Discussion Papers in Economics

Some results from the theory of optimal taxation and their relevance for increasing progressiveness of Indian tax structure.

Sushama Murty

Discussion Paper 19-03

Centre for International Trade and Development

School of International Studies

Jawaharlal Nehru University

India

*

Some results from the theory of optimal taxation and their relevance for increasing progressiveness of Indian tax structure

May 25, 2019

Abstract

Results in the theoretical literature on optimal taxation are reviewed to study their relevance for the design of income and commodity tax reforms in India. Given the recent steep increase in inequality, there is a need to increase significantly the overall progressiveness of the Indian tax structure. In the absence of a comprehensive income tax base in India, where much of income escapes taxation, this can be done by greater reliance on commodity taxation. GST slabs and GST rates have to be fixed in accordance with the "many person Ramsey rule" for commodity taxation that is based on both equity and efficiency considerations. Currently, there is not much variation in the marginal income tax rates faced by people at the high end of the income distribution, even though there is a lot of variation in their incomes. This implies that the existing income tax design gives nearly the same social welfare weights to both billionaires and salaried people with incomes close to Rs. $10,00,000$. Thus, to combat the growing inequality in India, the progressiveness of income taxation at the tail of the income distribution also needs to be significantly increased, as this will be equivalent to making social preferences more inequality-averse.

JEL Classification Codes: H21, H30
Key words: commodity taxes, income taxes, many person Ramsey rule, equity-efficiency trade-offs, marginal tax rates, progressiveness of tax structure

Some results from the theory of optimal taxation and their relevance for increasing progressiveness of Indian tax structure

1 Introduction.

Inequality in India has been on the rise. According to a recent survey by Oxfam International (Oxfam 2018), there was a steep increase in the wealth of the richest 1% from 58% of the country's total wealth in 2016-17 to 73% in 2017-18. Seventeen new dollar billionaires were added in this period taking the number of dollar billionaires in India to 101. On the other hand, the country's poorest half saw their wealth rise by only 1% percent. This alarming increase in inequality in India calls for major tax reforms that increase the progressiveness of the tax system in India. At the same time, it also calls for increase in targeted social expenditures that benefit the population at the lower end of the income spectrum.

The growth in tax revenue in India is hampered by low tax capacity and tax effort. According to data released by Central Board of Direct Taxes of India, only a miniscule 1% of the population pays taxes. Incomes generated in some major sectors like agriculture are constitutionally exempt from taxation and more than 90% of the population in India is too poor to be taxed, receiving incomes less than Rs. $2,50,000$, the minimum income level that is subject to income tax in India. With a large informal sector and a large class of unsalaried workers in the formal sector whose incomes are not perfectly observable, the incentives to under-report incomes to evade taxes and/or to gain benefits from government's targeted programmes for the poor are very high.

The major overhaul of the commodity tax machinery by way of shifting to the new Goods and Services Tax (GST) recently has also faced design problems. Important questions about the number of tax slabs and the associated levels of tax rates to be implemented have been debated quite often. Commodities have often been juggled in and out of different tax slabs in a somewhat arbitrary fashion.

There is a classic theoretical literature in public economics that studies tax design and rigorously lays out principles for optimal taxation. Results of this literature have greatly influenced the design of real-life tax policies, especially in much of the developed world. In this paper, we review some of the main results in this literature with a view to study their relevance for design of commodity and income tax reforms in India that increase the progressiveness of the overall tax structure. These results are highly pertinent because the central focus of the analyses leading to these results is on promoting equity with least loss in economic efficiency. For an inequality averse society, while the equity objective supports increasing the progressiveness of the tax structure, these analyses show that taxation principles for minimising efficiency losses are by-and-large regressive in nature. The optimal tax structure must strike the right balance between these two conflicting objectives.

A striking result in this literature that has influenced tax policies in many countries is by Atkinson and Stiglitz (1976), which has been interpreted to mean that non-linear income taxation is a far more superior redistributive and revenue generating policy instrument than commodity taxes. Based on this result, many OECD countries today levy only a single basic rate of VAT, sometimes supplemented by a couple of (one or two) additional minor rates. A non-linear income tax coupled with several carefully planned and implemented means-tested and universal benefit programmes that target the not so well-to-do of the society are the major fiscal tools that promote progressiveness of the tax structures in these countries.

We argue that the case of India is different. Given that a major part of the population escapes income taxation in India due to incentives for under-reporting and tax evasion, presence of constitutionally non-taxed sectors, and predominance of income levels that are too low for taxation, the role of income tax in promoting equity and generating revenue for the government is very limited. It has often been felt by policy makers that commodity taxes are inherently regressive. However, employing results from the theory of optimal commodity taxation, we argue that there is significant potential for increasing the progressiveness of the tax structure in India through commodity tax reforms. In particular, redistributive and revenue generation objectives can be promoted only through a system of differentiated commodity taxes, as opposed to a uniform rate of VAT that is popular in the more developed countries. Sound economic criteria based on considerations such as the consumption patterns of the rich and the poor and of people with different marginal propensities of tax expenditures and use of algorithms based on the "many person Ramsey rule" of commodity taxation need to be employed to group commodities into GST tax slabs and to determine the associated GST rates.[1]

Furthermore, a look at the Indian income tax schedule reveals that it does not differentiate enough between people at the higher end of the income spectrum. The marginal tax rates for people with incomes more than Rs. 10 lakhs vary between 30% and 35%. The effective marginal income tax rate taking into account the additional surcharge for people in the top income bracket is 34.5% as compared to most OECD countries, where it varies between 40% to 60%. Given the steep increase in inequality in India and the big number of dollar billionaires, there is a need to significantly differentiate rich from the super rich in India for taxation purposes, as this can be expected to generate more tax revenue and also increase the progressiveness of the tax structure in India. Economic theory shows that not doing so is tantamount to the society valuing an additional unit of income of a multi-billionaire same as an additional unit of income of a salaried employee drawing Rs. 10 lakhs annually, *i.e.,* giving the two the same social welfare weight for taxation purposes.[2]

In Section 2, we distinguish between first-best and second-best policies. While the first-best policies involve no equity-efficiency trade-offs in moving from a non-interventionist equilibrium to a first-best social optimum, they are not implementable given serious information constraints faced by the government, in other words, they are not incentive compatible. The second-best policies such as commodity and income tax are more incentive compatible but achieve equity

[1] The many person Ramsey rule is the characterisation of the optimal commodity tax structure, which is based on both equity and efficiency considerations. See Section 3.2.1.

[2] Assuming a constant elasticity of labour supply.

only at the cost of loss in some economic efficiency. In Section 3, we survey the main results in the literature on optimal commodity taxation, namely, the many person Ramsey rule and second-best production efficiency. The latter result justifies simplified commodity tax structures like VAT and GST that eliminate cascading of taxes by not allowing taxation of transactions between firms in intermediate inputs by implementing a system of input tax credits. We also draw attention to some earlier empirical works which have employed algorithms based on the tax reform methodology and the many person Ramsey rule to compute optimal commodity tax rates for India. Section 4 surveys the results in the literature on optimal income tax, which characterise the marginal income tax rates (the slope of the income tax schedule) based on the social weights attached to welfare of people with different incomes, the disincentive effects on labour supply of an increase in the marginal income tax rate, and the nature of the income distribution (especially at its upper tail). Section 5 surveys the literature of optimal mixed taxation, where both income and commodity taxes are simultaneously studied. It presents the main result by Atkinson and Stiglitz (1976) on the redundancy of commodity taxation in the presence of non-linear income taxation. Section 6 focusses on the applicability of this result to the context of India given stylised facts about inequality and the low tax capacity and tax effort in India. Section 7 concludes.

2 First-best versus second-best policies and equity-efficiency trade-offs.

While the first fundamental theorem of welfare formalises Adam Smith's conjecture on the non-wastefulness (Pareto efficiency) of a non-interventionist equilibrium of a private ownership economy with a complete set of perfectly competitive markets, it is silent about the social desirability of such an outcome. It is possible that, in this economy, the initial distribution of wealth is so highly skewed that it results in an equilibrium that is considered inequitable and hence undesirable by the society.

Acknowledging that there will usually be several ways to non-wastefully allocate the scarce resources of an economy to various end uses, the second fundamental theorem of welfare economics states that any such Pareto efficient allocation can be decentralised as a competitive equilibrium of a private ownership economy provided the government can interfere and implement a system of personalised lump-sum taxes and transfers. In particular, with unrestricted power to redistribute resources through use of such policy instruments, the government can, starting from the Pareto efficient but generally inequitable non-interventionist competitive equilibrium, decentralise that Pareto efficient allocation, which is considered just and desirable by the society. Thus, under the assumptions of the second welfare theorem, the movement from the non-interventionist competitive equilibrium to the desired Pareto-efficient allocation involves no trade-off between the twin objectives of promoting equity and efficiency.

These two fundamental theorems of social welfare provide the foundations of modern public economics. Much of public economics is concerned with the consequences of deviating from the ideal conditions under which these results operate. In particular, we will focus on the case where the government's power to do unrestricted lump-sum taxes and transfers is seriously inhibited by informational constraints such as lack of knowledge of true wealth positions/ abilities to pay of private individuals. If the rich know that moving to the social optimum will require heavy taxation of their wealth, then they will refrain from revealing their true wealth status to the government in order to avoid paying high lump-sum taxes. Similarly, if the poor know that transfers to them are contingent on their wealth status, then they too have the incentive to understate their wealth to benefit from higher transfers. Thus, in the presence of informational constraints, first-best policies such as personalised lump-sum taxes and transfers are not "incentive compatible."

Policies that take into account such information constraints, in addition to the usual resource constraints, while raising government revenue in an equitable and least distortionary manner will be at most second-best in nature. This is because the additional constraints restrict the set of allocations that can be attained, so that the social optima attained by such policies will be inferior and Pareto inefficient relative to those attained by the first-best policies. Nevertheless, in general, they tend to be ranked socially higher than the non-interventionist, inequitable but Pareto efficient competitive equilibrium allocation. Thus, choosing a second-best optimal allocation over a competitive equilibrium allocation involves promotion of the equity objective at the cost of loss in Pareto efficiency.

