# Optimal Tax Mix with Income Tax Non-compliance*

Jason Huang †
Stanford University

Juan Rios‡
Stanford University

September 2016

**Abstract**

Although developing countries face high levels of income inequality, they rely more on consumption taxes, which tend to be linear and are less effective for redistribution than a non-linear income tax. One explanation for this pattern is that the consumption taxes are generally more enforceable in these economies. This paper studies the optimal combination of a linear consumption tax with a non-linear income tax for redistributive purposes. In our model, households might not comply with the income tax code by reporting income levels that differ from their true income. However, the consumption tax is fully enforceable. We derive a formula for the optimal income tax schedule as a function of the consumption tax rate, the recoverable elasticities, and the moments of the taxable income distribution. Our equation differs from those of Mirrlees (1971) and Saez (2001) because households face a consumption tax and they respond to income tax not only through labor supply but also through misreporting their incomes. Both aspects are empirically relevant to our calibration of the optimal top rate in the Russian economy. We then characterize the optimal mix between a linear consumption tax rate and a non-linear income tax schedule. Finally, we find that the optimal consumption tax rate is non-increasing in the redistributive motives of the social planner.

*We are grateful to the editor, the two anonymous referees, Douglas Bernheim, Tim Bresnahan, Liran Einav, Caroline Hoxby, Matthew Jackson, Petra Persson, Florian Scheuer and Alex Wolitzky for their invaluable comments and advice.

†Stanford Economics Department; 579 Serra Mall, Stanford, CA 94305-6072. E-mail:*jhuang99* at *stanford.edu*

‡Stanford Economics Department; 579 Serra Mall, Stanford, CA 94305-6072. E-mail:*juanfrr* at *stanford.edu*

*JEL: D31, D63, H21, H24, H26, O21*
*Keywords: Labor Supply, Tax Non-compliance, Optimal Tax, Income Tax, Redistribution, Tax Evasion, Development*

# 1 Introduction

In developing countries, higher proportions of tax revenues come from consumption taxes, which are linear, rather than income taxes, which are typically non-linear. For instance, while in Mexico 49.2% of the tax revenue came from consumption taxes in 2013, this proportion was on average 30.7% among OECD countries and it was below 15% in the United States in the same year (OECD, 2015). One possible explanation for such pattern is that the usually non-linear income tax is more vulnerable to non-compliance than most linear taxes. There are at least three reasons for this. The first one is the simplicity of these linear taxes. The non-linear nature of the income tax requires from the government information on individuals' incomes, while linear taxes are impersonal and therefore do not require information on how much of each good was consumed by each individual. This milder information requirement makes enforcing linear taxes easier. The second reason is the self-enforcing mechanism of value added taxes (VAT) adopted in many economies across the world. Many countries have adopted VAT schemes because of their self-enforcing structures (Keen and Lockwood, 2010). This self-enforcement mechanism has been empirically documented in different settings (see for instance Naritomi (2013) and Pomeranz (2013)). Finally, consumption taxes are easier to enforce because there are fewer points of collection (firms) than in the case of income tax (workers).

Historically, we see a relationship between the maturity of a government's tax collection infrastructure and its reliance on non-linear income tax. The share of federal revenue in the US coming from excise taxes decreased from 12.6% in 1960 down to 2.7% in 2008, while the income tax share increased from 44% to 45.4% (Office of Management and Budget, historical tables, government receipts by source). The U.S. did not enact an income tax until 1861, while excise taxes have been in place since right after the ratification of the Constitution in 1789 (Historical Statistics of the United States Series). More generally, governments in early stages of development have relied on import tariffs, a form of linear consumption tax, because the authorities can focus all their collection efforts at the ports. Despite these enforcing advantages of consumption taxes, redistributing resource across different individuals using these taxes is harder because of their linearity.

In this paper, we leverage the strength of one tax to overcome the weakness of the other. We study the optimal use of a non-linear income tax, which is susceptible to non-compliance, and a linear consumption tax, which we assume is perfectly enforced. There are three main results.

First, we derive the optimal non-linear income tax schedule as a function of the linear

consumption tax rate, taxable and misreported income elasticities, and the moments of the taxable income distribution. Our schedule contains a corrective term that captures how the income tax adjusts to correct the limitation caused by the linear consumption tax. Intuitively, the marginal income tax rates are lower under the presence of the consumption tax, since less revenue needs to be collected through the income tax. Furthermore, as the non-compliance behavior becomes more responsive to the non-linear tax, the capacity of this tax to undue the linear tax distortions diminishes.

Second, we describe the optimal linear tax rate jointly with the optimal non-linear tax schedule. Perturbing the consumption tax affects social welfare through two channels. First, increasing consumption tax raises social welfare by expanding the set of implementable income tax schedules. Non-compliance puts a ceiling on marginal income tax rates, since too high rates lead households to report no income. Higher consumption tax reduces the marginal benefit of hiding income, allowing the social planner to impose higher rates before hitting the zero income constraint. Second, increasing the consumption tax decreases under-reporting and increases over-reporting as marginal value of labor decreases. These changes in mis-reporting behavior decreases all households' marginal mis-reporting cost. The optimal linear tax set the marginal benefit of more implementable income tax schedule from the first channel equal to net marginal cost of non-compliance from the second channel.[1]

The two channels through which the consumption tax affects welfare also offer some intuition for why developing economies that present high levels of income inequality but suffer from income tax enforcement tend to set high linear tax rates. The second channel affects all households but the first channel affects only those reporting zero income. While all households can mis-report their incomes, changes in the marginal income tax rates affect the mis-reporting behavior of only the households reporting positive incomes. Hence, we consider the households reporting positive income as *marginal* households and the ones reporting zero income as *infra-marginal*. As the infra-marginal set of households becomes larger, the optimal consumption tax rises in order to decrease the marginal utility of evasion, balancing the cost of non-compliance and decreasing the set of households reporting zero income. This change allows the planner to set higher marginal income tax rates.

Finally, the optimal linear consumption tax rate is non-decreasing in the redistributive mo-

[1]Even though perturbing the consumption tax changes reported incomes as well, the envelope theorem informs us that these changes have no effect on social welfare when an optimal income tax schedule is in place.

tives of the social planner. This result may appear surprising at first glance, since linear taxes are perceived as regressive. However, our result involves the joint optimal tax structure of the linear consumption tax and the non-linear income tax. A more redistributive social planner tends to implement higher marginal income tax rates, causing households to evade more taxes. Because the consumption tax cannot be evaded, a higher consumption tax rate discourages tax evasion by lowering the marginal benefit of an additional unit of evaded income. Hence, to combat the increase in evasion, the social planner sets a higher consumption tax rate.

