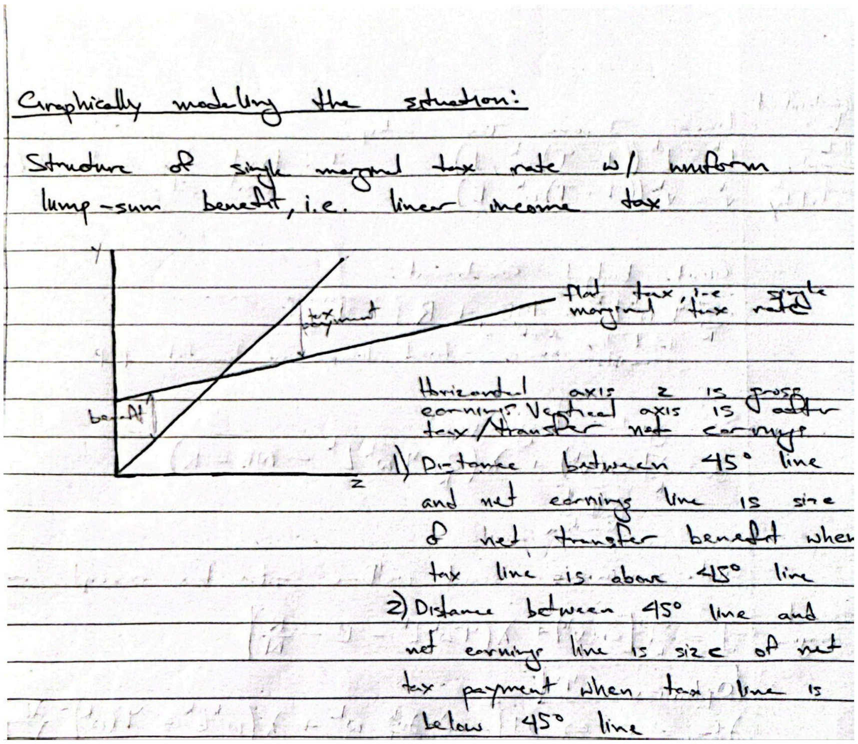
Setup: 

This graphically represents the government's simplified tax system. Because the question asks us to identify the optimal linear income tax, we cannot use the vast majority of the literature and lecture assigned this week. For example, the top-earner should face a zero marginal tax rate after a revenue-neutral tax adjustment, but that would definitionally not be a single marginal tax rate. So we cannot observe Saez's (2001) recommendation of "non-linear taxation is significantly more efficient than linear taxation to redistribute income" because of such liberties, or suggestions from Mankiw et al. (2009) for slight non-linearities.

Executive Summary: The theory from the lecture provides 3 implications for the optimal $\boldsymbol{\tau}$. Then, Stern (1987) provides empirical results with a slightly different setup that captures the exact same ideas. Using Stern's model, with appropriate values for ε (elasticity of social marginal utility of income), $\boldsymbol{\sigma}$ (compensated elasticity of labor supply), and R (normalized revenue target capturing public goods spending), the government can determine $\boldsymbol{\tau}$ and B (uniform lump-sum benefit).

Table of Contents (by page):

(individual) k's... Linear Income Tax Setup:

labor supply $L^k = L^k((1-\tau)w^k, B)$ — τ: fixed tax rate, b/w 0 and 1.

indirect utility $v^k = v^k((1-\tau)w^k, B)$

Govt. Budget Constraint:

$$\underbrace{\sum_k \tau w^k L^k}_{\text{total tax receipts}} = \underbrace{N}_{\text{\# of consumers}} B + R \Big\} \text{ exogenous rev target}$$

B: fixed lump-sum benefit p.p.

So govt. Lagrangian:

$$\mathcal{L} = \sum_k \underbrace{G(v^k)}_{\text{social welfare function}} + \lambda\left(\sum_k \tau w^k L^k - NB - R\right)$$

Govt. maximizes social welfare s.t. tax receipt constraint

rewrite Lagrangian

$$\mathcal{L} = \sum_k \left[G(v^k) + \lambda\left(\tau w^k L^k - B - \frac{R}{N}\right)\right]$$

$\boxed{\text{Optimize } \tau}$

$$\frac{\partial \mathcal{L}}{\partial \tau} = 0 \Leftrightarrow \sum_k \Big[\underbrace{-G' \frac{\partial v^k}{\partial w^k} w^k}_{\text{chain rule x2}} + \lambda\underbrace{\left(w^k L^k - \tau (w^k)^2 \frac{\partial L^k}{\partial w^k}\right)}_{\text{product rule \& chain rule.}}\Big] = 0$$

(Remember that L^k and v^k are functions of tax rate).