3 Incentive compatibility of commodity taxation and the equity-efficiency trade-off.

As opposed to personalised lump-sum taxes, which are imposed on private wealth and earning levels of individuals that are usually imperfectly observable by the government, commodity taxes are imposed on observable transactions in commodities between sellers and buyers. Hence, in contrast to the former policy instruments, the latter are incentive compatible. However, commodity taxes provide only a second-best means of raising governmental revenue and redistribution. This is because a commodity tax generates revenue for the government by driving a wedge between the price paid by the buyers and price received by the sellers. Thus, in competitive situations, it drives wedges between the marginal rates of substitutions of the buyers and the sellers, leading to a violation of the set of conditions that characterise first-best Pareto optimality.[3]

The loss in efficiency due to commodity taxation as opposed to a lump-sum tax to raise governmental revenue is called the dead-weight loss.[4] In a world with several commodities, all

[3]The conditions include (i) consumption efficiency –equalisation of the marginal rates of substitution between goods in consumption across all consumers, (ii) production efficiency –equalisation of the marginal rates of substitution between goods in production across all producers, and (iii) joint consumption and production efficiency – equalisation of the marginal rates of substitution between goods in consumption and production.

[4]In particular, in a partial equilibrium framework with one taxable good and an untaxed numeraire good,

of which could potentially be taxed to raise government revenue, the main intuition behind a system of commodity taxation that minimises dead-weight loss can be traced back to Pigou [1947] and has been dubbed as the "inverse elasticity rule for commodity taxation,"
Pigou [1947]: *"the best way of raising a given revenue ... is by a system of taxes, under which the rates become progressively higher as we pass from uses of very elastic demand and supply to uses where demand and supply are progressively less elastic."*

Thus, the inverse elasticity rule says that, if the objective of the government is to raise tax revenue with minimum efficiency losses, then goods with more inelastic demands should have higher tax rates. This minimises the price distortions involved in raising government revenue through commodity taxation.

The question of raising governmental revenue from commodity taxation with minimum loss in efficiency was first posed by Pigou to Ramsey. Ramsey (1927) studied the problem in abstraction from equity considerations. The general economy he studied has only a single consumer consuming several commodities. His results were resurrected and made more accessible to readers by Diamond and Mirrlees (1971a,b) (DM), who used the tools offered by modern duality theory in microeconomics to study commodity taxation not only as a means of raising governmental revenue but also a means of redistribution.

3.1 Defining a general equilibrium with commodity taxation and formulating the the second-best Ramsey problem.

Consider a many-person Ramsey economy with the following features:

- Private owned resources with perfectly competitive markets
- N commodities, one aggregated private firm, and H heterogenous consumers[5]
- Profit maximizing producer
- Either a constant returns to scale technology or a decreasing returns to scale technology with 100% profit taxation[6]
- Along with commodity taxation, the government can implement at most a uniform lump-sum transfer, which in the analysis below will be fixed equal to zero
- The societal preferences are represented by a Pareto inclusive and individualistic social welfare function (social welfare is increasing in utilities of consumers)

the imposition of a positive commodity tax to raise a given amount of governmental revenue simultaneously raises the consumer price and reduces the producer price, thereby restricting the demand and supply of the commodity below its socially optimal level. Under a lump-sum tax, however, the commodity continues to be produced at its socially optimal level. Tax is paid entirely out of the endowment of the numeraire commodity. (Recall that partial equilibrium analyses are conducted employing a quasi-linear preference structure, where the good under study is not subject to income effects. See, *e.g.,* Mas-collel et al (1995).)

[5]We will relax the assumption of a single aggregated firm in Section 3.2.2.

[6]The assumption of 100% profit taxation is relaxed in works such as Dasgupta and Stiglitz (1972) and Murty (2013).

The consumers and producer face different price vectors. The consumer price vector is denoted by $q \in \mathbf{R}_+^N$, while the producer price vector is denoted by $p \in \mathbf{R}_+^N$. The implied tax vector is $t = q - p$. The lump-sum transfer to the h^{th} consumer is denoted by $r^h \in \mathbf{R}$. The Marshallian demand vector of the h^{th} consumer net of his endowment is $\hat{x}^h(q, r^h) \in \mathbf{R}^N$, while the producer's net supply vector is $y(p) \in \mathbf{R}^N$ when the technology exhibits decreasing returns.[7] The indirect utility function of the h^{th} consumer is $V^h(q, r^h)$. The social welfare function is $W\left(u^1, \ldots, u^H\right)$, where u^h is the utility level of the h^{th} consumer.

In the literature, there are two main ways of formulating a general equilibrium model of commodity taxation. The first models market clearing in all commodity markets (which is defined as a tax equilibrium) without explicitly acknowledging the government budget balance,[8] while the second models a situation where the government wants to employ its taxation power to raise a fixed amount of revenue without explicitly modelling equilibrium in all commodity markets.[9]

3.1.1 Tax equilibrium and second-best welfare maximisation with decreasing returns to scale technology.

A tax equilibrium of an economy is a configuration of commodity taxes and a uniform lump-sum transfer r, such that all markets are in equilibrium when consumers and producers face these taxes.

Definition: A configuration $\langle q, r, p\rangle$ defines a *tax equilibrium (TE) of a many-person Ramsey economy if*

$$\sum_{h=1}^{H} \hat{x}^h(q, r) = y(p). \qquad (TE - MP)$$

The associated tax vector is $t = q - p$ *and the associated tax equilibrium allocation is* $\left\langle \hat{x}^1(q, r),\ \ldots\ ,\ \hat{x}^H(q, r);\ y(p)\right\rangle$.

Invoking the Walras law, it can be shown that the government's budget is balanced at a tax equilibrium: the total expenditure on the uniform lump-sum transfer is equal to the total tax revenue from commodity and profit taxation.

$$Hr = ty(p) + py(p).$$

It is to be noted that, given the homogeneity properties of consumer demands and producer supplies, $(TE - MP)$ is homogeneous of degree zero in $\langle q, r\rangle$ and is homogeneous of degree zero in p. Thus, if $\langle q, r, p\rangle$ is a tax equilibrium, then so is $\langle \mu q, \mu r, \kappa p\rangle$ for any $\mu > 0$ and $\kappa > 0$. Both lead to the same tax equilibrium allocation. This implies that a given tax equilibrium allocation

[7] If the net demand of a good by a consumer is negative, then this is a positive supply of the good to the firm by the consumer. If the net supply of a good by the firm is negative then it is a positive net input demand for the good by the firm. Under constant returns to scale, the supplies of goods in equilibrium is entirely determined by the aggregate consumer demands.

[8] See for example, Guesnerie (1995).

[9] See for example, Murty and Ray (1989), Atkinson and Stiglitz (1976), and Myles (1995).

is consistent with several tax vectors: The set of tax equilibrium vectors supporting the same tax equilibrium allocation consists of vectors such as $\mu q - \kappa p$ for $\mu > 0$ and $\kappa > 0$. Hence, the definition of a tax equilibrium admits two harmless price normalisations. For example the consumer and producer prices of the first commodity can both be fixed: $q_1 = p_1 = 1$, which implies that the first good is untaxed.

The relevant welfare maximisation problem is:

$$\begin{aligned} \max_{q,p,r^1,\ldots,r^H} \quad W(V^1(q,r^1),\ldots,V^H(q,r^H)) \\ \text{subject to} \\ \sum_h \hat{x}^h(q,r^h) = y(p) \\ q_1 = p_1 = 1 \\ r^h = 0 \quad \forall\ h = 1,\ldots,H. \end{aligned} \tag{1}$$

The above is a second-best welfare maximisation problem because the lump-sum transfers to all consumers has been uniformly pegged at zero.

3.1.2 Second-best welfare maximisation while raising a fixed amount of tax revenue.

In this second approach to modelling the second-best commodity tax problem the government wants to raise a fixed amount of tax revenue $\bar{R}$ through commodity taxation. The technology is usually assumed to exhibit constant returns to scale and the underlying production function is linear: $\sum_{i=1}^N p_i y_i = 0$. This implies that in a general equilibrium, the producer price vector has to be proportional to $\langle p_1,\ldots,p_N\rangle$. Profit maximisation under constant returns to scale results in zero profits with the producer supplies being determined in equilibrium by the aggregate consumer demand vector. The homogeneity of the demands and supplies imply that this model also permits two harmless price normalisation, so that one commodity can be left untaxed. So the second-best problem becomes

$$\begin{aligned} \max_{q,p,r^1,\ldots,r^H} \quad W\left(V^1(q,r^1),\ldots,V^H(q,r^H)\right) \\ \text{subject to} \\ \sum_{i=1}^N t_i \sum_h \hat{x}_i^h(q,r^h) - \sum_h r^h \geq \bar{R} \\ q_1 = p_1 = 1 \\ r^h = 0, \quad \forall\ h = 1,\ldots,H, \\ t_i = q_i - p_i\ \forall\ i = 1,\ldots,N. \end{aligned} \tag{2}$$

Again, by invoking the Walras law, it can be shown that all commodity markets are in equilibrium at the second-best optimum.

3.2 Features of the second-best optimum with commodity taxation.

3.2.1 The many and the one person Ramsey rules.

The characterisation of the optimal commodity tax system (the one that solves problems (1) or (2)) was first derived by Diamond and Mirrlees (1971a,b) and is referred as the many-person Ramsey rule (MPRR). It demonstrates in precisely what manner commodity taxation can be employed to serve the redistributive objectives of the government and highlights the presence of equity-efficiency trade-offs while designing optimal commodity taxes.

Note that the solution to the second-best commodity tax problem will almost always involve price distortions (Pareto inefficiency). Optimal commodity tax vector will not be zero unless $r^1, \ldots, r^H$ can be chosen freely, *i.e.,* unless we can implement personalised lump-sum transfers. In problems (1) and (2), however, r^h is required to be a constant (here equal to zero) for all consumers $h = 1, \ldots, H$. Hence, the optimal commodity tax vector will not be zero. We know that a non-zero commodity tax vector is distortive as it implies that the third condition of Pareto optimality, namely, joint consumption and production efficiency, is violated at the second-best.