Our work contributes to the literature in optimal taxation started by Mirrlees (1971). Part of this literature has addressed the possibility of evasion. Sandmo (1981) constructs a model with two groups of taxpayers (evaders and non-evaders) but restricts the income tax to be linear and set the probability of detection to be endogenous. Cremer and Gahvari (1996) incorporate tax evasion to the optimal income tax problem, where the social planner chooses not only the optimal tax schedule but also the optimal audit structure, restricting the analysis to 2 types. Alternatively, Schroyen (1997) models a two-class economy with an official and an unofficial labor market. The official economy is taxed non-linearly, while unofficial income is only observable after a costly audit upon which it is taxed at an exogenous penalty rate. Kopczuk (2000) considers an optimal linear tax problem in which individuals differ not only in their levels of labor productivity but also in their costs of avoidance. He finds that allowing for tax avoidance may improve social welfare since allowing people to avoid paying tax can serve as a redistributive mechanism. Sandmo (2005) reviews this literature. More recently, Piketty et al. (2014) derive the optimal income tax formula as a function of the labor supply, tax avoidance and compensation bargaining elasticities.

Another part of this literature has determined the optimal commodity taxes joint with the optimal income tax. Atkinson and Stiglitz (1976) show that if a general income tax function may be chosen by the government, no commodity tax should be employed on commodities if the utility function is separable between labor and all commodities. Jacobs and Boadway (2014) consider this problem, restricting the commodity taxes to be linear and writing the optimal tax formulas as a function of recoverable elasticities. However, these papers have not examined the trade-off between a linear tax and a non-linear evadable income tax. We restrict ourselves to a linear and uniform commodity tax over all consumption goods. Technically in our framework, there is no difference between consumption and income tax except for the functional form and the exposure to non-compliance.

The closest paper to the present work is Boadway et al. (1994). They study the use of a

non-linear income and a linear consumption tax when households can only evade income tax in a two-type economy. However, our model allows for a richer heterogeneity in the population. This richness allows us to conduct three exercises which were not feasible in the two-type framework. First, we derive a formula for the non-linear income tax schedule as a function of recoverable elasticities and moments of the income distributions. Second, we provide a formula for the asymptotic top income tax rate and apply the resulting formula to the Russian economy. This exercise shows that taking the consumption tax into account reduces the optimal tax rate by 28%. However, if one recognizes the mis-reporting behavior, a third of this decrease is undone. Finally, we provide a precise characterization of this linear tax, which allows us to perform comparative statics of the optimal linear tax rate with respect to the social planner's redistributive motives.

Finally, the present work also relates to the normative literature of taxation in developing countries. These papers take into account the tax evasion behaviors, common in these countries, to recommend tax policies. For instance, Kleven and Kopczuk (2011) solve for the optimal anti-poverty program in an income maintenance framework. Since they are interested on the trade-off between mis-targeting and take-up of the program, the government's objective is to maximize the number of deserving poor receiving the benefits given a budget. There are two papers directly related to ours in this literature. Gorodnichenko et al. (2009) analyze the welfare costs of income tax reform in Russia, taking into account the different margins of responses (real vs mis-reporting). Kopczuk (2012) conducts a similar exercise using a flat tax reform in Poland. In this paper, we discuss the implication of income tax evasion beyond their partial equilibrium analysis, in an optimal income tax context.

The paper is organized as follows. In section 2, we set up the model. In section 3, we characterize the optimal non-linear tax schedule for a fixed linear tax rate. Then in section 4, we characterize the joint optimal non-linear tax schedule and the linear tax rate. In section 5, we derive comparative statics results. We then build on this exercise in section 6 by conducting numerical simulation of the joint tax system. And in section 7, we conclude.

# 2 Model

We consider a unit mass of heterogenous individuals who differ in their levels of labor productivity, $\theta$, i.e. the amount of income generated with a unit of labor. Let $F(\theta)$ denote a differentiable cumulative distribution function with the probability distribution function

$f(\theta)$ with bounded support of $\theta \in [\underline{\theta}, \bar{\theta}]$ and $\underline{\theta} > 0$. We assume that the individuals have the following quasilinear preference for consumption, reported income and misreported income:

$$U(c, \tilde{y}, \bar{y}; \theta) = c - \psi\left(\frac{\bar{y} + \tilde{y}}{\theta}\right) - \phi(\tilde{y}),$$

where $c$ is consumption, $\bar{y}$ is the reported income, $\tilde{y}$ is the misreported income. The total income is the sum of the reported and misreported component, $\tilde{y} + \bar{y}$, and the labor supply is the total income divided by the household's productivity, i.e. $\frac{\tilde{y}+\bar{y}}{\theta}$. The continuously differentiable function $\psi(\cdot)$ captures the labor supply cost. We assume $\psi'(\cdot) > 0$ and $\psi''(\cdot) > 0$. The continuously differentiable function $\phi(\cdot)$ captures the cost of non-compliance. We can interpret $\phi()$ as the effort that one must exert to misreport its income or the expected disutility from the penalty he suffers when caught. We assume that $\phi(\tilde{y}) \geq 0$ and $\phi''(\tilde{y}) > 0$[2] for all $\tilde{y}$, $\phi'(\tilde{y}) > 0$ for $\tilde{y} > 0$ and $\phi'(\tilde{y}) < 0$ for $\tilde{y} < 0$.

In our model, the domain of $\phi(\cdot)$ is the set of the real numbers because the household can understate or overstate its income, i.e. $\tilde{y}$ can be either positive or negative. When $\tilde{y} > 0$, then $y > \bar{y}$, in which case the household hides income. When $\tilde{y} < 0$, then $\bar{y} > y$, in which case the household claims to have produced more than it actually did. A household may want to inflate its income if it faces a negative marginal tax rate, as in the case of the Earned Income Tax Credit. We need to allow $\tilde{y}$ to be negative to ensure that our solution does not trivially simplify to Mirrlees schedule. If hidden income was restricted to be non-negative, the Social Planner could set the consumption tax high enough to prevent any mis-reporting and set the income tax in order to obtain the Mirrlesian allocation. We formalize this intuition in Proposition 3 in Appendix A.