Next, recall Slutsky & Roy's Identity for Labor.

See $\boxed{0.0.2}$ and $\boxed{0.0.3}$ in Appendix

So...

$$\sum_k \Big[-G' \frac{\partial v^k}{\partial w^k} w^k + \lambda\left(w^k L^k - \tau (w^k)^2 \frac{\partial L^k}{\partial w^k}\right)\Big] = 0.$$

$\frac{\partial v^k}{\partial w^k} = \frac{\partial v}{\partial B} L$ (Roy's Identity)

$\frac{\partial L^k}{\partial w^k} = \frac{\partial L^c}{\partial w} + L\frac{\partial L}{\partial B}$ (Labor slutsky)

$$\Leftrightarrow \sum_k \Big[-G' \frac{\partial v^k}{\partial B} L w^k + \lambda\left(w^k L^k - \tau (w^k)^2\left(\frac{\partial L^k_c}{\partial w^k} + L\frac{\partial L^k}{\partial B}\right)\right)\Big] = 0$$

distribute

$$\sum_k \Big[-G' \frac{\partial v^k}{\partial B} L w^k + \lambda w^k L^k - \tau\lambda (w^k)^2 \frac{\partial L^k_c}{\partial w^k} - \tau\lambda (w^k)^2 L \frac{\partial L}{\partial B}\Big] = 0$$

Two bracketed terms: $-Lw^k\underbrace{\left[G'\frac{\partial v^k}{\partial B} + \lambda \tau w^k \frac{\partial L}{\partial B}\right]}_{\gamma^k\text{, social marginal utility of income.}}$

See Appendix (0.0.4)

So FOC reduces to:

$$\sum_k \left[-\gamma^k w^k L^k + \lambda w^k L^k - \lambda \tau (w^k)^2 \frac{\partial L_c^k}{\partial w^k} \right] = 0$$

divide both sides by λ $\iff$ $\frac{1}{\lambda} \sum_k \underbrace{\lambda w^k L^k}_{\text{sums to } N\overline{wL}} - \frac{1}{\lambda} \sum_k \gamma^k w^k L^k = \sum_k \tau (w^k)^2 \frac{\partial L_c^k}{\partial w^k}$

$\bar{\gamma}$ (see 0.0.1B)

population covariance formula: $\text{cov}(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$

So LHS becomes: $N \frac{\bar{\gamma}}{\lambda} \overline{wL} - \sum_k \frac{\gamma^k}{\lambda} w^k L^k$

$$\Rightarrow -\left(\sum_k \frac{\gamma^k}{\lambda} w^k L^k - N \frac{\bar{\gamma}}{\lambda} \overline{wL} \right)$$

$$\Rightarrow -\left[N \sum_k \left(\frac{\gamma^k}{\lambda}\right)\left(\frac{1}{N}\right) w^k L^k - N \frac{\bar{\gamma}}{\lambda} \overline{wL} \right]$$

$$\Rightarrow -N \underbrace{\left[\left(\frac{1}{N}\right) \sum_k \left(\frac{\gamma^k}{\lambda}\right) w^k L^k - \frac{\bar{\gamma}}{\lambda} \overline{wL} \right]}$$

$$-N \,\text{cov}\left(\frac{\gamma}{\lambda}, wL\right) \quad \Big\} \text{ LHS reduced}$$

Aside: $\frac{\gamma}{\lambda}$ is social marginal utility of income relative to social value of govt. funds

Can also reduce RHS: $\sum_k \tau (w^k)^2 \frac{\partial L_c^k}{\partial w^k} \Rightarrow \frac{\tau}{(1-\tau)} \sum_k L^k w^k \underbrace{\left[\frac{w^k(1-\tau)}{L^k} \right] \frac{\partial L_c^k}{\partial w^k}}_{\text{elasticity of labor } (\epsilon_L^k)^c}$