Note also that the homogeneity feature of the tax equilibrium leaves open the answers to the questions: "Which commodities should be taxed or subsidised and which should not? Which commodities are to be taxed more and which commodities are to be taxed less?" It follows from the discussion in Section 3.1.1 that the answers depend on the price normalisation rules adopted. Any commodity can be chosen to be untaxed and the extent of taxation of other commodities is relative to the commodity chosen for zero taxation.

Hence, as argued in Mirrlees (1976), one should not seek optimal tax formulae from the above second-best problem. Rather, one can hope to identify the extent to which net demands for commodities are discouraged or encouraged by the optimal tax system, as this characterisation of the optimal tax system is invariant to the choice of the untaxed commodity.

The basic feature of a optimal commodity tax structure is that it achieves the objectives of minimising efficiency losses and redistribution by differentially affecting the "compensated" consumptions of different goods: If the tax system t is intensified; in particular, all taxes and subsidies are changed by the same proportion, then demands for commodities are subject to both income and substitution effects. As noted by Mirrlees (1976), the income effects do not explain the way the tax system differentially affects commodity demands and hence are ignored by the many-person Ramsey rule when assessing the effects of optimal commodity taxes. The substitution effects, on the other hand, distinguish how the tax system individually affects (*i.e.,* discourages or encourages) consumer demands at the optimum.
Hence both the single and many-person Ramsey rules are expressed in terms of proportional reductions in compensated demands of various commodities.

DM's one-person Ramsey model of commodity taxation abstracts from equity issues by focussing on an economy with only one consumer. It derives a characterisation of optimal commodity tax structure that minimises dead-weight losses. Precisely, it states that, at a second-best optimum with commodity taxes, equi-proportionate changes in the commodity tax rates result in equi-proportionate discouragement in the compensated demands of all goods. The inverse elasticity rule is derived as a special case when all cross-price effects on demand for any commodity are zero.

The MPRR, on the other hand, shows that commodities should be differentially discouraged when both equity and efficiency considerations matter. It is given by

$$\begin{aligned}
\frac{\sum_h \sum_i \frac{\partial e_k^h}{\partial q_i} t_i}{\sum_h \hat{x}_k^h} &= -1 + \sum_h \gamma^h \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right) + \sum_h \left[\sum_i t_i \frac{\partial \hat{x}_i^h}{\partial r}\right] \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right) \\
&= -1 \quad + \quad A \quad + \quad B \\
&= -1 + \sum_h b^h \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right), \quad \forall\, k = 1, \ldots, N, \qquad (3)
\end{aligned}$$

where e_k^h denote the compensated demand for commodity k by the h^{th} consumer. For any consumer h, γ^h is a ratio of two terms: (i) the social marginal utility of income to consumer h, denoted by $\alpha^h = \frac{\partial W}{\partial V^h} \frac{\partial V^h}{\partial r}$ and (ii) the marginal cost of public funds, denoted by κ. The former measures the social value of an additional unit of income to consumer h. It is a product of the consumer's own valuation of an extra unit of income and the society's valuation of an extra unit of utility received by consumer h. If private marginal utility of income is diminishing in wealth and the society is inequality averse, then one would expect that α^h would be higher (respectively, lower) if consumer h is poor (respectively, rich). On the other hand, κ is the social marginal cost incurred if the government was to raise an extra unit of revenue. Hence, $\gamma^h = \alpha^h/\kappa$ is the social marginal utility of income to consumer h measured in units of government revenue. It is interpreted as the welfare weight given by the society to consumer h. A consumer with a higher value of γ is valued more by the society. He is more deserving of receiving additions to his income if the society has wealth to spare.

The term $\sum_i t_i \frac{\partial \hat{x}_i^h}{\partial r}$ in (3) is consumer h's marginal propensity of tax payments. It captures the responsiveness of tax payments of consumer h to changes in his income. Increases in commodity tax rates reduce real incomes of consumers, which can affect the commodity tax collections from consumers due to adverse behavioural responses to decreases in real incomes.

Term $A = \sum_h \gamma^h \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right)$ is a weighted average of the social marginal utilities of incomes to consumers measured in units of government revenue. The weights are the shares of different consumers in the total demand for good k. A is higher (respectively, lower) if good k is disproportionately consumed by people with high (respectively, low) γ's, *i.e.,* by people who the society values more (respectively, less).

Term $B = \sum_h \left[\sum_i t_i \frac{\partial \hat{x}_i^h}{\partial r}\right] \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right)$ is a weighted average of the marginal propensities of tax payments of different consumers, the weights being again the shares of different consumers in the total demand for good k. B is higher (respectively, lower) if good k is disproportionately consumed by people with high (respectively, low) marginal propensities of tax payments, *i.e.,* by people whose tax payments are more (respectively, less) responsive to changes in incomes.

Finally, the term $b^h = \gamma^h + \sum_i t_i \frac{\partial \hat{x}_i^h}{\partial r}$ is called the net social marginal utility of income to consumer h. It measures the total effects on social welfare (measured in units of government revenue) and on tax payment of a unit increase in income of consumer h.

Having defined all the terms present in (3), we are now ready to interpret the MPRR. The left side of (3) is the proportionate change in the total compensated demand for commodity k induced by a proportionate change in the tax vector at the optimum.[10] When this term is negative (respectively, positive) then the tax system discourages (encourages) commodity k.

The MPRR says that the discouragement of compensated consumption of commodity k is less (the left side of (3) is less negative)

- the higher (more positive) is term A, *i.e.,* the more the consumption of the k^{th} good is concentrated among people with high values of γ^h, *i.e.,* the people who are considered more important by the society. (In other words, discourage less the consumption of commodity k when a large proportion of the tax base for good k lies with people whom society values more)

- the higher (more positive) is term B, *i.e.,* the more the consumption of good k is concentrated among people with high marginal propensities of tax payments, *i.e.,* the people whose tax payments change considerably when incomes change. ((In other words, discourage less the consumption of commodity k when a large proportion of the tax base for good k lies with people whose tax payments will fall a lot when real incomes are reduced by commodity taxation.)

Equivalently, the MPRR says that, at a second-best optimum, the discouragement of compensated consumption of commodity k is less the more the consumption of the good is concentrated among people with low values of b^h, *i.e.,* among people with low net social marginal utilities of income. Atkinson and Stiglitz (1976) give examples where the equity efficiency trade-offs involved in designing commodity taxes are stark. These arise in the case when the cross price effects on compensated demands are zero, *i.e.,* $\frac{\partial e_i^h}{\partial q_k} = 0$ for all $i \neq k$. In this case, the MPRR yields the optimal ad valorem tax rate for any commodity $k = 1, \ldots, N$ as[11]

$$t_k^{adval} = \frac{t_k}{q_k} = \frac{\left[1 - \quad \sum_h b^h \left(\frac{\hat{x}_k^h}{\sum_h \hat{x}_k^h}\right)\right]}{\varepsilon_k} \qquad \left(\varepsilon_k = \sum_h \varepsilon_k^h \frac{\hat{x}_k^h}{\hat{x}_k}\right)$$

The numerator reflects equity considerations, while the denominator reflects efficiency considerations.

The more inelastic is the average compensated demand for good k (denoted by ε_k), the higher is the ad valorem tax rate for this good. Thus, this would seem to imply that efficiency considerations support taxing necessities more and luxuries less. (Basic intuition suggests that demand for necessities will tend to be more inelastic than demand for luxuries.) On the other

[10] See also Mirrlees (1976).

[11] $\varepsilon_k^h = \frac{\partial e_k^h}{q_k} \frac{q_k}{e_k^h}$ is the elasticity of compensated demand for good k with respect to its own price q_k for consumer h.

hand, the more is the consumption of the good concentrated among people with higher net social marginal utilities of income, the smaller should be the ad valorem tax on good k. Thus, equity considerations would tend to support taxing necessities less and luxuries more. (In an inequality averse society, our basic intuition suggests that demand for necessities will tend to be concentrated among consumers with higher social welfare weights (b^h's), while that for luxuries will tend to be concentrated among consumers with lower social welfare weights.) Atkinson and Stiglitz (1976) give examples where equity considerations dominate efficiency considerations and also examples where the reverse is true.

3.2.2 Second-best production efficiency

One of the remarkable features of the DM optimal commodity tax structure was that, despite its second-best nature, it still exhibits production efficiency, which is one of the characteristics of a first-best Pareto optimal allocation.[12]

An allocation of resources is production efficient if the corresponding aggregate production vector (obtained as the sum of production vectors of all individual producers) lies on the frontier of the aggregate technology of the economy, which is obtained by summing up technology sets of all individual producers. In particular, frontier points of the aggregate technology are obtained by summing frontier points of individual technologies with equal marginal rates of substitution in production.[13]

Profit maximising firms equate their marginal rates of substitution to various ratios of prices they face.[14] Hence, if different firms face different producer price vectors, then their marginal rates of substitution will not be equalised when they maximise profits leading to production inefficiency. The wedge between producer price vectors faced by two different firms is a tax on transactions that take place between the two firms. These transactions include buying and selling of intermediate inputs between the two firms. To allow also for taxation of transactions between two firms, define a tax equilibrium with firm-specific producer price vectors $p^1, \ldots, p^J$, where $p^j \in \mathbf{R}^j_+$ is the price vector faced by firm j:

Definition: A *tax equilibrium (TE) with firm-specific prices* is characterized by $\langle q, p^1, \ldots, p^J, r \rangle$ such that

$$\sum_h \hat{x}^h(q, r) = \sum_j y^j(p^j) \qquad (TE - OP2)$$

[12] This provided a powerful counterexample to the very influential claim by Lipsey and Lancaster (1956) that in the presence of a distortion in the economy, which makes attainability of at least one of the first-best Pareto optimal conditions impossible, the remaining Pareto optimal conditions, although still possibly attainable, are no longer desirable at a second-best optimum.

[13] For example,if there were only two goods, with one good being the input, say labour, and the other good being a standard consumption good like apples, then the relevant production marginal rate of substitution is the marginal product of labour. In that case, production efficiency is true if individual firms produce at frontier points of their individual technologies where their marginal products of labour are the same. The aggregate production will then lie of the frontier of the aggregate technology.