The households only pay income taxes on their reported income $\bar{y}$ so the amount collected by the government is $T(\bar{y})$. Since our model is static, households consume all their after tax income $\bar{y} + \tilde{y} - T(\bar{y})$, paying a linear tax over this amount at rate $t$. Although there is no distinction between consumption and income in our model, we interpret the linear tax as a consumption tax considering the usual pattern of linear consumption tax versus a non-linear income tax. The households choose both $\bar{y}$ and $\tilde{y}$ to solve the following problem:

$$\max_{\bar{y} \geq 0, \tilde{y} \geq -\bar{y}} (1-t)\left[\tilde{y} + \bar{y} - T(\bar{y})\right] - \psi\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) - \phi(\tilde{y}). \tag{1}$$

The first constraint, $\bar{y} \geq 0$, requires the household to report non-negative income. The

[2] We motivate this assumption by assuming that households exhibit decreasing marginal utility of honesty, since this would imply an increasing marginal cost of under and over-reporting income.

second constraint, $\tilde{y} \geq -\bar{y}$, ensures that the household cannot produce negative total income. The social planner's problem is the following:

$$\max_{T(\bar{y}),t} \int_{\theta\in\Theta} \Bigg\{ (1-t)\left[\tilde{y}_{t,T}(\theta) + \bar{y}_{t,T}(\theta) - T(\bar{y}_{t,T}(\theta))\right] \\ -\psi\left(\frac{\tilde{y}_{t,T}(\theta) + \bar{y}_{t,T}(\theta)}{\theta}\right) - \phi(\tilde{y}_{t,T}(\theta)) \Bigg\} d\tilde{F}(\theta)$$

with $d\tilde{F}(\theta)$ as the pareto weights the social planner puts on type $\theta$, and $\bar{y}_{t,T}(\theta)$ and $\tilde{y}_{t,T}(\theta)$ as the solution to Eq. (1), i.e. optimal amount of income to report and mis-report, respectively, for a given income tax schedule $T()$ and a linear tax rate $t$. The envelope theorem and the fact that the cross-product of the household problem with respect to $t$ and $\tilde{y}$ is negative, the optimal non-compliance decreases with $t$. The intuition is that higher consumption tax reduces the marginal value of hiding income when $\tilde{y}$ is positive or reduces the marginal cost of over-reporting income when $\tilde{y}$ is negative. On the other hand $\bar{y}_{t,T}(\theta)$ is not monotone in $t$, as this relationship will depend on the sign of $T'()$. The chosen tax system must satisfy the following resource constraint

$$\int_{\theta\in\Theta} \Bigg\{ \underbrace{t\left[\tilde{y}_{t,T}(\theta) + \bar{y}_{t,T}(\theta) - T(\bar{y}_{t,T}(\theta))\right]}_{\text{linear tax revenue}} + \underbrace{T(\bar{y}_{t,T}(\theta))}_{\text{non-linear tax revenue}} \Bigg\} dF(\theta) = 0,$$

which ensures that the social planner balances its budget.We can add an exogenous government spending $G$ to the budget constraint. However, this addition does not affect the optimal tax system, except for a lump sum tax $G$ on every household.

To characterize the jointly non-linear tax schedule and the linear tax rate, we proceed in two steps, following the approach developed in Rothschild and Scheuer (2013). First, we solve the inner problem: we characterize the optimal non-linear tax for a given linear tax rate. Such optimal schedule gives us a social welfare $W(t)$ that depends on the linear tax $t$. In the outer problem, we maximize the social welfare with respect to the linear tax rate.

# 3 Inner Problem: Optimal Income Tax Schedule for a Given Consumption Tax

In this section, we take the linear tax as given, and solve the Social Planner's problem for the non-linear tax. Rather than solving for the optimal $T(\cdot)$, we use the revelation principle from mechanism design. We consider the isomorphic problem in which the social planner offers a menu $[\bar{c}_t(\hat{\theta}), \bar{y}_t(\hat{\theta})]$ $\forall \hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ that can depend on the linear tax $t$, with $\bar{y}_t(\hat{\theta})$ as the income that a household that reports type $\hat{\theta}$ hands to the social planner and $\bar{c}_t(\hat{\theta})$ as the transfer in terms of real consumption given back to that household. Note that, because of tax non-compliance, reported type $\hat{\theta}$'s actual consumption is $\bar{c}_t(\hat{\theta}) + (1 - t)\tilde{y}$ rather than $\bar{c}_t(\hat{\theta})$. This feature of our model differs from the traditional application of the mechanism design approach in optimal non-linear taxation.

An individual that reports a type $\hat{\theta}$ receives a $[\bar{c}_t(\hat{\theta}), \bar{y}_t(\hat{\theta})]$ bundle and decides the optimal amount of income to mis-report by solving the following problem:

$$\max_{\tilde{y} \geq -\bar{y}_t(\hat{\theta})} \left\{ \bar{c}_t(\hat{\theta}) + (1-t)\tilde{y} - \psi\left(\frac{\bar{y}_t(\hat{\theta}) + \tilde{y}}{\theta}\right) - \phi(\tilde{y}) \right\} \tag{2}$$

Let:

$$\tilde{y}_t(\theta, \bar{y}) \equiv \text{argmax}_{\tilde{y} \geq -\bar{y}} \left\{ (1-t)\tilde{y} - \psi\left(\frac{\bar{y} + \tilde{y}}{\theta}\right) - \phi(\tilde{y}) \right\} \tag{3}$$

We add a subscript $t$ to $\tilde{y}_t(\cdot)$ because a household's misreporting decision depends on the linear tax rate. Because we assume that the households have quasilinear preferences, this amount of tax non-compliance does not depend on $\bar{c}$. Also, because the objective function of Eq. (2) is strictly concave in $\tilde{y}$ and continuously differentiable in $\tilde{y}$, $\bar{y}$ and $t$, the optimum is unique, continuous and differentiable in $t$ and $\bar{y}$. Since ability is the only source of heterogeneity in our model, $\tilde{y}_t(\theta, \bar{y})$ behaves as $\tilde{y}_{t,T}(\theta)$ and decreases with the consumption tax rate.