(multiply by $\frac{1-\tau}{1-\tau}$ and $\frac{L^k}{L^k}$)

$$\Rightarrow \frac{\tau}{1-\tau} \sum_k L^k w^k (\epsilon_L^k)^c \Rightarrow \boxed{N\left(\frac{\tau}{1-\tau}\right)\bar{\epsilon}}$$

when we define $\bar{\epsilon}$ as $\frac{1}{N} \sum_k L^k w^k (\epsilon_L^k)^c$

Then put LHS & RHS together: $-\cancel{N} \,\text{cov}\left(\frac{\gamma}{\lambda}, wL\right) = \cancel{N}\left(\frac{\tau}{1-\tau}\right)\bar{\epsilon} \Rightarrow \boxed{\frac{\tau}{1-\tau} = -\frac{1}{\bar{\epsilon}} \,\text{cov}\left(\frac{\gamma}{\lambda}, wL\right)}$

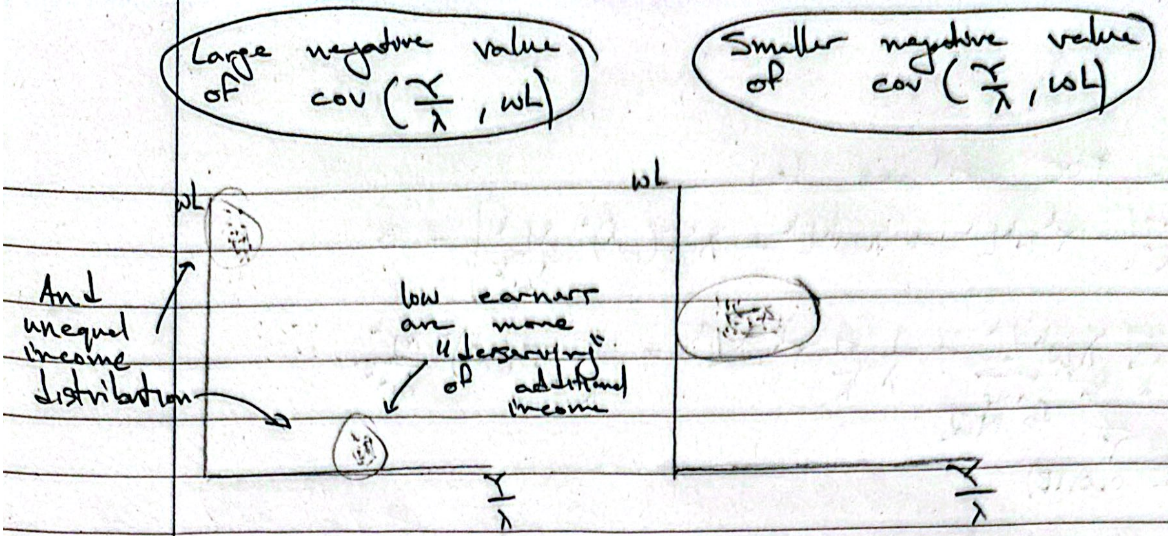
Define $\bar{\epsilon}$: $\frac{1}{N} \sum_k \underbrace{L^k w^k}_{\text{total gross earnings per worker}} \underbrace{(\epsilon_L^k)^c}_{\text{individual elasticity}}$

So: $\bar{\epsilon}$ is average of individual elasticities of labor supply, weighted by how much they earned!

Define $\text{cov}\left(\underbrace{\frac{\gamma}{\lambda}}, \underbrace{wL}_{\text{gross income}}\right)$

$\frac{\gamma}{\lambda}$: society's value for income for person, relative to govt. rev.

This covariance should be negative, b/c poorer people generally have larger γ.



$$\frac{\Upsilon}{1-\Upsilon} = -\frac{1}{\varepsilon} cov\left(\frac{\Upsilon}{\lambda}, wL\right)$$

So Υ depends on mean weighted labor elasticity and inequality of income distribution and society's perception of "deservingness" of additional income for various people.

Implications: 1) The more that high earners in society have more (compensated) elastic labor supply, the lower Υ is.

You could bring in some of the evidence on heterogenous labour supply elasticities here

2) The more unequal that income distribution is, the higher Υ is.

3) The more that society feels lower-income people "deserve" additional income, the higher Υ is.