[14] For example, at a profit maximising production bundle, marginal rates of technical substitution between inputs are equal to the input price ratios, marginal rates of product transformation between two outputs are equal to the output price ratios, and the marginal product of an input in the production of an output is equal to the ratio of the input and the output price.

This definition of a tax equilibrium allows (i) taxation of transactions between consumers and firms with $q - p^j$ being the vector of tax rates on transactions between consumers and the j^th firm and (ii) taxation of transactions between firms in intermediate inputs with $p^j - p^{j'}$ being the vector of tax rates on transactions between firms j and j'.

The second-best problem that allows for taxation of inter-firm transactions is given by

$$\max_{q,p^1,\ldots,p^J} W\left(V^1(q,r),\ldots,V^H(q,r)\right)$$
$$\text{subject to}$$
$$\sum_h \hat{x}^h(q,r) = \sum_j y^j(p^j). \tag{4}$$

Allowing for appropriate price normalisations and 100% profit taxation, it can be shown that, at a second-best optimum of such an economy, the producer price vectors are equalised, *i.e.,* $p^j = p$ for all firms $j = 1,\ldots,J$. Since all producers face the same price vector at a second-best, it follows from the above discuss that, profit maximisation by individual producers will imply equalisation of the marginal rates of substitution in production across all producers at a second-best, *i.e.,* production efficiency will hold at a second-best optimum.

Thus the structure of optimal commodity taxes does not require intermediate-input taxation (inter-firm transactions in intermediate inputs should not be taxed). The second-best optimal tax structure supports tax systems like VAT/GST (goods and services tax), where only transactions between consumers and producers are taxed. Under the GST in India, this is implemented as follows:

- Both consumers and producers pay retail prices q for goods purchased from firms (firms purchase intermediate inputs).
- But producers receive tax credit $q - p$ for inputs purchased.
- Hence, effectively, producers pay and receive producer prices in p for intermediate inputs purchased from and sold to each other.

The second-best production efficiency result of DM is based on assumptions like constant returns to scale or non-increasing returns plus 100% profit taxation. Dasgupta and Stiglitz (1972) and Murty (2013) generalise these results. In particular, Murty (2013) shows that, if profit incomes are partly distributed to consumers as dividends, then second-best production efficiency is true when the profit taxation power of government is well-aligned with the way profit incomes are distributed in the economy. In general, Murty (2013) characterises dividend distributions under which production efficiency holds when government can tax profits of different groups of firms at different rates. The Dasgupta and Stiglitz (1972) analysis follows a special case Murty (2013) where the government has perfect profit taxation power, *i.e.,* it can do firm-specific profit taxation. The DM case is also shown to be a special case where consumers receive no profit incomes at all.

3.3 Marginal tax reforms and computation of optimal commodity taxes using consumer expenditure data.

The literature on marginal tax reforms, starting from the pioneering articles of Guesneri (1977) and Feldstein (1976), recognises that economies will usually be operating at sub-optimal tax equilibria. Due to social, political, and institutional constraints it might not be possible to implement optimal tax rates in one go. Rather, what may be feasible are small (piecemeal) changes in the tax rates (marginal tax reforms) that, starting from the status-quo tax equilibrium, are welfare improving and equilibrium preserving (consistent with the resources of the economy).

Ahmad and Stern (1984), Murty and Ray (1989), employ the tax reform methodology to compute optimal commodity tax rates for the Indian economy.[15] Here we sketch the methodology employed in Murty and Ray (1989). We first define the marginal social cost of raising a unit of tax revenue by taxing the i^{th} commodity as[16]

$$\lambda_i := \frac{-\frac{\partial W}{\partial t_i}}{\frac{\partial R}{\partial t_i}} = \frac{\sum_h \alpha_h q_i \hat{x}_i^h}{q_i \hat{x}_i + \sum_j \frac{t_j}{q_j} \hat{x}_j \varepsilon_{ji}}, \tag{5}$$

where $R = \sum_{i=1}^N t_i \sum_h \hat{x}_i^h(q, r)$ is the commodity tax revenue, $\hat{x}_j = \sum_h \hat{x}_j^h$, $\alpha_h = \frac{\partial W}{\partial u^h} \frac{\partial V^h}{\partial r}$, and $\varepsilon_{ji} = \frac{q_i}{\hat{x}_j} \frac{\partial \hat{x}_j}{\partial q_i}$. Thus, λ_i is the change in social welfare due to a marginal change in i^{th} commodity tax rate per unit change in government revenue due to a marginal change in i^{th} commodity tax rate.

It can be shown that, at a solution to the second-best welfare maximisation problem (2), the marginal social cost of raising a unit of tax revenue by taxing the i^{th} commodity are equalised across all commodities: *i.e.,* at a second-best optimum

$$\lambda_i = \lambda, \quad \forall \ i = 1, \dots, N.$$

Hence, if λ_i's are not equalised across all commodities $i = 1, \dots, N$ at the current tax equilibrium (status-quo), then the status-quo is not a second-best optimum. Starting from a sub-optimal Murty and Ray (1989) provide an algorithm to recursively reform the tax system in a revenue neutral way till all λ_i's converge to a constant. Every round of reform involves revising up (respectively, down) the tax rate of any commodity i if the marginal social cost of raising a unit of tax revenue by taxing this commodity is lower (respectively, higher) than the average marginal social cost of raising a unit of tax revenue.

Data set on nine-commodity disaggregation of consumer expenditure from time series of NSS in India from the 7th and 28th round covering time periods 1953/54–1973/74 were employed to first estimate demand systems for two specifications of preferences in terms of indirect

[15] See also Myles 1995.

[16] The first equality is the definition of λ_i, while the second equality is derived from computing the numerator and denominator of $-\frac{\partial W}{\partial t_i} \Big/ \frac{\partial R}{\partial t_i}$.

utility functions, namely, the linear expenditure system and the restricted non-linear preference structure. The welfare weights α^h for $h = 1, \ldots, H$ in (5) are sensitive to specification of the social welfare function. An iso-elastic social welfare function was employed and analysis was carried out for two different values of inequality aversion.

4 Non-linear income taxation.

The seminal Mirrlees (1971) article is based on the premise that due to information constraints the government cannot observe the income-earning potentials of people and can at best tax observable variables that bear some correlation to them. In particular, Mirrlees (1971) studies the rich class of non-linear income tax schedules, under the assumption that incomes of people are perfectly observable and can be taxed. To be incentive compatible, such a tax policy has to influence consumer budgets in an anonymous as opposed to a personalised way, *i.e.,* the restrictions placed by such a tax policy on a consumer's budget opportunities should be the same for all consumers.

4.1 Basic Mirrlees (1971) model.

He considers a model where there are inherent differences in the income earning potentials of people. Productivity of one physical unit of labour time supplied differs among individuals. The model distinguishes between high and low productive individuals. Denote the ability of a consumer by n. The distribution of ability is given by the cumulative distribution $F(n)$, while the density function is $f(n) = F'(n)$. There is only one consumption good. Preferences of all consumers in the labour (l) and consumption (c) space are assumed to be identical and represented by the utility function $\bar{U}(c, l)$

The physical units of labour l supplied by a consumer are converted into effective units of labour $y = nl$ depending on the productivity of the consumer. If a consumer is more able, so that n takes a high value for him, then a given amount of physical units of labour supplied by him is converted into higher effective units of labour than if the consumer was less able (value of n is low). It is effective units of labour supplied by consumers that are converted into meaningful output in the production process. Firms pay for the effective units of labour consumers supply. Hence, if consumers supplied the same amount of physical units of labour, then those with higher productivity receive more remuneration as compared to those with lower productivity. Hence, the productive ability of a consumer can be interpreted as the implicit wage rate of the physical units of labour l supplied by him.

Preferences in the space of consumption and effective units of labour can be derived from the preferences defined in the space of consumption and physical units of labour:

$$U(c, y, n) = \bar{U}\left(c, \frac{y}{n}\right),$$

where $\frac{y}{n}$ is the physical units of labour supplied by an n man when he supplies y effective units of labour. The preferences in the space of consumption and effective units of labour will not be identical across all consumers and will vary depending on the productive ability of the consumers.

The production function is linear: $C = \frac{1}{\alpha}Y$, where C is the output of the consumption good and Y is the effective units of labour employed as the input. $\frac{1}{\alpha}$ is the marginal product of the effective units of labour. The linear nature of the production function makes the equilibrium producer price vector $\langle p_Y, p_c\rangle$ proportional to $\left\langle \frac{1}{\alpha}, 1\right\rangle$. We adopt a normalisation whereby the producer price of effective units of labour (the wage firms pay for effective units of labour input) is one ($p_Y = 1$). This implies that the producer price of the consumption good is fixed at $p_c = \alpha$. With this normalisation, the effective units of labour supplied by a consumer is also his gross (pretax) labour income.

An allocation in this economy is a profile of consumption bundles given by $a = \{\langle c(n), y(n)\rangle\}_{n=0}^{\infty}$, where $\langle c(n), y(n)\rangle$ is the bundle of consumer of type n. Suppose the government wants to raise a fixed amount of government revenue equal to $\bar{R}$ through taxation. Then the total consumption, including government consumption, and total effective units of labour supply under allocation a are, respectively, given by

$$C = \int_0^{\infty} c(n)f(n)dn + \bar{R}, \qquad Y = \int_0^{\infty} y(n)f(n)dn$$

Allocation a is a feasible allocation if it satisfies the technological constraint:

$$\int_0^{\infty} c(n)f(n)dn + \bar{R} = \frac{1}{\alpha}\int_0^{\infty} y(n)f(n)dn.$$

4.2 Non-interventionist competitive equilibrium and comparison of first and second-best policies.

If there was no government intervention in this economy and competitive forces prevailed, then at a market equilibrium it can be shown that more productive consumers will be much better off as compared to the low productive workers. In particular, more productive consumers will also consume more and it is possible that they may be working less. Under assumption of diminishing marginal utility of consumption, this implies the marginal utilities of consumption of highly productive workers will be lower than those of the low productive workers.

