Inequality and Redistribution: The Need for New Perspectives.

Agnar Sandmo

Abstract.

Recent years have seen a marked increase in inequality in the OECD area, and since the early 1990s the trend has been unmistakable in Norway also. This paper offers a brief review of the main causes behind this development, in particular globalisation, skill-biased technical change, tax reform and deregulation, and a greater role for non-competitive labour markets. As an example of the last point it reviews a tournament model of wage formation and its implications for tax incidence and inequality. Lastly, the paper argues that the expenditure tax may be a better alternative than the dual income tax for combining concerns for economic efficiency and a more even distribution of income after tax.

JEL Classification: D3, H2.

---

This paper is based on a presentation at the Oslo Autumn Conference in November 2004. The presentation was a follow-up to a lecture by Tony Atkinson; see Atkinson (2004). I am indebted to Mats Persson and Bertil Tungodden for helpful comments on an earlier version.
Inequality and Redistribution: The Need for New Perspectives.

Recent years have seen a marked increase in statistical measures of inequality in the OECD area. The trend has been most pronounced in the United States and the United Kingdom, but it has been clear in other countries as well; see e.g. the documentation in Gottschalk and Smeeding (1997) and Atkinson (1995, especially chapters 1 and 2; 2004). In Norway inequality was for a long time more stable than in many other countries, but at least since the early 1990s the trend has been unmistakable here also. From 1986 to 2002, according to the social indicators of Statistics Norway, the income share of the lowest decile in the distribution of income fell from 4.1 to 3.6 per cent\(^2\). Over the same period the share of the top decile rose from 18.6 to 23.6 per cent. It is also of interest to note that the share of labour earnings in total income fell from 78 to 71 per cent, while income from capital increased its share from 5 to 8 per cent. The percentage share of transfer income rose during the same period from 17 to 21\(^3\).

It has of course to be kept in mind that while income inequality is increasing, Norway is still near the bottom of the scale in international comparisons of income inequality. One should also remember that the distribution of disposable income is an imperfect measure of the distribution of the standard of living. For the standard of living it is also of crucial importance to take account of the degree of free access to publicly supplied goods like health and education, and this is likely to vary both between countries and over time. Another weakness of this measure is that it is based on annual income, while the distribution of lifetime income, especially in rich countries with well-developed capital markets, may be a better indicator of the distribution of the standard of living. However, such data are hard to construct, and we simply have to keep these complications in mind when considering the information on income inequality.

In the following I wish to present some reflections on two issues. The first concerns the sources of increased inequality in Western countries over the past few decades; why is it that the long period of decreasing inequality after World War II has been followed by many years

\(^2\) These figures refer to the distribution of income after tax per consumption unit, computed as household disposable income divided by the square root of the number of persons in the household.

\(^3\) The numbers are taken from the tables in Samfunnspellet, no. 4, 2004. See also the article by Kleven and Mørk (2004) in the same issue.
of increase? The second set of reflections relate to redistributive taxation. What is the scope for progressive taxation, and what should the tax base be?

**The sources of increased inequality.**

The causation behind the recent increase in inequality in the Western world is unlikely to be a simple one; I believe that it must be understood in terms of the interaction of a number of forces whose relative importance may vary from one country to another as well as over time. The following list is brief and selective.

First of all, it is often maintained that the increase in inequality of factor income, particularly income from labour, is due to globalisation. A short version of this theory is that globalisation has liberalised trade and factor movements between the rich and the poor countries of the world. The poor countries have a relative abundance of unskilled labour, while the rich countries have a relatively large endowment of skilled labour. As countries utilize their comparative advantage, the demand for unskilled labour increases in the poor countries, whereas skilled workers in the rich part of the world experience increased demand for their services. Therefore, the skilled-unskilled wage differential increases in the rich countries, while it decreases in the poor countries. If trade unions in the rich countries try to resist the fall in the wages of the unskilled, the result will be increased unemployment - another source of increased inequality. For further discussion and references see e.g. Atkinson (1999) and Sandmo (2003).

