggests that growth reduces inequality, rather than contributing to increases in it. The effect vanishes, however, when one eliminates the ECA spells from the sample. This is brought out vividly by Figure 3 below, taken from Ravallion and Chen (1997, p.370).
It may be reasonable to conclude that the negative link between growth and inequality detected in this study was brought about by the rather specific circumstances of transition in Eastern Europe and Central Asia, where negative growth and increasing inequality both prevailed since 1990, not necessarily one because of the other. As the horizontal regression line in the above diagram suggests, for countries outside that region there is no evidence of a systematic relationship between growth and changes in inequality. Furthermore, the line’s position at zero change in inequality suggests that, independently of the rate of growth, the average change in inequality in this sample of ‘spells’ was nil.

Results of this nature have led most economists to adopt a more skeptical view about systematic causal links running from economic growth (a rising distribution mean) to inequality (changes in the dispersion of the distribution). The skepticism seems to be justified whether we look for links in levels or changes. Recent evidence seems to refute the Kuznets hypothesis about an inverted-U relationship between the level of income and the level of inequality. And outside transition economies, growth episodes do not on average seem to be associated with contemporaneous increases in inequality.

But it is important to recall that one should not confuse inequality and poverty. By shifting the distribution function to the right, there is overwhelming evidence that economic growth contributes to reductions in poverty. Figure 4a and b below (from Ahuja et.al., 1997, pp.40-41) illustrate the impact of economic growth on poverty in Thailand, between 1975 and 1992.
Figure 4a

**Thailand: Cumulative Distribution Functions, Detail of Low-Income Range, 1975-92**

Source: Ahuja et al (1997)

Figure 4b

**Thailand: Cumulative Distribution Functions, 1975-92**
Of course, it is possible that individual episodes of distributional changes combine a growth in the mean with an increase in inequality large enough to raise poverty. This took place, for instance, in the Philippines, between 1988 and 1991 (See Ahuja et. al., p.47). However, these episodes are definitely the exception. Figure 5 below, taken from Ravallion and Chen, 1997, p.377, plots the regression lines of changes in log poverty (with the poverty line set at 75% of the initial mean welfare indicator) on growth rates, both for ECA and other countries. The (statistically significant) negative slope for both groups of countries indicates the strong average association between growth and poverty reduction. (Notice that the horizontal distance between the two lines can be interpreted as the extra growth needed to obtain the same reduction in poverty in ECA, given its increases in inequality.)

Lest there be any lingering doubt about the power of the effect of growth on poverty reduction, notice that the first and third quadrants of Figure 5 (i.e. positive growth and rising poverty; and negative growth and falling poverty) are effectively empty. While we can not find an empirical link between growth and contemporaneous inequality, there is now no doubt that growth helps to reduce poverty.
III. THE EFFECT OF DISTRIBUTION ON GROWTH

But does the fact that contemporaneous changes in mean incomes and inequality do not seem to be systematically related necessarily imply that there is no link between the two variables at all? Could it be, for instance, that countries with higher initial (i.e. *ex ante*) inequality grow more slowly (or faster) than others? And if any such pattern emerged, what would explain it? Are there causal links between the dispersion of an income (or expenditure, or wealth) distribution and economic growth, or indeed other aspects of economic performance?

In the 1990s, the classical view that distribution (one aspect of which is measured by inequality indices) is not only a final outcome, but in fact plays a central role in determining other aspects of economic performance, has come back into fashion. While many economists often start working on a topic at the same time, much of the credit for pioneering this line of enquiry must go to Oded Galor and Joseph Zeira, whose 1993 paper on “Income Distribution and Macroeconomics” concluded thus: “In general, this study shows that the distributions of wealth and income are very important from a macroeconomic point of view. They affect output and investment in the short and in the long run and the pattern of adjustment to exogenous shocks. It is, therefore, our belief that this relationship between income distribution and macroeconomics will attract more studies in the future.” (1993, p.51).

They were certainly right in their last prediction. And most of the studies that followed concurred that they were right on the first point as well. Empirically, the proposition that initial inequality seemed to be associated with lower growth rates was put forward by Persson and Tabellini (1994) and Alesina and Rodrik (1994). Using the data sets available to them, both studies found that inequality variables had significantly negative coefficients in growth regressions, when controlling for a number of the usual right-hand side variables, such as initial income, schooling and physical capital investment. A survey by Benabou (1996a) listed a number of other cross-country empirical investigations of this relationship, and reported that the vast majority of them reached the same conclusion.