4.2.1 First-best policies and incentive compatibility.

If society is inequality averse then, starting from a market equilibrium, gains in social welfare can be achieved through redistribution. For example, if social welfare is sum of individual utilities of people, then it will increase if consumption by the rich (whose marginal valuation of consumption is low) could be reduced and consumption by the poor (whose marginal valuation of consumption is high) could be increased. The ideal (first-best) government policy instruments that can achieve this redistribution are personalised *lump-sum* taxes and transfers. They would involve taxing away some lump-sum incomes from the high-type rich and using the tax revenue to finance transfers to the low-type poor. These first-best instruments achieve redistribution with no loss in Pareto efficiency: marginal rates of substitution between goods in both consumption and production are equalised.

This first-best solution to the problem of maximising social welfare is however not feasible in real life. Due to informational constraints the government cannot distinguish between consumers of different abilities. Only the consumers know their true abilities. Since first-best redistribution requires lump-sum taxation of the highly able types, these types of consumers will deny that they have high ability. Since the government cannot verify the true abilities of people, redistribution through first-best instruments is not "incentive compatible:" High ability types have the incentive to lie and "mimic" people of low ability to avoid paying taxes.

4.2.2 Non-linear income tax as an incentive compatible second-best policy.

Although the government cannot observe the abilities of consumers, Mirrlees (1971) assumes that the government can observe how much people are earning (*i.e.,* it can observe their gross incomes or effective units of labour supplied). It considers taxing the observable wage incomes of people in lieu of their unobservable abilities. The government announces a "common" tax schedule $\bar{T}(y)$ for all people, which specifies the tax to be paid at every level of gross income. The people then have to choose – *self-select* – how much they would like to work and earn. The income tax schedule, which could potentially be highly non-linear, defines a common budget set for all consumers:

$$c \leq \frac{y - \bar{T}(y)}{\alpha} =: C(y) \tag{6}$$

A consumer with ability n maximises utility subject to this common budget set:

$$\max_{c,y} \{U(c, y, n) \mid c \leq C(y)\}.$$

For every ability type n, the first order condition of this problem implies equalisation of the marginal rate of substitution in consumption (denoted by $MRS_{y,c}$) to the slope of the consumption schedule (when it is differentiable):[17]

$$MRS_{y,c}\left(c(n), y(n)\right) = C'\left(y(n)\right) \tag{7}$$

Suppose the resulting allocation is $a = \{\langle c(n), y(n)\rangle\}_{n=0}^{\infty}$. By a revealed preference argument it follows that, at this allocation, their is no welfare gain for a consumer of ability n from

[17] If the tax schedule is kinked, then the consumption schedule is also kinked. It is possible that, in such a situation, consumers of multiple ability types choose the same optimal consumption bundle. This phenomenon is called bunching in the income tax literature.

mimicking any other consumer of a different ability, say n'.[18] Hence, such a non-linear income tax is an incentive compatible tax policy instrument.

However, the consequence of using such an income tax is the loss in first-best Pareto efficiency. The marginal tax rate (MTR), the slope of the income tax schedule, is the addition to tax revenue due to an extra rupee of income earned. Since (6) implies that

$$\bar{T}(y) = y - \alpha C(y),$$

if $\bar{T}$ is differentiable then, employing (7), the MTR for consumer of ability n is given by

$$MTR(y(n)) = \bar{T}'(y(n)) \;\; = \;\; 1 - \alpha C'(y(n)) \;\; = \;\; 1 - \alpha MRS_{y,c}(c(n), y(n)).$$

Measured in units of the consumption good, the MTR is given by

$$\frac{MTR(y(n))}{\alpha} = \frac{1}{\alpha} - MRS_{y,c}(c(n), y(n)).$$

Clearly, the MTR (in units of the consumption good) is the wedge between the the marginal rate of substitution in production as defined by the marginal product of efficiency unit of labour $\frac{1}{\alpha}$ and the marginal rates of substitution in consumption denoted by $MRS_{y,c}$. While non-linear income taxation results in an incentive compatible allocation, this will generally not be Pareto efficient, as the marginal tax rates of consumers of different abilities may not all be zero. There may be consumers for whom the MTR is non-zero.

A source of loss in Pareto efficiency due to non-linear income taxation is the disincentive effect on labour supply that it induces. A positive MTR implies the gross wage paid by the producers is higher than the net wage received by the consumers. This could potentially reduce the incentive to work depending on the relative magnitudes of the substitution and income effects induced by the tax. If leisure is assumed to be a normal good, then the substitution and the income effects on labour supply will work in opposite directions, and a decrease in wage rate due to the income tax leads to a reduction in labour supply if the former effect dominates the latter. When this is true, the magnitude of the disincentive effect of wage tax depends on the responsiveness (elasticity) of labour supply to changes in the wage rate. Intuitively, efficiency losses due to income taxation will be minimised when the wage tax (MTR) is higher for people whose labour supply is more inelastic (*i.e.,* less responsive to decreases in wages).

[18] Utility maximising bundles $\langle c(n), y(n)\rangle$ and $\langle c(n'), y(n')\rangle$ for consumers of abilities n and n' are both points in the budget set (6), but since consumer of ability n chooses the former when the latter is also available, it must be the case that

$$U\left(c(n), y(n), n\right) \geq U\left(c(n'), y(n'), n\right), \;\forall\; n, n'.$$

4.3 Second-best optimal non-linear income tax schedule.

Assuming a social welfare function of the form $W = \int_0^\infty G(u(n))f(n)dn$,[19] the second-best welfare maximisation problem with non-linear income taxation boils down to choosing an income tax schedule $\bar{T} : \mathbf{R}_+ \longrightarrow \mathbf{R}_+$ with image $\bar{T}(y)$ and an allocation $\{y(n), c(n)\}_{n=0}^\infty$ to solve:

$$\begin{aligned} \max \int_0^\infty G\left(U(c(n), y(n), n)\right) f(n)dn \\ \text{subject to} \\ \int_0^\infty \bar{T}(y(n))f(n)dn \leq \bar{R} \\ \text{and } \forall\ n \in [0, \infty),\ \langle y(n), c(n)\rangle \text{ solves} \\ \max_{y,c}\ \left\{U(c, y, n) \mid c \leq \frac{y - \bar{T}(y)}{\alpha}\right\} \end{aligned} \tag{8}$$

This problem was first posed and solved by Mirrlees (1971). He demonstrated that the constraint set by utility maximisation in problem (8) can be replaced by a differential equation that captures the requirement that the optimal tax policy should be incentive compatible. Once this replacement is made, Mirrlees demonstrated that the problem can be readily solved by using known tools in the mathematical theory of optimal control.

The main features of the second-best optimal non-linear income tax schedule derived as a solution to problem (8) include:

- It results in an incentive compatible allocation where welfare, consumption, and gross income are non-decreasing in ability. Thus, while better endowed people are better off at the second-best optimum, they also supply more effective units of labour so that the size of the GDP is enhanced and more resources are potentially available for redistribution.

- Production efficiency holds at the second-best, so economy produces on the frontier of the aggregate technology.

- It supports income subsidies (positive transfer payments/negative levels of tax payments) to people with very low ability people. But these transfers/income subsidies are gradually phased out as income level increase.

- MTRs are nonnegative, in particular, they are bounded between 0 and 1. Thus, it is not possible to tax away more than 100% of an increment in income.

- MTRs are positive for all consumers who work at the optimum, except for the consumers with the highest ability in the economy. This shows that the non-linear income tax is distortive and is truly a second-best policy instrument.

- Seade (1977) and Sadka (1977) considered cases where productive ability n is bounded. In these cases, they show that the MTR for the highest and most productive type of

[19] A special case of this social welfare function is the utilitarian social welfare function $W = \int_0^\infty u(n)f(n)dn$, where social welfare is sum of individual utilities.

workers is zero. Removing price distortions and hence removing disincentives for supplying more labour for the most able type is rewarding as consumers with this ability will then contribute greatly to the GDP.

The following equation from Mirrlees (1971) characterises the MTRs for people of different abilities at the second-best optimum.

$$\left(\frac{1}{\alpha} - MRS_{y,c}(c(n), y(n), n\right) = -\frac{l(n)\bar{U}_l\left(c(n), l(n)\right)}{n^2 f(n)} \int_n^\infty \left(\frac{1}{\bar{U}_c(c(m), l(m))} - \frac{G'(u(m))}{\lambda}\right) T_{nm} f(m) dm \quad (9)$$

where the optimal supply of physical units of labour by consumer of ability n is $l(n) = \frac{y(n)}{n}$, $u(m) = \bar{U}(c(m), (l(m))$ is the welfare of consumer of ability m at the second-best optimum, and $T_{nm} = \exp\left[-\int_n^m (l(s)\bar{U}_{cl}(c(s), l(s))/(s\bar{U}_c(c(s), l(s))ds\right]$. If utility function $\bar{U}$ is additively separable, then $T_{nm} = 1$.

The left side of (9) is the MTR faced by consumers of ability n at the second-best optimum. The term $\frac{1}{\bar{U}_c(c(m),l(m))}$ on the right side of (9) captures the increase in government revenue when welfare of consumer of ability m is reduced by one unit by reducing his consumption. λ is the social marginal cost of public funds (social marginal cost of raising an additional unit of government revenue). Hence, the term $\frac{G'(u(m))}{\lambda}$ is the marginal social cost in units of government revenue of reducing welfare of consumer of type m by one unit.

Thus, the term $\left(\frac{1}{\bar{U}_c(c(m),l(m))} - \frac{G'(u(m))}{\lambda}\right)$ is the increase in government revenue net of revenue-equivalent social cost because of reducing consumer of type m's welfare by one unit. Focusing on the case when $\bar{U}$ is additively separable, (9) says that the MTR of consumer of type n depends upon (i) the total net gain in revenue from reducing welfare of every consumer with ability greater than n by one unit, (ii) the total disutility of work as measured by $-l\bar{U}_l$, and (iii) ability n and the number of n type people, $f(n)$.

4.4 Modern reformulation of Mirrlees (1971) characterisation of optimal income tax schedule.

The Mirrlees (1971) characterisation of optimal MTR in (9) is difficult to interpret and not readily useful for empirical application. Employing a methodology based on tax reforms, this characterisation has been presented in a form that is more amenable for interpretation and empirical applications by Saez (2001), Brewer, Saez, and Shephard (2010), and Picketty and Saez (2013).