Now we introduce the indirect utility function, $u_t(\cdot)$, defined as

$$u_t(\bar{c}, \bar{y}; \theta) = \bar{c} + (1-t)\tilde{y}_t(\theta, \bar{y}) - \psi\left(\frac{\bar{y} + \tilde{y}_t(\theta, \bar{y})}{\theta}\right) - \phi(\tilde{y}_t(\theta, \bar{y})). \tag{4}$$

with $\tilde{y}_t(\cdot)$ as defined above, $\bar{c}$ as the official consumption allocation chosen by the social planner and $\bar{y}$ as the income collected by the social planner. This modified utility function reflects the household's preference by accounting for the act of tax non-compliance.

The utility of type $\theta$ when he reports $\hat{\theta}$ is $u_t(\bar{c}(\hat{\theta}), \bar{y}(\hat{\theta}); \theta)$. The social planner wants to ensure that

$$\theta \in \underset{\hat{\theta}}{\operatorname{argmax}} u_t(\bar{c}_t(\hat{\theta}), \bar{y}_t(\hat{\theta}); \theta),$$

i.e. each type chooses to report its true type. Let $v_t(\cdot)$ be the value function of type $\theta$, which we define as

$$v_t(\theta) = u_t(\bar{c}_t(\theta), \bar{y}_t(\theta); \theta) =$$
$$\bar{c}_t(\theta) + (1-t)\tilde{y}_t(\theta, \bar{y}_t(\theta)) - \psi\left(\frac{\tilde{y}_t(\theta, \bar{y}_t(\theta)) + \bar{y}_t(\theta)}{\theta}\right) - \phi(\tilde{y}_t(\theta, \bar{y}_t(\theta))). \tag{5}$$

Using the envelope theorem with respect to $\bar{c}$, $\bar{y}$ and $\tilde{y}_t$ the local incentive constraint is expressed as the following:

$$v_t'(\theta) = \frac{\bar{y}_t(\theta) + \tilde{y}_t(\theta, \bar{y}_t(\theta))}{\theta^2}\psi'\left(\frac{\bar{y}_t(\theta) + \tilde{y}_t(\theta, \bar{y}_t(\theta))}{\theta}\right). \tag{6}$$

Since Eq. (6) reflects the marginal information rent a type $\theta$ household extracts for not emulating a household less productive by an arbitrarily small amount, this formula provides intuition how the linear tax affects non-compliance and total production, which in turn impacts the incentive constraints. To see this, note that when $t$ increases, the marginal benefit of income decreases and the total income in any equilibrium of an incentive compatible tax schedule also decreases. This decreases $v'(\theta)$ and relaxes the local incentive constraint. Figure 1 displays $v_t(\theta)$ for $t_L < t_M < t_H$. As the consumption tax rate increases, the relationships between the value functions and the types become flatter, and the social planner pays less informational rent to the higher types.

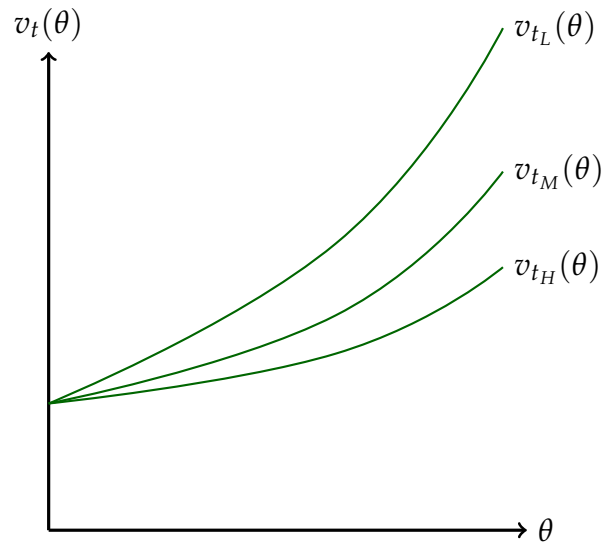


Figure 1: Value Function of Different Types

The following lemma shows that, with monotonicity of $\bar{y}(\theta)$, these intuitions generalize to global incentive compatibility constraints.

**Lemma 1.** *Given the preference in Eq. (4), an allocation is incentive compatible if and only if it satisfies Eq. (6) and $\bar{y}(\theta)$ is non-decreasing.*

*Proof.* See Appendix A. □

The intuition for the lemma is as follows. Because the cost of non-compliance through $\phi()$ is the same for all types but the cost of supplying such misreported income is relatively lower for higher types, the marginal disutility of $\bar{y}$, even accounting for the non-compliance behavior, is decreasing in $\theta$. And since the preference is quasilinear, we have that the marginal substitution between $\bar{c}$ and $\bar{y}$ is decreasing in $\theta$ as well.

With Lemma 1, we proceed with our direct mechanism approach with the new preference (Eq. (4)), replacing the incentive compatibility constraint with local incentive compatibility constraint and monotonicity constraints.

For a given $t$, the Social Planner Problem (SPP) is

$$\max_{v_t(\theta),\bar{y}_t(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} v_t(\theta) d\tilde{F}(\theta)$$

with $d\tilde{F}(\theta)$[3] as the pareto weight, subject to Eq. (6)

$$v_t'(\theta) = \frac{\tilde{y}(\theta, \bar{y}_t(\theta)) + \bar{y}_t(\theta)}{\theta^2} \psi' \left( \frac{\tilde{y}(\theta, \bar{y}_t(\theta)) + \bar{y}_t(\theta)}{\theta} \right)$$

and to the resource constraint, which is

$$\int_{\theta \in \Theta} \{\bar{y}_t(\theta) - \bar{c}_t(\theta) + t\tilde{y}_t(\theta, \bar{y}(\theta))\} dF(\theta) \geq 0. \tag{7}$$

The term $t \int_{\theta \in \Theta} \tilde{y} dF(\theta)$ is the revenue the social planner has from the linear tax in the direct mechanism that it can use for redistribution. The solution of this problem yields the following proposition:

**Proposition 1.** *The first order condition for the optimal income tax rate at positive reported income level $\bar{y}$ with linear tax rate $t$ can be written as either*

$$\frac{T'(\bar{y}_t(\theta)) + \frac{t}{1-t}\left(\frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T}} + 1\right)}{1 - T'(\bar{y}_t(\theta))} = \left(\frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T}} + 1\right) \frac{\tilde{F}(\theta) - F(\theta)}{f(\theta)\theta} \left(1 + \frac{1}{\mathcal{E}_{y,1-T'}}\right) \tag{8}$$