But the debate is not as settled as that survey implied. Since then, a number of studies drawing on the Deininger-Squire database - which is superior to those available to Persson and Tabellini or to Alesina and Rodrik, both in number of countries and in time-spans covered – have questioned the budding consensus. Adding a number of econometric methodological improvements to this better data set, Forbes (1997) actually finds a positive and significant relationship between inequality and growth. Others have echoed her concerns and cautioned against a premature acceptance of the inverse relationship between initial inequality and growth as a new stylized fact of development economics.

However, the econometric problems that seem to beset the negative relationship in the newer data sets appear to be specific to inequality variables defined in the *income* (or expenditure) space. Deininger and Squire themselves, for instance, find that the negative coefficient on initial income inequality in their growth regressions becomes insignificant only when a variable for *asset* inequality (the Gini coefficient for land ownership) is introduced. Birdsall and Londono (1997) investigate a similar relationship for the other asset crucially important for the poor,
namely human capital. Also using a subset of the Deininger-Squire database, they conclude that: “…initial inequalities in the distribution of land and of human capital have a clear negative effect on economic growth, and the effects are almost twice as great for the poor as for the population as a whole.” (p.35) Once again, once these asset distribution variables are included, the significance of income inequality disappears.

The empirical issue is clearly not settled. Nevertheless, it would seem fair to report the current state of the debate as follows: while initial income inequality may not directly affect an economy’s aggregate growth potential, others thing being equal, it does proxy for more fundamental inequalities of wealth. Once measures for those are included, there seems to be a significant negative relationship between asset inequality and growth.

Now, why would that be the case?

**III.1. POLITICAL ECONOMY CHANNELS**

One of the mechanisms though which inequality was suggested to affect growth relied heavily on political economy considerations. The models suggested by Alesina and Rodrik (1994), Bertola (1993) and Persson and Tabellini (1994) were early examples of this strand of the literature. While each paper is unique, one might nevertheless seek to summarize the basic insights of this strand as follows.

Policy decisions are not taken by a benign government seeking to maximize some social welfare function. Instead, they are the result of political interactions which can be modeled as votes on the values that certain policy variables (such as tax rates) are to take. A key result of this political economy literature is the ‘median voter theorem’ which, simply put, states that if: (I) preferences for some such policy variable (say: a proportional tax rate \( \tau \)) vary monotonically across the distribution of some attribute of the population (say, their incomes), and (II), each person has one vote (with equal weight), then the preference of the voter which occupies the median position in that distribution will be the outcome of the voting process.

Now, if the government is such that the net effect of its taxation and expenditure policies is redistributive (i.e. the difference between the individual’s tax bill and any benefits she receives from government expenditure increases with income), then it is easy to show that preferences for the value of \( \tau \) decline monotonically with income. And if one measures inequality as the distance between the median and the mean voter, then for a given mean income, it follows that greater inequality (i.e. a lower income for the median) leads to the choice of a higher tax rate. Add to this any distortionary effect of taxation – such as a disincentive to exert effort, or to save - and you have a model for why greater inequality leads to lower growth:
This argument was sufficiently persuasive to spawn a considerable literature. In addition to the aforementioned Benabou (1996a) paper, this literature is also surveyed in Bertola (1998), a chapter in the *Handbook of Income Distribution*, edited by Atkinson and Bourguignon (forthcoming). Unfortunately, it encountered a rather serious empirical problem, namely that the cross-country correlation between inequality and tax rates is negative. This may well be because, as Perotti (1996) shows, the effect of the fiscal system in many high-inequality countries is actually regressive. Be that as it may, the empirical evidence on the intermediate role of distortionary taxation as the channel linking higher inequality to lower growth is not particularly supportive.

The proponents of the importance of political economy considerations as central to the relationship between inequality and growth have thus turned away from tax choice, and towards the role of political instability and its effect on investment behavior. Alesina and Perotti (1996) is a prominent example.

But even if political channels are likely to play some role, they are not the only channel that economists have suggested to account for the empirical association between wealth inequality and lower growth. We now turn to the possible role of market – in particular capital market – failures.