So all the government must do is estimate $cov\left(\frac{\Upsilon}{\lambda}, wL\right)$ and ε from data to find (optimal) Υ.

From here, finding lump-sum benefit size is simple.

Recall constraint: $\sum_k \Upsilon w^k L^k = NB + \underset{\text{PG}}{R}$ revenue target to fund public goods, treat as exogenous

$$\Rightarrow B = \frac{\sum_k \Upsilon w^k L^k - \overset{PG}{R}}{N} \quad \text{(case w/ public goods)}$$

$$B = \frac{\sum_k \Upsilon w^k L^k}{N}$$

So size of lump-sum benefit is tax receipt (after funding public goods, if applicable), equally divided among all.

Appendix:

0.0.1

Important Asides:

A government can optimize its SWF for B (lump-sum benefit), one that is uniform.

$$\mathcal{L} = \sum_k G(v^k(p^c, m+B)) + \lambda\left(\sum_k T^k - \sum_k B - P_G G\right)$$

indirect utility per consumer function of product prices & budget m + lump-sum benefit; total tax revenue; total lump-sum benefits; total public goods spend

Find $\frac{\partial \mathcal{L}}{\partial B}$ when population size is N.

$$\sum_k G' \frac{\partial v^k}{\partial m} + \lambda\left(\sum_k \frac{\partial T^k}{\partial m} - N\right) = 0$$

chain rule, b/c v^k is function of B; chain rule b/c taxes are function of income m; from B·N in Lagrangian

$$\Leftrightarrow \sum_k (\gamma^k - \lambda) = 0$$

N moved into summation

$\boxed{\mu^k = \gamma^k - \lambda}$ social marginal utility of transfer

intermediate step: $\gamma^k : G' \frac{\partial v^k}{\partial m} + \lambda \frac{\partial T^k}{\partial m}$ social marginal utility of income

From simple $\mathcal{L} = \sum_k G(v^k(p_c, m^k)) + \lambda(\sum_k T^k - R)$

$\therefore$ dividing both sides by N : $\frac{1}{N}\sum_k \mu^k = 0$

At optimal B, mean SMW of transfer = 0.

0.0.1 B $\quad \sum_k (\gamma^k - \lambda) = 0 \Rightarrow \sum_k \gamma^k = N\lambda \Rightarrow \lambda = \frac{\sum_k \gamma^k}{N} \Rightarrow \lambda = \overline{\gamma}$

0.0.2

Slutsky for Labor:

$$\frac{\partial L}{\partial w} = \frac{\partial L_c}{\partial w} + L\frac{\partial L}{\partial B}$$

change in labor supply resulting from wage rate change; substitution effect; income effect; positive b/c labor is "bad"

$L = L(w, B)$

i.e. amount of labor supplied depends on wage rate & benefits size

0.0.3

Roy's Identity for Labor:

$v(w, B)$: Indirect utility is function of wage rate & lump-sum benefit.

$$\frac{\partial v}{\partial w} = \frac{\partial v}{\partial B} L$$

normally, Roy's Identity:

$$\underset{\text{Marshallian demand}}{x_i(p,m)} = -\left(\frac{\frac{\partial v(p,m)}{\partial p_i}}{\frac{\partial v(p,m)}{\partial m}}\right)$$

in this case:

$$L(w,B) = \frac{\frac{\partial v}{\partial w}}{\frac{\partial v}{\partial B}} \Rightarrow \frac{\partial v}{\partial w} = \frac{\partial v}{\partial B} L$$

(labor is a "bad" so no negative)

0.0.4 Social Marginal Utility of Income (γ^k)

Govt. Lagrangian: $\mathcal{L} = \sum_k G(v^k((1-\tau)w^k, B)) + \lambda(\sum_k \tau w^k L^k - NB - R)$

So recall $\gamma^k = G' \frac{\partial v^k}{\partial B} + \lambda \frac{\partial T^k}{\partial B}$ and $T^k = \tau w^k L^k$

(from general case) $\Rightarrow \gamma^k = G' \frac{\partial v^k}{\partial B} + \lambda \frac{\partial(\tau w^k L^k)}{\partial B}$ ← recall this is function of B

$$\Leftrightarrow \gamma^k = G' \frac{\partial v^k}{\partial B} + \lambda \tau w^k \frac{\partial L^k}{\partial B}$$