Suppose cumulative distribution function of income at the second-best optimum is $H(y)$ and the corresponding density function is $h(y) = H'(y)$. For simplification, let us assume that

- $\alpha = 1$, so that the budget constraint is $c = y - \bar{T}(y)$, and c can be interpreted as the disposable income and y as gross (pre-tax) income.

- Preferences are quasi-linear in consumption so that there are no income effects on choice of y.[20]

Let $\bar{y}$ be an arbitrary level of gross income. By employing a marginal tax reform methodology, Saez (2001) shows that the optimal top marginal tax rate for incomes above $\bar{y}$ is given by

$$\frac{\bar{T}'(\bar{y})}{1-\bar{T}'(\bar{y})} = \frac{(1-\bar{g})\left(\frac{y_m}{\bar{y}}-1\right)}{\varepsilon \frac{y_m}{\bar{y}}}, \tag{10}$$

where $\bar{g}$ is the average social marginal utility of consumption in units of government revenue (welfare weight) of top bracket individuals, *i.e.,* of people with income at least equal to $\bar{y}$, ε is the elasticity of labour supply, and y_m is the mean income of people with income more than $\bar{y}$.

The left side of (10) is the MTR at $\bar{y}$ as a proportion of increment in disposable income per unit increase in gross income. The right side shows that $MTR(\bar{y})$ is zero when $\bar{y}$ is equal to y_{max}, which is defined as the highest income level in the economy.

The Pareto distribution characterised by $\bar{y}$ and a is given by

$$H(y) = 1 - \left(\frac{y}{\bar{y}}\right)^{-a} \quad \forall\ y \geq \bar{y},\ a > 1.$$

For Pareto distributions, the ratio $\frac{y_m}{\bar{y}}$ is exactly equal to the constant $\frac{a}{a-1}$, independent of $\bar{y}$. Vilfredo Pareto discovered that for most empirical income distributions, the ratio $\frac{y_m}{\bar{y}}$ is a constant different from one beyond a certain level of income, say $\bar{y} = y^*$ (see Pareto (1965). Thus, tails of most empirical income distributions (*i.e.,* for values of income greater than y^*) behave like a Pareto distribution. Therefore, (10) shows that the top marginal tax rate $MTR(\bar{y})$ is a non-zero constant when $\bar{y}$ is greater than y^* but not equal to y_{max}. Hence, the zero top marginal tax rate is applicable only for the topmost able guy in the economy.

After deriving a expression for the optimal top marginal tax rate, we consider the optimal marginal tax rate at any level of income l y. This is given by

$$\frac{\bar{T}'(y)}{1-\bar{T}'(y)} = \frac{1}{\varepsilon} \cdot \frac{1-H(y)}{y\ h(y)} \cdot (1-g(y)), \tag{11}$$

where $g(y)$ is the average social marginal welfare weight for individuals with incomes above y and ε is a weighted average of effective labour supply elasticities of individuals with income more than y. Thus, (11) shows that the optimal MTR is high at income level y, when $g(y)$ is low. As expected, (11) shows that the optimal MTR is inversely related to the labour supply elasticity ε – the more inelastic is labour supply, the less are the disincentive effects of higher MTRs on labour supply.

When $1-H(y)$ is high then the number of individuals with incomes above y is high. A high MTR at this level of income increases significantly the tax revenue collections from individuals

[20] For more general preference structures see Saez (2001).

with incomes above y. At the same time, for people with income equal to y,[21] a higher MTR at y implies a greater disincentive to work and hence a reduction in tax revenue for the government. This disincentive effect on labour supply can be shown to depend upon $yh(y)$. Hence, optimal MTR is high when $1 - H(y)$ is high relative to $yh(y)$. The fraction $\frac{1-H(y)}{yh(y)}$ is called the hazard ratio. It is constant for a Pareto distribution. For most empirical income distributions, it initially decreases and then increases and converges to a constant value after a certain level of income say y^*.

Saez (2001) and Brewer, Saez, and Shephard (2010) have employed (10) and (11) to compute optimal marginal tax rates for the US and UK, respectively. Saez (2001) finds that for incomes starting from $y^* = \$150,000$ to close to the very top, the ratio $\frac{y_m}{\bar{y}}$ is roughly a constant around 2 for the US. For the UK, the hazard ratio is very high at the bottom of the income distribution and falls as income increases and then rises slightly until it becomes flat around 0.6. This implies a value of the Pareto parameter a equal to 1.67. Various estimates of the elasticity parameter ε in (11) along with this value of a yield various estimates of the optimal top MTR (for the top 1% in the income distribution) for the UK. For example, with an estimate of $\varepsilon = 0.46$, and for a Rawlsian social welfare function, the optimal top MTR is 56.6%, which is close to the actual top MTR equal to 52.7%. If values of elasticities are taken around 1 standard deviation of 0.46, the optimal top MTRs lie in a range, 50.4% to 64.5%. The MTRs at all income levels have also been simulated for different estimates of elasticities and specifications of the welfare weights.

5 Optimal tax-mix: Case for uniform rate of commodity taxation under non-linear income taxation.

Here we consider general tax schedules that allow taxation of both income and commodities, *i.e.,* tax is a function both of the amounts of commodities purchased and of income earned: $T(y, x_1, \ldots, x_N)$. This formulation allows non-linear taxation of both commodities and income. A special case, which is commonly observed in real life, is one where linear commodity taxes are combined with non-linear income taxation:

$$T(y, x_1, \ldots, x_N) = \sum_{i=1}^{N} t_i x_i + \bar{T}(y).$$

An important theoretical result in the literature on mixed taxation that has greatly influenced design of real-life income and commodity tax policies is by Atkinson and Stiglitz (1976).

Theorem 1 *(Atkinson and Stiglitz (1976)) When non-linear income taxation is also an instrument available to the government, then employing (a highly differentiated system of) commodity taxation is redundant when consumer utility functions are weakly separable between labour and all consumption goods taken together.*

[21] Intuitively, their number is given by the density $h(y)$.

Under the assumptions of the theorem, its conclusion implies that government's redistributive objective and revenue requirements are best met with least loss in efficiency by non-linear income taxation supplemented by a uniform rate of commodity taxation. This is in contrast to the MPRR for optimal commodity taxation, where commodity tax rates will generally vary across different commodities.

Although limited in its application to only separable preferences, this theoretical result by Atkinson and Stiglitz (1976) has been invoked several times in the literature and by policy makers to argue that commodity taxation provides only an indirect means of tackling inequality and that redistributive goals can be more effectively achieved by measures that directly tax incomes of people from various sources, while also simultaneously conferring benefits in the form of transfers and tax credits to the really needy groups in the society.[22] These theoretical findings on optimal tax mix greatly simplify the design and implementation of commodity and income tax policies providing relief to tax administration. Enforcing a finely differentiated system of commodity taxation along the lines of the MPRR with a redistributive motive in mind involves huge administrative and monitoring costs and can be expected to result in reduced levels of compliance.

Thus, many OECD countries tend to adopt only a single or at most a few VAT rates. The EU does not permit most member states to levy more than three VAT rates – a standard rate and two reduced rates, where the latter can include zero rating of commodities such as food items and children's clothing. The Mirrlees Review, in its chapter on VAT by Crawford, Keen, and Smith (2010) recommends movement to a uniform rate of VAT for UK. These authors study the impact of a reform package for UK that replaces the existing differentiated VAT structure (which includes a standard rate of 17.5% and reduced rates of 5% and 0%) for a 17.5% uniform rate of VAT, whose effect on equity is neutralised by a 15% increase in all income support, income-based jobseeker's allowance and tax credit rates, and in the associated housing benefit and council tax benefit thresholds. They argue that, while the switch to a single rate of VAT results in financial losses for all income groups, when the compensating measures for the targeted groups are taken into account, households in the lowest income group gain on average £2.50 per week, whereas those at the top lose around £25 per week. The package leads to a net increase in tax revenue to the government, as the VAT reform raises much more than the cost of the compensating package.

These arguments thus lay the onus of achieving redistributive goals mainly on direct tax/ transfer measures, leaving VAT and other forms of commodity taxation as instruments that primarily contribute to the government's tax revenue. Focus then shifts away from the study of redistributive impacts of commodity taxes to the study of their revenue impacts, *e.g.,* to the study of whether introduction of the VAT leads to generation of higher tax revenues and higher tax-GDP ratio (*i.e.,* to questions such as: "Is VAT a money machine?")[23] However, even as a tool of tax revenue collection, in many developed countries, the role of VAT is secondary to

[22] See Crawford, Keen, and Smith (2010).

[23] See for instance Keen and Lockwood (2006, 2010).

instruments of direct taxation. For example, the average share of revenue from taxing income, corporate profits, and capital gains in the total tax revenue in OECD countries was 34% in 2014/15 (with 49% in USA, 48% in Canada, over 63% in Denmark, and over 40% in most other Scandinavian and Western European countries). While, the average share of revenue from taxing goods and services in OECD countries was 32% (with 17% in USA, 23% in Canada, and ranging from 20 to 33% in many Western European and Scandinavian countries.)[24]

6 Case for highly differentiated commodity taxation in India and increasing progressiveness of the income tax at the top.

6.1 Incentive compatibility of income taxation.

The results discussed above in the theoretical literature on income taxation and optimal mixed taxation rested on two key assumptions, namely, (i) income levels of people are perfectly observable by the government and (ii) incomes generated in all sectors of the economy are taxed.

However, as is evident in the real world, the government cannot perfectly observe incomes of people. Tax evasion and build-up of black money is a ubiquitous problem faced by both developed and developing economies. There is a tendency to misreport (usually under-report) true income to reduce tax liability and/or claim higher means tested government transfers. Thus, in reality, non-linear income tax too fails to be a perfectly incentive compatible tax policy. Secondly, incomes generated in some sectors of the economy may be constitutionally exempt from income taxation.

These two issues have significant relevance for understanding the trends in income taxation in India.