*or*

$$\frac{T'(\bar{y}) + \frac{t}{1-t}\left(\frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T}} + 1\right)}{\mathcal{E}_{\bar{y},1-T'}\bar{y}T''(\bar{y}) + (1 - T'(\bar{y}))} = \frac{\tilde{H}(\bar{y}) - H(\bar{y})}{h(\bar{y})\bar{y}} \frac{1}{\mathcal{E}_{\bar{y},1-T'}}. \tag{9}$$

*The $H(\bar{y})$ is the cumulative reported income distribution and $h(\bar{y})$ is the density of the reported income.*

*Proof.* See Appendix A. □

Our result is similar to those found in the optimal income tax literature, except we have a linear tax rate and non-compliance elasticity. More specifically, suppose $t = 0$. Then Eq. (8) is the same as in Piketty (1997) and Eq. (9) is the same in Lans Bovenberg and Jacobs (2005). However, the schedule is a function of not only the elasticity of taxable $\mathcal{E}_{\bar{y},1-T'}$ income but also of the mis-reported income $\mathcal{E}_{\tilde{y},1-T'}$. These two objects are related. The cost of avoidance is one of the key determinants of the second elasticity at the same time that it affects the first elasticity as shown in Kopczuk (2005).

The optimal non-linear tax adjusts in the presence of the linear tax, since the linear tax

[3] We do not require $d\tilde{F}(\theta)$ to be differentiable. This generalization allows us to consider a Rawlsian social planner who puts all the weight on the lowest type $\underline{\theta}$.

introduces a wedge that the non-linear tax can partially undo. The extent of this adjustment depends on household's non-compliance behavior. In the extreme case when non-compliance is costless, i.e. $\phi() = 0$, we have $-\mathcal{E}_{\tilde{y},1-T'}\tilde{y} = \mathcal{E}_{\bar{y},1-T'}\bar{y}$, since the first order condition of the household problem with respect to $\tilde{y}$ implies $\frac{d\tilde{y}}{d\bar{y}} = -1$. Hence, the optimal non-linear tax does not need to account for the effect of the linear tax. However, the optimal non-linear schedule still depends on the linear tax since this tax influences households' reported and misreported income elasticities. In particular it affects the marginal disutility of reported income relative to the mis-reported income by reducing the marginal utility of consumption. In the other extreme case in the absence of non-compliance, we have $\mathcal{E}_{\tilde{y},1-T'} = 0$, so the non-linear tax must fully account for the linear tax. The intuition is that as misreported income becomes more elastic, the non-linear tax becomes less effective in undoing the linear tax wedge. In fact, when $t > 0$, the optimal income tax schedule that properly accounts for tax non-compliance behavior is more progressive than the schedule that ignores such behavior.

If the linear tax $t = 0$, we see from proposition 1 that we get back the "no distortion at the top" result, i.e. $T'(\bar{y}_{max}) = 0$. This result differs from the one found in Grochulski (2007), which states that, under certain non-compliance cost structure and even with bounded support for skill types, the optimal income tax is progressive. The difference arises because the incomes of households in our model come from supplying labor while they are exogenously endowed in Grochulski (2007).

So far, we have characterized the optimal income tax for strictly positive reported income. There are two interpretation for why $\bar{y}_t(\theta) = 0$. First, the usual condition used to rule out a corner solution in other optimal income tax problems (the marginal disutility from labor when supplying no labor is 0) does not ensure an interior solution because of the presence of tax non-compliance. The intuition is the following. Without the tax non-compliance, reported income equals total income, and at zero reported income, the marginal change in social welfare when slightly increasing production is positive since marginal disutility of production is 0. However, with positive misreported income, even when a household reports zero income, its actual level of production is positive and the marginal disutility from labor might be high. As a result, the marginal social welfare of increasing reported income may be negative, even at level zero. Another interpretation for why the non-negative constraint on $\bar{y}$ may bind is that the presence of non-compliance restricts the set of implementable income tax schedules. When the marginal income tax rates exceed some threshold, households choose to report zero income.

Second, the set of types that report zero income must consist of an interval starting from

the lowest type. This result arises by the monotonicity of $\bar{y}_t(\theta)$. This result agrees with the general findings that the informal markets are less efficient than the formal markets.[4]

In summary, the characterization of the optimal non-linear tax schedule offers two seemingly opposing implications. On one hand, the presence of non-compliance imposes an upper envelope for the feasible tax schedule. Gordon and Li (2005) note this fact when they argue that developing countries need to set lower marginal tax rates so that firms do not move into the informal economy. On the other hand, for the region of the schedule with positive reported income and positive consumption tax, the presence of non-compliance makes marginal tax rates higher than the optimal rates that ignore tax non-compliance.

## 3.1 Top Income Optimal Tax Rates

Once we have fully characterized the non-linear schedule that accounts for the linear tax rate and mis-reporting behavior, we use the result to look at the asymptotic tax rate at the top income of the reported income distribution. We derive this asymptotic rate in the following corollary.

**Corollary 1.** *Assume that the income distribution above $\bar{y}_{top}$ has a Pareto distribution with parameter $a$. Furthermore, the relevant elasticities and the ratio of reported and mis-reported income converges as reported income approaches infinity, i.e. $\mathcal{E}_{\bar{y},1-T'} \to \bar{\mathcal{E}}_{\bar{y}}$, $\mathcal{E}_{\tilde{y},1-T'} \to \bar{\mathcal{E}}_{\tilde{y}}$, and $\frac{\tilde{y}}{\bar{y}} \to b$, then the optimal top income tax converges to a linear rate $\tau$ characterized by*

$$\tau = \frac{1 - \bar{g} - a\frac{t}{1-t}[\bar{\mathcal{E}}_{\bar{y}} + \bar{\mathcal{E}}_{\tilde{y}} b]}{1 - \bar{g} + a\mathcal{E}_{\bar{y}}}, \tag{10}$$

*where $\bar{g} = \frac{1-\tilde{H}(\bar{y}_{top})}{1-H(\bar{y}_{top})}$ is the social marginal utility of giving an extra dollar to a tax payer at the top of the distribution.*

*Proof.* We take the limit of Eq. (9) in Proposition 1 as $\bar{y}$ goes to infinity, use the fact that $\bar{y} \geq \bar{y}_{top}$ follows a pareto distribution with parameter $a$ and hence $\frac{\bar{y}h(\bar{y})}{1-H(\bar{y})} = a$, the definition of $\bar{g}$ and rearrange terms. □

For an economy without consumption tax $t = 0$, this formula reduces to Eq. (9) of Saez

[4] La Porta and Shleifer (2008) examine World Bank firm-level surveys and find that firms in the informal sector are smaller, much less productive, and managed by less educated managers than compared to even small firms in the formal sector.