Second, it is widely believed that the development of technology has a built-in inequality bias. To an increasing extent, both manufacturing and services require high-skill employees, workers who are able to handle advanced technological equipment. Since the required skills are scarce, those who possess them will be able to sell their services at an increased premium in the labour market. Consequently, the skilled-unskilled wage differential increases. An econometric study by Krusell et. al. (2000) interprets this general hypothesis as *capital-skill complementarity*, i.e. as the hypothesis that capital accumulation raises the marginal productivity of skilled labour but reduces the marginal productivity of the unskilled. They find considerable support for this hypothesis in United States data for 1963-1992, even though there was a substantial increase in the supply of skilled workers during the same period. Acemoglu (2002), whose analysis is based on a broad survey of the literature, confirms this.
conclusion but adopts a different view of the chain of causation than most of the previous literature. He argues that although technical change in the United States has been skill-biased during most of the 20th century, this should primarily be interpreted as the technological response to an increased supply of skilled workers.

Third, during the past couple of decades Western countries have gone through a period of deregulation and tax reform with increased emphasis on economic incentives. It is highly unlikely that individuals respond in a uniform way to improved incentives. Some will react to the substitution effects of lower marginal tax rates by working harder and saving more. Others will change their behaviour only to a very small degree; in other words, the income effects will wholly or partly dominate the substitution effects. The consequence of this could be that the former group of individuals will increase their incomes relative to the latter. In this perspective, some increase of inequality in factor incomes is an expected and unavoidable consequence of the creation of improved incentives for economic efficiency. Blomquist et. al. (2001) present evidence indicating that the Swedish tax reform of the 1980s did indeed contribute to increased inequality. For the United States it has been suggested that part of the increased inequality of labour incomes following the Tax Reform Act of 1986, has been due to income shifting from other tax bases as a response to lower marginal tax rates; see e.g. the discussion in Slemrod (1998). This source of increased inequality is of a different nature than the two previously mentioned, since they are the results of economic policy reforms, not of developments that are due to causes that are exogenous, at least to the national economy. To the extent that this kind of increased inequality can be interpreted as the outcome of some sort of social welfare maximization, it could even be argued that it is desirable, being the outcome of a rational trade-off between efficiency and equality.

These three causes of increased inequality can all be interpreted in terms of the standard competitive model; in particular, they assume that wage rates reflect the marginal productivities of different types of labour. But is the competitive theory of wages a realistic one? There is no simple answer to this question. There can be no doubt that for a number of problems, especially perhaps at a high level of aggregation, the competitive theory is a good guide to understanding how the labour market works. But in applications of a less aggregative kind, such as the development of the personal distribution of income, the answer to the question is much more in doubt, and there have always been economists who have been sceptical to the marginal productivity theory of wages. Indeed, even before the advent of that
theory, John Stuart Mill (1848; 1965, 199-200) thought that while what he called the “laws of production” were essentially determined by technology, the “laws of distribution” were determined by the “laws and customs of society” and “human institutions”.

In addition to the three sets of causes mentioned above, we may have witnessed a change in social attitudes to inequality, Mill’s customs of society. The labour market is a complex mechanism whose operation reflects more than just the impersonal forces of supply and demand. Wage differentials – at least at the level of the individual working place – must be seen as being reasonably fair to be socially acceptable. If they are not, they may give rise to social frictions, e.g. in the form of strikes, which will reduce efficiency. Thus, both workers and management may have an interest in holding down inequality in the workplace; see Atkinson (1999) and Agell (1999, 2002). But attitudes towards fairness and social acceptability may change. It may be that they have indeed changed in favour of more tolerance towards wage and income differentials, and that part of the increased inequality of gross labour incomes may be a reflection of this. However, it should be noted that polls taken in connection with the Norwegian parliamentary elections during the 1990s fail to show such a trend in attitudes; the degree of support for a statement like “taxation of large incomes should be reduced” was remarkably stable during the period.4

**Wage formation under non-competitive conditions.**

One of the assumptions that textbooks often mention as a condition for use of the competitive framework, is that the commodity in question must be homogenous - or at least reasonably so. Applied to the labour market this implies that the labour supply of different individuals must be close, if not perfect, substitutes within the particular firm or industry that we study. Is this condition likely to be met? And is it more or less likely to be met as countries grow richer?