**III.2. CAPITAL MARKET IMPERFECTIONS**

The effects of income distribution on taxation and government expenditure are not the only mechanisms through which it is likely to exert an influence on economic efficiency. A more basic channel, in fact, is that productive opportunities might vary along the wealth distribution. Put simply: poor people may not have the same chances in life as richer people, and may thus never quite realize their full productive potential. This may be because they do not get as good an education as those afforded by richer families, or because they can’t get loans to start up a
business as easily, or because they can’t afford the insurance they would require to undertake some risky – but productive – venture. In this case, a distribution with lots of poor people, or unequally distributed opportunities, would under-utilize its aggregate productive potential to a greater degree than a distribution with relatively fewer poor people, or one where opportunities were more equitably distributed. In static terms, it would be less efficient and have a smaller production possibility set. In a dynamic setting, it would be likely to grow more slowly.

During the last decade, these intuitive ideas were formalized in various models. Most of them formalize the differences in opportunities along the income distribution by assuming non-convex production sets.¹ In Galor and Zeira (1993) or Aghion and Bolton (1997), there is a single, fixed scale of production, or ‘project size’. In Ferreira (1995) there is a minimum scale or project size, above which the production set is convex. Whatever the particular details, the idea is that there are fixed costs, or minimum ‘lumpy’ investments, which are required to gain access to some productive activity. This could be a school or college fee; it could be the price of the smallest viable plot of land in an agricultural community; or the permit to operate a stall in a market. Wherever you look, the argument goes, you can find people who would be engaged in a more productive activity, had they only had the minimum initial ‘investment’ required to enter it.

These models all share another basic property; at least one capital market must be either missing or imperfect. The most obvious case is a missing or imperfect credit market, where repayment is not costlessly enforceable, leading either to a wedge between lending and borrowing interest rates (e.g. Galor and Zeira, 1993) or to collateral requirements (e.g. Banerjee and Newman, 1993). In either case, the equilibrium distribution may be characterized by a group of people who fail to invest in the most productive sector of the economy, despite being as able and entrepreneurial as anyone else, simply because they either do not have enough wealth for the collateral, or are trapped in a low-return, high borrowing rate situation. In this group of models, the poor are prevented from choosing the most productive activity available given their skills, because imperfect information and incomplete contracts cause a credit market failure: loans that would have been good are not made, and applicants that are turned down remain poorer than they would otherwise be.

In some cases, the equilibrium distribution is ergodic (Ferreira, 1995), implying that a poor person (or lineage) will eventually claw its way out of poverty, just as some rich people (or lineages) will experience downward social mobility. In others, the limiting distributions are not ergodic, and poor people may remain forever trapped in poverty (Galor and Zeira, 1993).

Another capital market which may be imperfect, and cause inequality both to persist in the long run and to lead to inefficiencies, is the insurance market. Models of this type usually have risk-averse agents who wish to hedge the risk of their own projects by selling shares in them, and buying shares in other ventures (project risk is assumed to be entirely idiosyncratic). For a given degree of risk aversion, there will generally be a minimum share of the project which the agent must sell in order to be willing to take it on. But if the disutility of effort supply increases in wealth, and effort is not monitorable, it is possible that the upper tail of the distribution needs to retain ownership of a share of the project – so that their commitment to supply effort is credible – which is too large in terms of their risk aversion. In this case, this rich group becomes a class of rentiers, lending their wealth rather than investing. In this case, it is an insurance market failure

¹ One exception is Piketty (1997), who assumes perfectly convex production sets.
which leads to a group (now the rich) not investing in the most productive activity. This is the essence of Banerjee and Newman (1991).

Aghion and Bolton (1997) and Piketty (1997) also belong to this ‘moral hazard in effort supply’ class of models. But in their cases, it is the credit market which is imperfect, and the poor who are excluded from the desirable activity. Effort is not fully observable. Once again, it depends on the share of returns to the project which accrue to the enterprising agent, rather than to the lender. Hence, the lower the initial wealth of the borrower (and thus the larger the loan required to enter), the less likely it is that she will supply the required effort. This leads lenders to establish a maximum loan size (or minimum initial wealth) for lending.\(^2\) Agents poorer than that minimum wealth level are excluded from borrowing, and hence from the most profitable sector.