(2001).[5] The new term $a\frac{t}{1-t}[\bar{\mathcal{E}}_{\tilde{y}} + \bar{\mathcal{E}}_{\tilde{y}}b]$ captures the readjustment made by the income tax as a result of the linear tax. As before, the larger is the mis-reported income elasticity the smaller is this readjustment because the income tax is less effective in overcoming the wedge created by the linear tax.

Table 1 provides optimal asymptotic rates using Eq. (10). We make this calculation for the Russian economy using a set of pareto parameters of the income distribution and welfare weights on the top income group under different consumption tax and non-compliance behavior scenarios. In the first three columns, we assume no linear tax. In the next three columns, we assume 18% consumption tax (Russian VAT rate according to the Federal Tax Service of Russia) without non-compliance behavior. In the last three columns, we assume the same consumption tax and allow for income mis-reporting. We choose Russia for our analysis because it has the best estimates of mis-reported income elasticities. Gorodnichenko et al. (2009) estimate a reported income elasticity of 0.47, a mis-reported income elasticity of $-0.376$ and $\tilde{y}/\bar{y}$ to be on average 0.39.[6] We assume this is the average value also above the top threshold so that $b = 0.39$. Finally, because we do not have a access to the top income distribution of the top taxable income of Russia, we calculate the optimal rate under three different values for the Pareto parameter $a = 1.5, 2, 2.5$.

The comparison between columns 1-3 and columns 4-6 shows that the consumption tax rate drives the optimal top rate downward by around 28% in the absence of mis-reporting behavior. However this large drop is partially undone when we incorporate mis-reporting in the model (last three columns). For the values estimated to the Russian economy, including mis-reporting in the model would undo 30% of this drop. In the extreme case, in which mis-reporting was the only margin of response ($\bar{\mathcal{E}}_{\tilde{y}} = -\bar{\mathcal{E}}_{\tilde{y}}\tilde{y}_m/\bar{y}_m$), the last term in the numerator of Eq. 10 vanishes, and the adjustment to the consumption tax disappears. As usual, the optimal rate decreases with the welfare weight in the top bracket $g$ and with the Pareto parameter of the top reported income distribution $a$.

# 4 Outer Problem: Characterizing Optimal Consumption Tax

In the previous section, we solved the inner problem by characterizing the optimal income tax schedule for a given linear tax rate. Now we solve the outer problem, i.e. we maximize

[5] More precisely, it corresponds to this equation when there are not income effects and hence the compensated and uncompensated elasticities are identical.

[6] They report that evasion is 28% of total income, which implies $\bar{y}/\tilde{y} = 0.39$, under the assumption that evasion is the unique source of mis-reporting.

Table 1: Optimal income tax rates for high income earners

| $g$ | $t=0$ | | | $t=0.18$ and $\tilde{y}/\bar{y}=0$ | | | $t=0.18$ and $\tilde{y}/\bar{y}=0.39$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $a=1.5$ | $a=2$ | $a=2.5$ | $a=1.5$ | $a=2$ | $a=2.5$ | $a=1.5$ | $a=2$ | $a=2.5$ |
| 0 | 58 | 51 | 46 | 49 | 41 | 34 | 52 | 44 | 37 |
| 0.25 | 51 | 44 | 39 | 41 | 32 | 25 | 44 | 36 | 29 |

Note: $g$ is the ratio of social marginal utility with infinite income over marginal value of public funds. The pareto parameter of the income distribution $a$ takes values 1·5, 2, 2·5. In the first three columns we assume no consumption tax in the economy. In columns 4-6, we assume the consumption tax is 18% with no mis-reporting. In the last three columns, we assume mis-reported income to be 39% of reported income. Optimal rates are computed according to formula (10).

the welfare with the optimal income tax in place with respect to the consumption tax rate.

**Proposition 2.** *The optimal consumption tax rate, with the optimal non-linear income tax schedule derived in Proposition 1, must satisfy the following condition:*

$$\int_{\underline{\theta}}^{\bar{\theta}} \left( \frac{\phi'(\tilde{y}_t(\theta,\bar{y})) + \lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta,\bar{y}))} \right) \Bigg|_{\bar{y}_t(\theta)} f(\theta) d\theta = 0, \tag{11}$$

*with $\bar{y}_t(\theta)$ defined by as the solution to Eq. (14) and with $\lambda_t(\theta)$ as the lagrange multiplier on the constraint $\bar{y}_t(\theta) \geq 0$. The above expression can be rewritten as*

$$\int_{\hat{\theta}}^{\bar{\theta}} \frac{A_t(\theta)(B_t(\theta)(1-t)-t)}{1-t(1-A_t(\theta))} y_t(\theta) \mathcal{E}_{y,1-T'} dF(\theta) + \int_{\underline{\theta}}^{\hat{\theta}} \frac{\phi'(\tilde{y}_t(\theta,0)) + \lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta,0))} dF(\theta) = 0 \tag{12}$$

*with*

$$A_t(\theta) \equiv \left( \frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T'}} + 1 \right)$$

$$B_t(\theta) \equiv \frac{\tilde{F}(\theta) - F(\theta)}{f(\theta)\theta} \left( 1 + \frac{1}{\mathcal{E}_{y,1-T'}} \right)$$

$$\hat{\theta} = \max\{\theta : \bar{y}_t(\theta) = 0\}.$$

*and with $y_t(\theta)$ defined by Eq. (19).*

*Proof.* See Appendix A. □

Our optimal condition for the linear tax rate decomposes into two terms. We express the first term as a function of the moments of income and elasticities. However, we leave the second term with primitives, since it composes of infra-marginal types that report no income and do not respond to marginal non-linear tax rates. This result distinguishes

between two types of black market activities: the parts that are responsive to non-linear tax, and the ones that are not. Highlighting this distinction help to explain the difference in the use of linear taxes across economies.