To give an idea of empirical optimal linear tax rates, Stern (1976) is useful. I have written the framework below. **(Read that, then proceed.)** Stern (1976) suggests a $\boldsymbol{\sigma} \approx 0.4$ for married males in the United States. He notes that when $\boldsymbol{\sigma} < 1$, there is a positive relationship between wage rate and labor supply for low wages and a negative relationship between wage rate and labor supply for higher wages. For the value of ε, he claims that it tends to be $1 < \varepsilon < 2$. In other words, he writes that ε=0 corresponds to a utilitarian social welfare function (SWF), where there is "an absence of aversion to inequality in incomes" and ε=∞ being the Rawlsian maxi-min, where a society is only as well-off as its poorest citizen.

$\boxed{\sigma = \frac{1}{1+\rho}}$, where σ is the compensated elasticity of labor supply.

From an empirical estimate of σ, can back out ρ, a component of every consumer's constant elasticity of substitution $\boxed{\text{utility function:}}$ $\boxed{u(c,\ell) = [\alpha(1-\ell)^{-\rho} + (1-\alpha)c^{-\rho}]^{-\frac{1}{\rho}}}$ $\boxed{\text{(per consumer)}}$

$\boxed{\text{Social Welfare Function}}$: $\boxed{\frac{1}{1-\epsilon}\int_0^\infty [u(c,\ell)]^{1-\epsilon} \underbrace{f(w)}_{\text{density function, distribution of wage rates}} dw}$ $\boxed{(c \text{ is consumption, } \ell \text{ is labor supply})}$

Note that ϵ has a different meaning here, compared to lecture notes. Here, ϵ is the elasticity of social marginal utility of income Y.

To complete the model:

$\boxed{c = (1-t)w\ell + g}$ ← $\boxed{\text{individual budget constraint}}$

t: linear income tax aka constant marginal rate

g: lump sum grant, analogous to B

Finally, the $\boxed{\text{government's budget constraint}}$:

$\boxed{t\int w\ell f(w)\,dw = g + R}$ ← R: exogenous revenue req.

Note that this model is very similar to that in lecture, just different notation.

Table 2-1. *Calculations of Optimal Linear Marginal Tax Rates*

σ	ε = 0 t	ε = 0 g	ε = 2 t	ε = 2 g	ε = 3 t	ε = 3 g	ε = ∞ t	ε = ∞ g
			R = 0 (purely redistributive tax)					
0.2	36.2	0.096	62.7	0.161	67.0	0.171	92.6	0.212
0.4	22.3	0.057	47.7	0.116	52.7	0.126	83.9	0.167
0.6	17.0	0.042	38.9	0.090	43.8	0.099	75.6	0.135
0.8	14.1	0.034	33.1	0.073	37.6	0.081	68.2	0.111
1.0	12.7	0.029	29.1	0.062	33.4	0.068	62.1	0.094
			R = 0.05 (equivalent to about 20 percent of GDP)					
0.2	40.6	0.063	68.1	0.135	72.0	0.144	93.8	0.182
0.4	25.4	0.019	54.0	0.089	58.8	0.099	86.7	0.139
0.6	18.9	0.000	45.0	0.061	50.1	0.071	79.8	0.107
0.8	19.7	0.000	38.9	0.042	43.8	0.051	73.6	0.082
1.0	20.6	0.000	34.7	0.029	39.5	0.037	68.5	0.064
			R = 0.10 (equivalent to about 45 percent of GDP)					
0.2	45.6	0.034	73.3	0.110	76.7	0.119	95.0+	—
0.4	35.1	0.000	60.5	0.065	65.1	0.076	89.3	0.112
0.6	36.6	0.000	52.0	0.036	57.1	0.047	83.9	0.081
0.8	38.6	0.000	46.0	0.016	51.3	0.026	79.2	0.057
1.0	40.9	0.000	41.7	0.002	47.0	0.011	75.6	0.039

Notes: In each pair of columns, the first entry is the marginal tax rate (percent) and the second is the lump-sum grant. ε = 0 corresponds (roughly) to an absence of aversion to inequality in incomes and ε = ∞ to the Rawlsian maxi-min. In Stern (1976), (1 − ε) is used in place of ε. A central estimate of the elasticity of substitution σ might be 0.4. Total output in these models is about 0.25 (it is endogenous), so that R = 0.05 corresponds to government spending (excluding transfer payments) of about 20 percent of gross national product. Where the optimal tax rate is above 95 percent, the precise level and g were not calculated (a dash is shown for g). The level of the uniform grant g was not presented in the Stern (1976) tables. It satisfies the government budget constraint 2-17, that is, $tY = g + R$, where Y is national income per capita. Hence Y can be calculated from the values of t, g, and R presented.