Other authors have built upon these models to investigate how robust is the basic result that, in the presence of capital market failures, greater ex-ante inequality leads to lower efficiency. Ferreira (1995) investigates the possibility that governments are not necessarily only a source of inefficiency and distortion, by introducing a role for public capital in private production functions. In that model, because richer agents can substitute private alternatives for public capital, but the poor are more dependent on State expenditures in services like health, education, public transport and basic infrastructure, public expenditure can play an equalizing and efficiency-augmenting role. Benabou (1996b) also considers the possibility that government intervention may be beneficial rather than only distortionary. He combines capital market failures with a political economy setting, to show that political-economic systems may display multiple equilibria. If taxation and expenditure have a positive net effect on efficiency, then there may exist both a high inequality, low redistribution, low efficiency equilibrium; and a low inequality, high redistribution, high efficiency equilibrium.

This strand of the literature does not suffer from the empirical problems associated with the earlier political economy channel, because it does not predict a positive correlation between inequality and average tax rates. It does predict a negative correlation between ex-ante wealth inequality and economic growth which, as described in Section III above, many authors have now found evidence of. It has gained further empirical support from a recent study by Barro (1999), which finds that the negative effect of inequality on growth is significant for poorer countries, but vanishes above a certain per capita income threshold. This is consistent both with the fact that financial markets are least developed in these poor countries, and with the fact that they are likely to have distributions with more mass below the relevant production set non-convexity (as long as this grows by less than mean income, during the process of development).

### III.3. Social Conflict Channels

Imperfect credit and insurance markets; government redistribution with a positive net effect on output; and the outcome of voting systems can offer alternative – or complementary – explanations for the observation that wealth (or asset) inequality is negatively related to economic growth. But they are not the only reasons. A more recent strand of the literature emphasizes the importance of social conflict as a link between inequality and efficiency. Alesina

\(^2\) An example of an incentive-compatibility constraint.
and Perotti (1996) argue that inequality can lead to less political stability, and this in turn can lead to sub-optimal investment levels.

Rodrik (1997) suggests a related but distinct mechanism. Greater inequality, he argues, may increase the share of resources in a political system which are dedicated to bargaining over the distribution of rents (or of the burdens of a negative shock), thus reducing its effectiveness in terms of agreeing a rapid response to those shocks. He suggests that this may have been one reason why East Asian economies adjusted to the Debt Crisis of the 1980s so much more rapidly and effectively than Latin America. The magnitude of the external shocks hitting ‘representative’ countries in both regions was broadly similar, but Latin America was substantially more unequal.

But social conflict does not affect economic performance only by increasing political instability and unrest, thereby reducing investment levels (Alesina and Perotti, 1996), or by crippling the ability of political systems to respond effectively to external shocks (Rodrik, 1997). It may also have high opportunity costs caused by violence. Violence levels, as measured by recorded homicide rates, have recently increased sharply in both of the most unequal regions in the world (Latin America and sub-Saharan Africa), and in the one where its growth has been fastest (Eastern Europe, Russia and Central Asia). Fajnzylber et. al. (1998) document these global trends, and find evidence to suggest that income inequality is significantly associated with violence levels, across countries. Bourguignon (1998) and others have documented the growing importance of the social and economic burden imposed on society by this rising violence, both in terms of the direct costs in lives and medical resources, and in terms of the opportunity costs of (both public and private) resources diverted from other activities towards preventing and fighting crime.

IV. CONCLUDING COMMENTS

In sum, income and wealth distributions can no longer be seen as mere outcomes of the general equilibrium of an economy. The central processes that determine resource allocation – through capital markets, through the political system, and through social circumstances – are influenced by the distribution of wealth in important ways. More unequal societies tend to develop larger groups of people who are excluded from opportunities others enjoy – be they a better education, access to loans, or to insurance – and who therefore do not develop their full productive potentials. Both theory and empirical evidence suggest that these incomplete realizations of economic potential are not of concern only to those who care about equity per se. They also affect aggregate economic potential, and therefore aggregate output and its rate of growth.

The inverted-U relationship between growth and inequality suggested by Kuznets has not survived recent empirical scrutiny terribly well. Instead, it is gradually being replaced by a perception that the main flow of causation may be in the other direction, with inequality hampering the rate and quality of economic growth. The debate is not over, either conceptually or empirically. But its very liveliness attests to the importance of the question. To paraphrase Tony Atkinson, inequality is unlikely to go back out into the cold periphery of economic analysis any time in the foreseeable future.
REFERENCES