This decomposition provides some intuition for why developing countries with more concerns for redistribution set higher linear tax rates. Those countries with weaker tax enforcement may have a larger infra-marginal black market sector, which our model captures in the second term of Eq. (12). That equation sets the social marginal welfare with respect to $t$ to zero. Assuming that optimal $t$ is unique, the marginal social welfare crosses zero once from above. Since the second term in that equation is strictly positive [7], it pushes up the marginal social welfare curve, leading to a higher optimal $t$[8]. Figure 2 provides a graphical interpretation of this argument, with the lower (upper) curve representing an economy with small (large) second term.

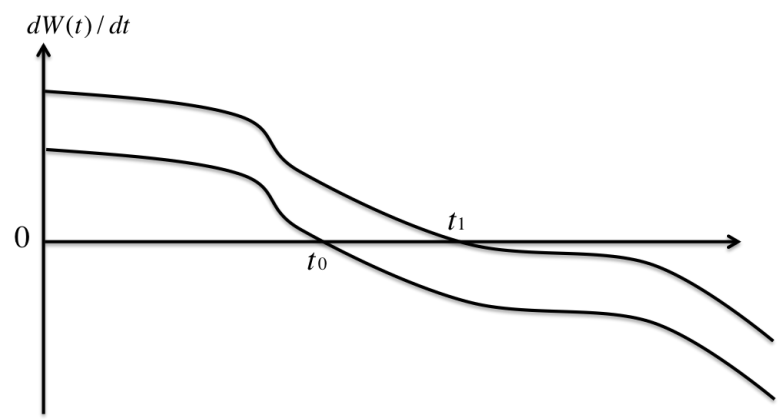


Figure 2: Marginal Social Welfare With Respect to $t$

Note: $dW(t)/dt$ denotes the marginal social welfare with respect to $t$ as a function of $t$.

Notice that all the effects that depend on total income, i.e. $\psi()$ and its higher derivatives, disappear when characterizing the optimal linear tax rate $t$. The social planner has full control over them through the non-linear tax. Here, we see how the flexibility of the income tax complements the linearity of the enforceable tax. And since the social planner has already optimized the non-linear tax in the inner problem, the envelope theorem states that changing $t$ slightly does not impact the social welfare effect through this total income.

However, the social planner cannot fully control tax non-compliance, and non-compliance

[7] Note that when $\bar{y} = 0$, $\tilde{y} > 0$, so that $\phi' > 0$.

[8] We do not formalize this intuition into a proposition since our argument rests on endogenous terms rather than primitive. Changing the primitive that increases the proportion of household reporting zero income can change the endogenous elasticities.

changes the solution in two ways. First, an increase in $\tilde{y}$ increases the social cost of non-compliance by $\int \phi'(\tilde{y}(\theta))d\tilde{F}(\theta)$. Second, even when $\psi'(0) = 0$, non-compliance may cause the constraint $\bar{y} \geq 0$ to bind. We can think of this second force as non-compliance's impact on the social planner's ability to implement a desired non-linear tax schedule. Hence, a change in $t$ only affects social welfare through the cost of non-compliance and the boundary conditions, the two channels that the social planner does not fully control.

To fix ideas, consider an exogenous increase of the social cost of evasion $\phi(\cdot)$ which increases this cost to $k \cdot \phi(\cdot)$ for some $k > 1$. This would directly affect welfare by reducing the utility of all households mis-reporting in equilibrium, but it would also affect the set of households that would choose to report zero income. This would in turn change the possible set of schedules the social planner can choose in order to raise the revenue necessary for redistribution.

Finally, note that the social planner cares about the breakdown of the total income into reported and misreported to the extent that cost of non-compliance impacts the social welfare. To better illustrate this point, consider an economy in which a $1 - \alpha$ fraction of the cost of non-compliance is transferred back to the social planner, as considered in Chetty (2009). For example, if part of the costs of non-compliance consists of fines households must pay when caught, these fines can be sources of additional revenue that the social planner can use for redistribution. On the other hand, the $\alpha$ fraction of the cost can reflect the efforts to learn about the tax code or the expense of hiring tax lawyers. We can instead characterize the optimal income tax schedule for a given linear tax $t$ as the solution to the following point-wise maximization.

$$\begin{aligned}\bar{y}_t(\theta) = \arg\max_{\bar{y}\geq 0} \bar{y} + \tilde{y}_t(\theta,\bar{y}) - \psi\left(\frac{\tilde{y}_t(\theta,\bar{y}) + \bar{y}}{\theta}\right) - \alpha\phi(\tilde{y}_t(\theta,\bar{y})) \\ + \frac{\tilde{y}_t(\theta,\bar{y}) + \bar{y}}{\theta^2}\psi'\left(\frac{\tilde{y}_t(\theta,\bar{y}) + \bar{y}}{\theta}\right)\left(\frac{F(\theta) - \tilde{F}(\theta)}{f(\theta)}\right). \end{aligned} \tag{13}$$

This problem is analogous to Eq. (14) found in the proof of proposition 1 with an $\alpha$ in front of $\phi()$ term. In the case that $\alpha = 0$, all the misreporting cost incurred by the households are transferred back to the social planner. If the non-negativity constraint on $\bar{y}$ does not bind, the optimal allocation only depends on the total production and the linear tax has no role. To see this more precisely, when $\alpha = 0$, the objective function of Eq. (13) is a function of $\bar{y} + \tilde{y}_t(\theta,\bar{y})$. As long as the constraint $\bar{y} \geq 0$ does not bind, $\bar{y} + \tilde{y}_t(\theta,\bar{y})$ is strictly increasing with respect to $\bar{y}$, regardless of the non-compliance behavior of the household. Hence the social planner can choose the appropriate $\bar{y}$ to yield the desired

total income. Intuitively, when evasion does not generate resource costs ($\alpha = 0$), the planner can implement redistribution without caring about the enforcement of each tax instrument. However, when mis-reporting is wasteful ($\alpha > 0$), the planner trades off the enforceability of the consumption tax against the progressivity of the income tax which yields a mix of both instruments.

# 5 Comparative Statics

In this section, we discuss how the optimal linear tax rate varies with the redistributive motives of the social planner. Our analysis uses proposition 2. First, we formalize the degree of a redistributive motive with the following definition.