6.2 Tiny share of income tax payers in India.

The data released by Central Board of Direct Taxes (CBDT) of India reveals that only a small percent of the Indian population pays income taxes: In the assessment year 2015-16, only 1.7% of Indians filed income tax returns, and 93.3% of these declared incomes between Rs. 1-2.5 lakhs and so were exempted from paying any tax. Thus, the actual percentage of people paying income tax in India is very small.[25]

This could be attributed to both low tax capacity and low tax effort in India. Tax capacity has been defined as “the ability of a government to raise tax revenues based on structural factors including the level of economic development, the number of 'tax handles' available, and

[24] See *https://stats.oecd.org*, Comparative Tables of the OECD Revenue Statistics.

[25] See *https://www.incometaxindia.gov.in/Pages/Direct-Taxes-Data.aspx* and *https://timesofindia.indiatimes.com/business/india-business/toi-budget-2018-special-income-tax-payers-in-india-are-minuscule/articleshow/62538779.cms*.

the ability of the population to pay taxes," see Chelliah (1971). Tax effort, on the other hand, is "a measure of how well a country is using its taxable capacity, that is, tax effort is the ratio of actual tax revenue to taxable capacity[26]" see Bahl (1971).[27] According to Oxfam Annual Report for India (2016-17), India performs poorly on both counts. Its tax effort is 16.7% of the GDP, which is around 53% of its tax capacity.

According to OECD (2016), raising tax revenue poses many challenges in developing countries due to factors such as weak tax administrations, low taxpayer morale and compliance, corrupt and poor governance, prevalence of hard-to-tax sectors, a small tax base, and missing reciprocal link between tax and public and social expenditures. These factors contribute to tax evasion and prevalence of significant sectors that are untaxed. These combined with predominance of low levels of income reduce tax capacity and tax effort in developing economies.

In India, incomes generated in the huge informal sector by workers such as electricians, contractors, masons, hair-dressers, cleaners, hawkers, caterers, restaurateurs, etc., are not verifiable. It is possible that a significant number of self employed in this sector earn incomes above the maximum permissible level that is untaxed but do not report them. There are also incentives for many workers in this sector to under-report their incomes to claim benefits from targeted welfare programmes of the government. As noted by Hanna and Olken (2018), most developing countries in the world rely on proxy-means tests based on directly verifiable and observable information on household assets or amenities rather than on self-reported incomes to classify and target households that are deserving of benefits. However, in the case of India, these authors note that targeting is implemented without a proxy-means test.

A large proportion, nearly 95% of rural families earned less than Rs. 2.5 lakhs in fiscal year 2014-15 and were exempt from income taxation. There are also several categories of income earners such as farmers who do not have to pay any income tax even if they earn incomes above above the maximum permissible level that is untaxed. Thus, the entire agricultural sector in India, which consists of more than half the population, is out of the tax net. In the urban areas, self-employed professionals and consultants such as doctors and lawyers pay a lower rate of income tax as compared to salaried employees. It is intuitive that under-reporting will be very common in the former group as their earnings will not be directly observable, while compliance will be high in the latter group whose taxes on wages are directly withheld.

6.3 Income inequality in India and income taxation.

A recent survey by Oxfam International (Oxfam (2018)) shows that India witnessed an unprecedented increase in inequality between the years 2016-17 to 2017-18. While in the former year, the richest 1% of the population held 58% of the country's total wealth (which was above the global figure of 50%), in the latter year, the top 1% of the population held 73% of the country's total wealth. The survey also showed that the wealth of India's richest 1% increased by over Rs 20.9 lakh crores during 2017, an amount equivalent to total budget of the central

[26]Both as percentage of GDP

[27]Mukherjee (2017) estimates tax capacities and tax efforts of Indian states.

government in 2017-18. On the other hand, the country's poorest half percent saw their wealth rise by just 1%. According to the Oxfam survey, there were 101 dollar billionaires in India in 2018, of which seventeen were added during the year 2017-18. During this period, wealth of the dollar billionaires increased by over Rs. 4.89 lakh crores to Rs. 20.67 lakh crores (reference economic times summary of Oxfam).

According to the World Inequality Report by Alveredo et al (2018), the share of national income accruing to the top 1% earners touched 22% in 2014, while the share of top 10% was around 56%. The report states that inequality has risen substantially from 1980s onwards, following the profound transformations in the economy that centred on the implementation of deregulation and opening-up reforms.

This rise in inequality is reflected also in the income tax collections in the economy. In the fiscal year 2015-16, 90% of population earned income less than Rs. 2.5 lakhs, the maximum permissible level of income that is untaxed. Income tax data of India also indicates that only 0.35% of the tax payers declared incomes above Rs 50 lakhs, while 6.28% declared a taxable income between Rs. 2.5 and 25 lakhs. Around 11% of the total tax payers contribute to 76% of the total tax revenue. Thus, a very small proportion of the tax payers pay disproportionate amount of taxes.

Moreover, even within this class, evasion levels are very high. According to the report by Oxfam (2018), the top 1% of the global earners is evading an estimated $ 200 billions in tax. In particular, developing countries are losing at least $170 billion each year in forgone tax revenues from corporations and the super rich.

6.4 Increasing the progressiveness of tax system in India.

The alarming increase in inequality in India in the recent past that has been discussed in the previous section (Section 6.3), coupled with the low tax effort and capacity, call for a redoubled effort to increase progressiveness of the tax system in India and to improve tax administration in order to garner more tax revenue to meet both redistributive and developmental objectives.[28]

6.4.1 Increasing progressiveness of income taxation at the top of the income distribution.

Given the problems of tax evasion, significant sectors that are constitutionally untouched by income taxation, and predominance of levels of incomes too low to be taxed, there are limits to which income tax can be used as an effective policy tool for redistribution and raising government revenue. Nevertheless, there seems to be still some scope for increasing the progressiveness of income taxation for India at the higher end of the income distribution.

[28]Progressiveness of a tax structure measures the rate at which the average tax rate or the tax burden (the share of tax in total income) increases with increase in income. If the average tax rate (ATR) increases with increase in income then the tax system is progressive. If it decreases (respectively, stays constant) then the tax system is regressive (respectively, proportional).

The design of the current income tax brackets (2019-20) can help understand this scope. The high tax brackets (when we include the surcharges at the rate of 10% of assessed income tax for incomes between Rs. 50 lakhs and Rs. 1 crore and at the rate of 15% of assessed income tax for incomes above Rs. 1 crore) imply that, effectively, the high income tax brackets in India are

Taxable income	**Tax rate**
Above Rs. 10 lakhs to 50 lakhs	30%
Above Rs. 50 lakhs to 1 crore	33%
Above Rs. 1 crore	34.5%

Thus, for the huge range of incomes above Rs. 10 lakhs, there is very little variation in the income tax rates. This range includes not only the salaried people with incomes very close to Rs. 10 lakhs but also the growing set of Indian billionaires (including dollar-billionaires). Recalling the formula for the optimal marginal tax rates in (11) and assuming that the hazard rate and the effective labour supply elasticity ε are nearly constant across all the three high income brackets,[29] this implies that the social welfare weights on people with incomes above Rs. 10 lakhs do not vary much. The social welfare weight given to a billionaire seems not to be so different from the welfare weight of a salaried employee earning Rs. 10 lakhs.

This implies that, under the assumptions made above, increasing progressiveness of the income tax at the upper end of the income distribution is tantamount to a more fine differentiation of social welfare weights attributed to people earning incomes in these high brackets. The topmost marginal tax rates in more egalitarian countries are much higher than the 34.5% in India. In 2018, most OECD countries had top marginal tax rates above 40%, with Sweden topping this list with top marginal tax rate of 60% followed by Denmark, Japan, France, Austria, Greece, and Canada, all of which have top marginal tax rates above 50%. The top marginal tax rates in UK and US were 45% and 43.7%, respectively.[30]

Better estimates of optimal top marginal tax rates and marginal tax rates in general can be obtained by employing formulae given in (10) and (11), when we also have good estimates of effective-labour supply elasticities (responsiveness of gross earnings to changes in the marginal income tax rates) and a knowledge of the tail of the empirical income distribution in India, which will permit computation of the hazard rate for India. Assuming inequality aversion, this information needs to be coupled with an appropriate system of welfare weights, which differentiates incomes of tax payers and attributes lower and lower weights to higher and higher levels of incomes. The rate at which the society discounts welfare of the super rich depends on the extent of society's aversion to inequality at the tail of the income distribution for countries like India, where it is people with incomes in this range that predominantly contribute to income tax revenue. Tax reforms that can increase progressiveness of the income tax at the tail of the income distribution, may require an increase in societal preferences for inequality aversion. It is then that the marginal tax rates on the super rich will be significantly higher than the rest

[29]Recall that, empirically, the hazard rate takes a constant value after a certain level of income.
[30]See $https://stats.oecd.org/Index.aspx?DataSetCode=TABLE\_I7$.

of the population (including the rest of the population in the tail of the income distribution).[31]

6.4.2 Reliance on commodity taxation.

The applicability of the Atkinson and Stiglitz (1976) result on the redundance of commodity taxation when income tax is available as a policy tool is severely undermined when income is imperfectly observable and where there exist sectors in the economy, where incomes are constitutionally non-taxable, as in the case of India. See Sections 6.1 and 6.2.

While it may be difficult to observe individual incomes, transactions in commodities between firms and consumers are more transparent in real life. This makes commodity taxation a more incentive compatible instrument than income taxation. While there is a leakage in the income tax base due to unobservability of income or presence of constitutionally non-taxed sectors, the commodity tax base is more immune to such a leakages. Income that escapes taxation can be taxed from the expenditure side using commodity taxes.

Thus, tax revenue can be collected in the form of commodity tax revenue even from the informal and agricultural sectors where incomes of people are not observable for taxation or are exempt from income taxation in India. Income-tax evasion at the higher end of the income distribution can also be partly neutralised by commodity taxation.

Thus, it is not surprising that economies like India tend to rely more on commodity taxation than the OECD countries. The share of commodity taxes in total tax revenue in India was around 67% in 2014, while that of direct taxes was only 33%.