As countries grow richer and the share of skilled labour in the labour force increases, it also means that the degree of differentiation in the skills possessed by workers necessarily increases. As each worker’s skill becomes more unique, there is reason to believe that his bargaining power in wage negotiations increases, so that the differentiation of skills becomes reflected in increased differentiation of wages. This tendency may be more pronounced in

---

4 For an analysis of the 2001 election see Aardal (2003), which also provides references to previous studies.
services than in manufacturing (although it is probably of importance there also). The reason
is that in manufacturing there is likely to be a tighter organization of work which is dictated
by the technological organization of the production process, implying a much higher degree
of capital-labour complementarity. In the service industries, by contrast, individual
characteristics of the employee count for more in terms of the amount and quality of the
output. This has implications for the homogeneity, or non-homogeneity, both of jobs and
workers. Employers in the service industries offer jobs with more variety than in
manufacturing. Similarly, each worker can offer a set of characteristics that make him
different from other workers. This implies that in the bargaining process - broadly defined -
the power to influence wages and other parameters of the labour contract is greater both for
the employer and the employee. The competitive assumption is the limiting case where each
agent’s bargaining power is virtually zero. In the service sector, compared to manufacturing,
the individual worker may have much more reason to believe that the set of qualifications that
he offers to the employer is a unique one, and the employer has jobs to offer that from the
worker’s perspective are different from those offered by other employers. The implication of
this is that the competitive model becomes less credible as a realistic picture of the process of
wage formation.

It is a well-known feature of the economic growth process that as countries grow richer, an
increasing share of the labour force is employed in the service sector, and this may imply that
the part of the economy where wage differentials can be explained by the competitive model
is shrinking over time\(^5\). This development will be reinforced by the fact that the relative
growth of the service \textit{industries} probably underestimates the relative growth of service \textit{jobs},
since service-type jobs tend to account for a larger share of the employment in many
manufacturing industries. This is an additional reason to broaden the theoretical perspective
on the process of wage formation.

To explain wage formation and wage differentials in the expanding service sector, we need, in
view of the above, to focus more on non-competitive or imperfectly competitive models. But
the number of imperfectly competitive models is large, and it is difficult to identify just one of
them as being the obvious choice for a better understanding of wage bargaining in a

\(^5\) According to Official Statistics of Norway (1995) the share of the labour force employed in the public and
private service sectors (exclusive of public administration and personal services) increased from 2.6\% to 25.9\%
over the period 1900-1990.
predominantly service economy. What we need may be a variety of models that can capture the different systems of wage formation in the economy.

Can this be expected to make a major difference to the way in which we view the labour market? Let me try to indicate an answer to this broad question by way of an example. This is taken from an article by Mats Persson and myself (forthcoming) in which we look both at an imperfectly competitive theory of wage inequality as well as the incidence of a redistributive tax and the characteristics of an optimal income tax system.

The theoretical framework of this analysis is taken from the theory of tournaments, as originally formulated by Lazear and Rosen (1981); for a more recent exposition see Lazear (1995). A tournament is a procedure for picking the best candidate for promotion to a higher step on the career ladder in an organization in which the candidate who wins will be rewarded by a higher wage than the loser. In the simple but instructive version of the model that we use, there are just two candidates for promotion, equal both in terms of preferences and productivities. The tournament consists simply in the employer observing the outputs of the two workers in the first period and giving the promotion to the worker with the higher output. The output of each of the workers is determined in part by his effort, in part by a random element. Since the workers are identical, they will in equilibrium supply the same amount of effort, so that the outcome of the tournament is decided by luck; the winner is the worker whose random element in output was more favourable to him. The difference in wages after the tournament does not reflect differences in individual ability or productivity (although the average wage equals productivity). Nevertheless, the wage difference can be justified in terms of productivity because the wage premium offered to the winner calls forth a larger effort on the part of both competitors. Since the whole of the wage premium goes to the winner, this is an example of what Frank and Cook (1995) refer to as winner-take-all markets.

The wage spread is an incentive to effort, but what determines the magnitude of the wage spread? In brief, and somewhat paradoxically, the more important is the random element in the tournament, the larger must the wage spread be in order to induce a given level of effort. In a labour market of this type, the observation that there is a large dispersion of wages should not lead us to believe that there is a correspondingly large variation in individual productivity or effort; the explanation may instead be that luck is more important than effort for the outcome of the tournament.
Into this framework we now insert the simplest type of redistributive tax in the form of a linear income tax. This has a constant marginal tax rate combined with a uniform transfer to all taxpayers. It implies an increasing average rate, so that the tax schedule is progressive. We then perform a comparative statics experiment; we increase progression in the sense of increasing the marginal tax rate while at the same time increasing the transfer payment. What is the effect of increased progression on the wage dispersion before and after tax?