Source: Based on Stern (1976), table 3.

This table simply puts numbers attached to the implications we already derived from the lectures. For example, implication 1 that the more elastic weighted labor supply is, the lower the (optimal) $\boldsymbol{\tau}$ is: we see this with $\boldsymbol{\tau}$ strictly decreasing as $\boldsymbol{\sigma}$ increases. Implication 3 is also quantified: the more that society feels lower-income people "deserve" additional income, the higher $\boldsymbol{\tau}$ is, seen as ε increases from 0 to ∞. Analogously, the size of the lump-sum benefit (B in lecture, g here), increases with ε. And obviously, $\boldsymbol{\tau}$ increases with the government's exogenous revenue target R.

From this table, a government has almost everything it needs to determine its optimal tax rate. We know that realistically, R cannot equal 0, because there needs to be national defense spending at the very least. In my country, the United States, total government expenditure (federal+state) amounted to 36.2% of GDP in 2023. However, in our simple model, only a fraction of this is spending on public goods, which should be equal to the exogenous revenue target. So we should exclude Medicare/Medicaid, SNAP, Social Security, and more when calculating R. (All of this would be part of B.) Then, R will likely be 0<R<0.05—the US government spending less than 20% of GDP on pure public goods sounds plausible. Next, I would argue that American society has a lower ε than most of the OECD. Finally, I would argue that $0.4<\boldsymbol{\sigma}<0.6$ today, because even though Stern claimed 0.4 as a central estimate, this only includes married men. Women, and especially women with children, generally have more elastic labor supply, and also represent a larger portion of the workforce than in 1976.

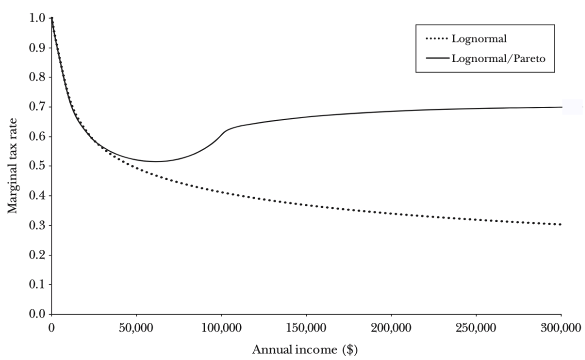
Good

A government would achieve more definite values for R, $\boldsymbol{\sigma}$, and ε than my gut instincts. Then they would have to extend this chart to pinpoint the $\boldsymbol{\tau}$ and g corresponding to those values. This would get them the optimal single marginal tax rate and lump-sum benefit, pursuant to all the assumptions made (e.g., identical consumer utilities). Indeed, Mankiw, et al. (2009) writes "the optimal tax schedule is sensitive to assumptions about the inputs…the shape of the distribution of abilities, the social welfare function, and labor supply elasticities. None of these three components of the problem is easily pinned down."

But Mankiw, et al. (2009) concludes that the optimal marginal tax rate is between 48 to 50 percent for all but the lowest and highest-skilled workers, thus providing a universal lump-sum grant of ~60% the average income per worker. So Mankiw's optimal linear tax rate is close to Stern's 54% marginal tax rate at ε=2, R=0.05, and σ=0.4. Even then, Mankiw's work is less relevant than Stern's in directly answering this question of uncompromisingly constant marginal tax rates. Nonetheless, these numbers provide an estimate for answering Question 5. The hypothetical government's $\boldsymbol{\tau}$ may be close to these.

Figure 3

Optimal Marginal Tax Simulations, with Different Ability Distributions



Note: The figure shows optimal marginal tax rates given two different ability distributions: one lognormal; and one lognormal until approximately $43 per hour and Pareto thereafter.

Your essay demonstrates wide reading and a good understanding of the linear income tax model. A critical discussion of some of the assumptions that go underly the linear income tax model would have given you some more marks.