6.4.3 Achieving progressiveness through commodity taxtion.

Given the Atkinson and Stiglitz (1976) result on the superiority of income tax over commodity taxes, there has been little focus on exploring the potential of commodity taxation for increasing the overall progressiveness of the tax system. It has often been thought that commodity taxes are inherently regressive.[32] Such views tend to focus mainly on the efficiency considerations involved in the design of the commodity tax structure, which are based on the inverse-elasticity concept requiring high rates of taxes on goods with relatively more inelastic demands. As argued in Section 3.2.1, these generally tend to be necessities, which are known for forming a sizeable part of the consumption baskets of the not so wealthy.

However, the MPRR characterises the optimal commodity tax structure based on both equity and efficiency considerations. Assuming an inequality averse society, in principle, progressiveness of the overall tax system can be enhanced through commodity taxation by taxing more and more heavily goods whose consumptions are concentrated among the more and more

[31]This is assuming that the disincentive effects of high rates of taxation are small or not too significantly different at high levels of income. Or, even when there is a variation in the elasticity of earnings of people with respect to marginal tax rate, it is possible that, analogous to examples given in for commodity taxes, equity considerations can offset loss in welfare due to the disincentive effects of higher marginal tax rates. This remains to be verified by further empirical work in this area in the Indian context.

[32]See, for instance, $https://economictimes.indiatimes.com/news/economy/policy/india-cant-have-single-rate-gst-3-slab-structure-possible-arvind-subramanian/articleshow/64952714.cms$.

affluent in the society. Since people in all sectors – formal, informal, constitutionally untaxed – incur expenditure on commodities subject to taxation, designing commodity taxes on the basis of the intuition provided by the MPRR can go a long way in ensuring that the tax burden on consumers increases with increase in income in all sectors. Thus, commodity tax structure can be designed to enhance progressiveness of the tax system.

There is a lot of debate today about both the number of tax slabs to be implemented and the corresponding levels of tax rates to be implemented under the new system of commodity taxation GST in India. When we acknowledge that commodity taxes are significant policy instruments for both redistribution and revenue generation in India (where income taxation has severe limitations), then a lot of thought has to go into designing the tax slabs and the corresponding commodity tax rates to be implemented under the new GST.

Enhancing progressiveness of the tax system through commodity taxation requires a greater degree of differentiation of commodities for taxation purposes than is seen in the more developed economies, where it is common to see a single main VAT.[33] In the absence of compensatory means-tested social-benefit packages along the lines seen in the developed countries, GST in India based on a uniform or a couple of tax slabs with commodities assigned to tax slabs without a sound economic criterion will tend to be very regressive.

In particular, MPRR implies that the differentiation of commodities for commodity tax purposes needs to be based on consumption patterns: Commodities need to be categorised on the basis of the income slab(s) in the economy that are consuming them disproportionately. Luxury goods and necessities can then be defined on the basis of this categorisation. Given these categories, a revenue-neutral algorithm along the line suggested by the Murty and Ray (1989), which is based on the MPRR, can guide the policy makers on the tax rates to be implemented under the new GST. With sufficient inequality aversion in the society, this algorithm is likely to lead to a commodity tax structure that will enhance the progressiveness of the overall tax structure in India.

7 Conclusions.

There is a rich theoretical literature on optimal taxation, which emphasises on focusing on equity and efficiency considerations while designing optimal taxes and tax reforms. A good understanding of this literature is an important pre-requisite for designing tax policies in the real world. The relevance of these results and the way they need to be applied differ from context to context. For example, based on one such theoretical result (Atkinson and Stiglitz (1976)), non-linear income taxation has come to be regarded as a far more superior tax policy instrument than commodity taxation in countries with a comprehensive and more leak proof income tax base.

However, this will not be a suitable tax design for India. The Atkinson and Stiglitz (1976) conclusion on the redundancy of commodity taxation in the presence of non-linear income

[33] See Section 5.

taxation is not true for India where much of income escapes taxation. Hence, in the absence of income tax reforms and reforms in tax administration that increase the income tax base and in the absence of comprehensive means-tested welfare programmes for the poor and needy, the reliance on commodity taxation for meeting both revenue generation and redistributive objectives has to be greater in India. The public economic literature has results that indicate how progressiveness of the tax system can be increased through commodity taxation. Sound economic criteria based on considerations such as the consumption patterns of the rich and the poor and of people with different marginal propensities of tax expenditures combined with use of algorithms based on the many person Ramsey rule of commodity taxation need to be employed to group commodities into GST tax slabs and to determine the associated GST rates.

Two stylised facts about income taxation in India stand out. Firstly, most of the income tax revenue comes from taxing people with high incomes. Secondly, there is not much variation in the marginal tax rates faced by these people, even though there is a lot of variation in incomes at the higher end of the income distribution starting from Rs. 10 lakhs. From the theoretical results in the income tax literature, this implies that the existing income tax design gives nearly the same social welfare weight to all people with incomes above this level, who include both billionaires and salaried people with incomes close to Rs. 10 lakhs. Thus, combatting inequality and increasing governmental tax revenue in India requires also increasing significantly the progressiveness of income taxation at the higher end of the income distribution as this will be equivalent to making social preferences more inequality averse.

References

Ahmad, E., and N. H. Stern, "The Theory of Reform and Indian Indirect Taxes," *Journal of Public Economics* **2**, 1984, 259-98.

Alveredo F., L. Chancel, T. Piketty, E. Saez, and G. Zucman, *World Inequality Report 2018* $https://wir2018.wid.world/$

Atkinson, A. B., and J. Stiglitz, "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics* **6**, 1976, 55-75.

Bahl, R. W., "A Regression Approach to Tax Effort and Tax Ratio Analysis", IMF Staff Papers, 18(3), 1971, 570-612.

Brewer, M., E. Saez, and A. Shephard, "Means Testing and Tax Rates on Earnings," Chapter 2 in J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds) *Dimensions of Tax Design: the Mirrlees Review*, Oxford University Press, 2010.

Chelliah, R. J., "Trends in Taxation in Developing Countries," IMF Staff Papers, 18(2), 1971, 254-331.

Crawford, I., M. Keen, and S. Smith, "Value Added Tax and Excises," Chapter 4 in J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds) *Dimensions of Tax Design: the Mirrlees Review*, Oxford University Press, 2010.

Dasgupta, P. and J. Stiglitz, "On Optimal Taxation and Public Production" *The Review of Economic Studies* **39**, 1972, 87-103.

Diamond, P. and J. Mirrlees (a), "Optimal Taxation and Public Production I: Production Efficiency" *The American Economic Review* **61**, 1971, 8-27.

Diamond, P. and J. Mirrlees (b), "Optimal Taxation and Public Production II: Tax Rules" *The American Economic Review* **61**, 1971, 261-278.

Feldstein, M., "On the Theory of Tax Reform," *Journal of Public Economics* **6**, 1976, 77-104.

Guesnerie, R., "On the Directions of Tax Reforms," *Journal of Public Economics* **7**, 1977, 179-202.

Guesnerie, R., *A Contribution to the Pure Theory of Taxation*, Cambridge, UK: Cambridge University Press, 1995.

Hanna R. and B. A. Olken, "Universal Basic Income Versus Targeted Transfers: Anti-Poverty Programs in Developing Countries," *Journal of Economic Perspective*, **32**, 2018, 201-226

Keen, M. and B. Lockwood, "Is the VAT a Money Machine," *National Tax Journal* **59**, 2006, 905-928.

Keen, M. and B. Lockwood, "The Value Added Tax: Its Causes and Consequences," *Journal of Development Economics* **92**, 2010, 138-151.

Lipsey, R. G. and K. Lancaster, "The General Theory of Second Best," *Review of Economic Studies* **24:**, 1956, 11-32.

Mas-Colell, A., M. D. Whinston, and J. R. Green, *Microeconomic Theory*, Oxford University Press, Oxford, New York, 1995

Mirrlees, J., "An Exploration into the Theory of Optimum Income Taxation," *Review of Economic Studies* **38**, 1971, 175-208.

Mirrlees, J., "Optimal Tax Theory: A Synthesis," *Journal of Public Economics* **6**, 1976, 327-358.

Mukherjee, S., "Changing Tax Capacity and Tax Effort of Indian States in the Era of High Economic 2001-2014" NIPFP Working Paper Series, Working Paper No. 196, 2017.

Murty, M. N. and R. Ray, "A Computational Procedure for Calculating Optimal Commodity Taxes with Illustrative Evidence from Indian Budget Data," *The Scandinavian Journal of Economics* **91**, 1989, 655-670.

Murty, S., "Production Efficiency and Constraints on Profit Taxation and Profit Distribution in Economies with Ramsey Taxation," *Social Choice and Welfare* **41**, 2013, 579-604.

Myles, G. D., *Public Economics*, Cambridge, UK: Cambridge University Press, 1995.

OECD, "Tax Administrations and Capacity Building: A Collective Challenge," OECD Publishing, Paris, 2016. $http://dx.doi.org/10.1787/9789264256637-en$

OXFAM India, "An Economy That Works for Women," Annual Report 2016-17

OXFAM India, "The Bottom Line: Fighting for the Marginalized Indians," Annual Report 2017-18

OXFAM International, "Reward Work Not Wealth," Annual Report 2018
$https://www-cdn.oxfam.org/s3fs-public/file\_attachments/bp-reward-work-$

$not - wealth - 220118 - en.pdf$

Pareto, V. Ecrits sur la Courbe de la Rèpartition de la Richesse (Genève: Librairie Droz) 1965

Piketty T. and E. Saez, "Optimal Labour Income Taxation," *Handbook of Public Economics*, Edited by A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez, Volume 5, Chapter 7, 391–474 Elsevier B.V., 2013.

Pigou, A. C., *A Study in Public Finance*, Macmillan, London, 1947.

Ramsey, F., "A Contribution to the Theory of Taxation" *The Economic Journal* **37**, 1927, 47-61.

Sadka, E., "On Income Distributions,Incentive Effects and Optimal Income Taxation" *The Review of Economic Studies* **42**, 1976, 261–268.

Seade, J. K., "On the Shape of the Optimal Tax Schedules" *Journal of Public Economics* **7**, 1977, 203–236.

Saez, E., "Using Elasticities to Derive Optimal Income Tax Rates" *Review of Economic Studies* **68**, 2001, 205–229.