Defining the wage dispersion as the ratio of the highest to the lowest wage, we find that an increase in progression unambiguously increases the dispersion in income before tax. What is more striking is that we also find that, under fairly reasonable assumptions, increased progression may increase the dispersion of wage income after tax - in other words, a more progressive tax schedule may increase inequality. These are results that stand in contrast to most competitive models of tax incidence. In particular, they are at variance with models that are used to study optimal redistribution via a progressive income tax, particularly the models that are descendants of the famous analysis of Mirrlees (1971). In the Mirrlees framework, wage rates net of tax are given, so that inequality can only increase through differences in labour supply responses between low-wage and high-wage groups. In the Persson-Sandmo model taxpayers respond in identical ways to increased progression, and inequality is affected through the equilibrium wage rates.

Our model is based on a number of special assumptions, so that one should be careful about drawing strong conclusions from it. It is probably most significant in drawing attention to the kind of hypotheses that may emerge from increased application of non-competitive models to problems of wage formation, tax incidence and inequality. The example suggests that non-competitive conditions may give rise to new types of inequality that are not easily explained by factors such as globalization or skill-biased technical change, and that tax incidence in this type of labour market should be given more attention both in theoretical and empirical research.

Redistribution policy.

As pointed out in Atkinson (2004), the theory of optimal income tax progressivity identifies the main determinants of the marginal tax rate as
(1) the elasticity of labour supply,

(2) the dispersion of income or wages before tax,

(3) the degree of inequality aversion in the social welfare function.

The causes of increased inequality that I have discussed above influence the three determinants in different ways. Skill-biased technical change increases the dispersion of wages and provides an argument for more redistribution through a higher marginal tax rate. A greater social tolerance of inequality can be interpreted as less inequality aversion, and this should lead to a reduction of the marginal tax rate. Globalisation affects both the dispersion of incomes and, if migration is a problem of significant proportions, the elasticity of labour supply; it therefore provides arguments both for and against more tax progressivity. The analysis of optimal taxation in the Persson-Sandmo tournament model, implies that the marginal tax rate should be higher, the larger is the random element in the determination of wage differentials. If all causes of increased inequality are present simultaneously, then, on balance, it becomes unclear whether the recent increase in inequality favours more or less progressivity in the personal tax schedule.

The redistributive effects of direct taxation depend not only on the formal degree of progressivity, but also on the legal definition of the tax base. The Norwegian tax reforms that took place in the 1980s and '90s were similar to tax reforms in many other countries in that they combined a reduction in the degree of progressivity with a broadening of the tax base (see Christiansen (2004) for an excellent survey of the history of Norwegian tax reform). A major change was introduced in 1992 with the so-called dual income tax, which limited progressivity to the taxation of labour income, while capital income was taxed at the flat rate of 28 per cent, equal to the rate of corporate income tax. The main argument for proportional taxation of capital income was based on two features of the previous system; first, the tax-favoured status of a number of household assets such as housing, and second, the full deductibility of interest payments against a high marginal tax rate. By reducing the marginal tax rate on capital, one would move closer to a system of neutral taxation of capital income. But while to a large extent achieving this, one also introduced a substantial gap between the marginal tax rates on capital and labour.
Why should capital income be taxed at lower rates than wage income? From a distributional point of view there seems to be no good reason for it, so the justification must be sought in efficiency considerations. One argument has already been mentioned, viz. the desire to discourage socially unproductive investments through arbitrage transactions. Another concern has been that high taxation of capital income might lead to capital flight, given the improved conditions for international capital mobility. A higher international mobility of capital might be interpreted as an increase in the elasticity of saving with respect to the rate of return, and this would strengthen the argument for a low tax rate on income from capital; see Atkinson and Sandmo (1980).

Someone who is concerned with equity in taxation has to admit that the dual income tax is a system with limited attractions. Although the principles can be defended from the point of view of efficiency, the major weakness of the system is that the individual burden of the tax is lower, the higher is the share of capital income in the individual’s total income. This may not only be considered unfair from the viewpoint of distributive justice; it also encourages unproductive rent-seeking activities as individuals devote time and expenditure to find ways to convert labour into capital income. It seems unlikely that a reform of the system can be devised that completely eliminates the incentives for such conversion. The question therefore arises whether there exists an alternative tax system that would be better suited for a rational combination of the concerns for efficiency and equity. More precisely: Can we device a tax system that avoids the distortion in the capital market while at the same time taxing the consumption resources of those who derive a substantial portion of their income from capital? To the best of my knowledge, there is only one well-developed system of taxation that achieves this, and this is the expenditure tax that was extensively discussed both at the theoretical and practical level around 1980.6

Let me briefly recapitulate the main features of the expenditure tax. This is a direct tax levied on total income received minus saving. The tax base will therefore be lower than with the income tax when saving is positive and higher when saving is negative. The main arguments in favour of such a tax relate partly to redistributive concerns, partly to efficiency considerations. As regards redistribution, the expenditure tax taxes consumption, which is a

---

6 For practical proposals see Lodin (1976), Meade Committee (1978) and Norman and Sandmo (1985), and for theoretical discussions Atkinson and Sandmo (1980), King (1980) and Christiansen and Sandmo (1981).
better indicator than annual income of the consumer’s standard of living. Just like the conventional income tax, this tax can be made progressive and so help to even out inequalities in the standard of living between consumers. As regards efficiency, it can be shown that under certain conditions the expenditure tax is neutral with respect to capital income, so that there are no tax distortions with regard to saving and portfolio decisions.

The point about the expenditure tax that I wish to concentrate on here is related to a common misunderstanding of this tax system. It is often maintained that the expenditure tax avoids the distortion in the capital market by exempting capital income from taxation. The first part of this statement is true, while the second is false. Let me demonstrate this by means of a simple two-period model of consumer choice, avoiding all analytical details.

The representative consumer maximizes a utility function that has present and future consumption ($C_1$ and $C_2$) as well as labour supply ($H$) as arguments. The consumer works in the first period and lives off his savings in the second period. He maximizes the function $U=U(C_1, C_2, H)$ subject to two alternative sets of budget constraints.

The dual income tax

Let $t_w$ and $t_r$ be the tax rates on wage and interest income. I assume for simplicity of exposition that there is proportional taxation of both types of income, but the conclusions are equally valid for progressive taxation of wage income. With $S$ being the amount of saving in the first period, the budget constraints for the two periods can be written as

$$C_1 + S = wH(1-t_w),$$  \hspace{1cm} (1)

$$C_2 = S[1+r(1-t_r)].$$  \hspace{1cm} (2)

These can be combined to yield the lifetime budget constraint,

$$C_1 + [1+r(1-t_r)]^{-1} C_2 = wH(1-t_w).$$  \hspace{1cm} (3)

One sees immediately that the dual income tax implies two distortions of relative prices, one in the labour market and one in the capital market. Assuming in this context that the economy
is competitive, the wage rate and the rate of interest reflect the marginal productivities of labour and capital. Then the tax system causes the consumer to receive distorted information about the social value of his factor supplies.

The expenditure tax.

With an expenditure tax at rate $\tau$, all saving in the first period is tax exempt, so that the first-period budget constraint is

$$C_1 + S = wH - \tau(wH - S).$$  \hspace{1cm} (4)

In the second period the consumer’s expenditure is equal to his savings plus interest, and this is accordingly the tax base.

$$C_2 = S(1 + r)(1 - \tau).$$  \hspace{1cm} (5)

We can now write the lifetime budget constraint by eliminating $S$ from the last equation. We then obtain

$$C_1 + (1 + r)^{-1}C_2 = wH(1 - \tau).$$  \hspace{1cm} (6)

Comparing this with the corresponding equation (3) for the case of the dual income tax, we see immediately that there is no distortion in the capital market; only the labour market distortion remains. Moreover, it is natural to conclude that the expenditure tax is simply a special case of the dual income tax where the tax on interest income is set equal to zero. But this is a misleading conclusion that mixes up the tax effects on relative prices and the tax bases. In order to see the difference between the two systems it is instructive to look at government tax revenue under the two alternative systems. To do this we imagine a context of overlapping generations where, in each period, two generations are living side by side, and the government is collecting taxes from both of them. Assuming the generations to be of equal size, the tax revenues are as follows:

**Dual income tax:** $t_w wH + t_r rS$.  \hspace{1cm} (7)
Comparing the expressions (7) and (8) we see that the expenditure tax is in fact levied on both sources of income to finance consumption; it is simply not true that the expenditure tax leaves out capital income from the tax base. It is also not correct to say that setting $t_r=0$ under the dual income tax makes it equivalent to the expenditure tax. It is true that under this special assumption the two systems become equivalent in terms of the number and nature of distortions, but they remain different in terms of the bases that they tax, and therefore their distributional implications are different. The dual income tax can only avoid the distortion in the capital market by excluding capital income from the tax base. The expenditure tax achieves non-distortion in the capital market by a neutral taxation of capital income. This is a fundamental difference between the two tax systems. Of the two systems, it is the expenditure tax that can achieve neutrality in the taxation of capital income while at the same time treating both sources of income in the same way for purposes of redistribution.

The comparison between the two systems can be put slightly differently. What matters for efficiency is the marginal tax rate; what matters for redistribution is the average rate. In the dual income tax both average and marginal tax rates are in general positive. Under the expenditure tax the average and marginal taxes on labour income are positive, but for capital income it is only the average tax which is positive; the marginal tax rate is zero\(^7\).

My discussion may have given the impression that neutral taxation of capital - in the sense of the marginal tax rate being zero - is obviously desirable from an efficiency point of view. As shown e.g. by Atkinson and Sandmo (1980) this is in general not correct. As second best tax theory tells us, where there are more than one tax distortion, there is no guarantee that removing one of them will get us closer to the social optimum. In moving from an income to an expenditure tax system, the distortion of the labour supply decision remains. Because of this it may be desirable to preserve the distortion of the interest rate, for a lower net of tax interest rate may stimulate labour supply and thus counteract some of the reduction of labour supply that is caused by the marginal tax rate on labour earnings. But it should be noted that

\(^7\)This conclusion must be modified for an expenditure tax which is progressive in the sense that the marginal tax rates are increasing. If the amount of expenditure is increasing over time, the marginal tax rate in period 2 will be higher than in period 1, and this implies a tax distortion of the same nature as the tax on interest income. But for taxpayers whose expenditure decreases with time, the expenditure tax acts like a subsidy to saving. For the average taxpayer the neutrality assumption may not, therefore, be very misleading.
our article, in concentrating solely on the total amount of saving, left out some complications; in particular, it did not take account of the difficulties involved in taxing all types of capital income at the same effective rate. This is an important aspect of what the expenditure tax achieves and strengthens the case for it as a means of achieving a better taxation of capital income in a redistributive context. Also, we did not discuss the intra-generational distributional case for an expenditure tax, concentrating instead on the issue of justice between generations.

Concluding remarks.

The trend towards increased inequality calls for a deeper understanding of its sources, including changes in the structure of the labour market. It also calls for a reconsideration of the systems of taxation that we use to finance public expenditure while achieving a just and fair distribution of the tax burden between citizens. In my view the time is ripe for a reconsideration of the principles involved in the expenditure tax.

The attention in this paper has been limited to inequality as conventionally understood, and not specifically to poverty. However, in a broader picture of inequality in society it is necessary to take into account the new poverty related to drug addiction, illegal immigration and other social issues that have recently been brought to the foreground of public debate. For these groups of people the issue of the proper tax treatment of capital income is of little interest and relevance. This is a reminder that policies that aim to achieve a more equal distribution of economic resources involve more than tax policy; they also involve issues of social security and social assistance. More generally, the analysis of the redistributive effects of public policies must also take account of the differential impact of public expenditure, both transfers and services, on different economic and social groups.
References.

Aardal, Bernt, ed. (2003), Velgere i villrede. N. W. Damm & Søn, Oslo.


Kleven, Laila and Eiliv Mørk (2004), ”Inntekt og forbruk,” *Samfunnsspeilet* 18, no. 4, 44-50.