**Definition 1.** Pareto weights $\tilde{F}(\theta)$ are more redistributive than pareto weights $\tilde{F}'(\theta)$ when $\tilde{F}(\theta)$ is first order stochastically dominated by $\tilde{F}'(\theta)$, i.e. $\tilde{F}(\theta) \geq \tilde{F}'(\theta)$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

If we assume that $\phi'()/\phi''()$ is non-decreasing, a condition that is satisfied by an iso-elastic mis-reporting cost function $\phi(\tilde{y}) = \frac{|\tilde{y}|^{\delta}}{\delta}$, and that we assume $W(t)$ is differentiable and single-peaked, we can show that the optimal linear tax rate is weakly increasing with the planner's redistributive motives (corollary 2). The interpretation of the first condition is that the marginal impact of increasing $t$ on social welfare through the cost of non-compliance does not decrease as mis-reported income increases. The interpretation for the second condition is straightforward, but to fully specify the set of functional forms and parameters to ensure the condition is quite cumbersome. For all the cases that we numerically checked when assuming $\phi(\tilde{y}) = \frac{|\tilde{y}|^{\delta}}{\delta}$ and $\psi(x) = \frac{x^{\gamma}}{\gamma}$, functional forms commonly assumed in the optimal tax literature, the necessary conditions are also sufficient conditions for optimality.

**Corollary 2.** *Suppose that $\phi'()/\phi''()$ is a non-decreasing function and $W(t)$ is differentiable and single peaked, i.e. optimal $t$ is unique. Then the optimal $t$ never decreases when the social planner becomes more redistributive.*

*Proof.* See Appendix □

As discussed in more detail in Appendix A, the condition that $\phi'()/\phi''()$ is non-decreasing ensures that the marginal social welfare with respect to changes in $t$, i.e. $\frac{dW(t)}{dt}$ is non-decreasing with respect to the redistributive motive of the social planner. Consider figure 3. Suppose $W(t)$ and $W'(t)$ are the social welfares as functions of $t$ under and $\tilde{F}(\theta)$ and $\tilde{F}'(\theta)$, respectively, and that $\tilde{F}(\theta)$ is more redistributive than $\tilde{F}'(\theta)$. As we show in the

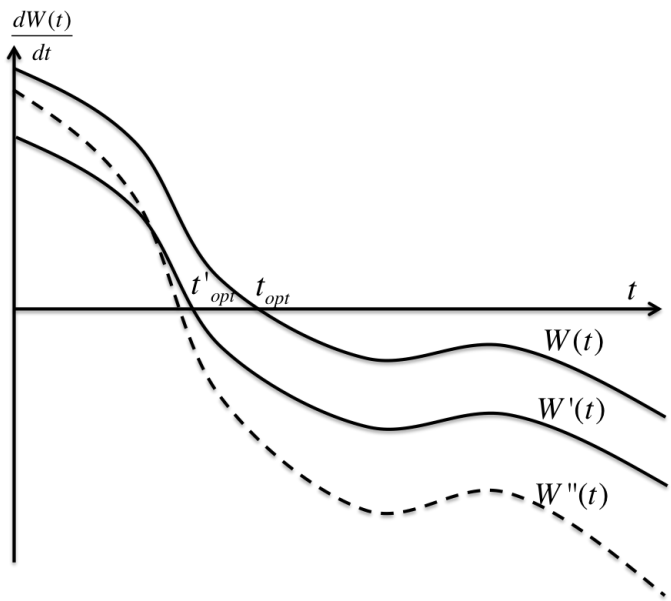


Figure 3: Condition Ensured by the Monotonicity of $\phi'()/\phi''()$

Note: The two solid lines represent two possible marginal social welfare curves under different Pareto weights. The dashed line represents a marginal social welfare curve ruled out by the monotonicity condition of $\frac{\phi'()}{\phi''()}$.

proof, $dW(t)/dt$ always lie above $dW'(t)/dt$, and the two curves do not cross. Hence, a social welfare function like $W''(t)$ cannot exist, since its derivative curve crosses the derivative curve of $W'(t)$. Therefore, $t_{opt}$ is larger under the more redistributive weights.

Corollary 2 may seem surprising, as linear tax is perceived as "regressive." However, when social planner becomes more redistributive, the marginal income tax rates increase. These increases in the rates result in higher mis-reported income $\tilde{y}$. More specifically, households who evade incomes will evade more and the households who inflate their incomes will inflate less. As a result, the social planner needs a higher linear tax rate to offset the rise in $\tilde{y}$.

# 6 Numerical Simulation

We have shown that when social planner cares more about the lower types, linear tax rate increases to combat the potentially higher non-linear tax rates. This result begs the question of how the marginal income tax rate moves *relative* to the consumption tax across the income distribution. We use numerical simulation of our optimal tax system to answer this question.

Following Mankiw et al. (2009), we use a log normal as the skill distribution, parametrized with $(\mu, \sigma) = (2.757, 0.5611)$, which calibrates to US wage distribution in 2007. We also assume that $\frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\tilde{y},1-T'}} = d\tilde{y}/d\bar{y} = -0.05$ for all $\theta$. To calculate the optimal linear tax, we make a parametric assumption on $\psi(x) = \frac{(x)^{1/\mathcal{E}_{y,1-T'}+1}}{1/\mathcal{E}_{y,1-T'}+1}$ to estimate the counter-factual $y_t(\theta)$ under different $t$. Following them again, we assume $\mathcal{E}_{y,1-T'} = 0.5$ for our simulation. Furthermore, we assume that the cost of non-compliance is such that $\lambda_t = 0$ for all type, so every household reports non-negative income. Hence, the linear tax does not change the set of implementable tax system and only affects the total welfare through the marginal channel. This simplified exercise offers us a lower bound to the optimal linear rate under more realistic conditions that have households reporting zero income.

Figure 4 shows the consumption and the income marginal tax rates under different Pareto weights. The figure shows that for the linear tax rates seem to change more relative to non-linear marginal rates. While the linear tax always increases with the strength of the redistributive motive, the income marginal rate does not exhibit this monotonicity. Another stylistic feature is that the linear tax rate is strictly above the non-linear marginal tax rates, with the non-linear ones dipping below zero. This fact is partly explained by our assumption that the consumption tax is perfectly enforceable. The social planner sets relatively high linear tax to fight non-compliance, and it offsets this distortion with negative non-linear marginal tax rates.

# 7 Conclusion

We view the contribution of this paper as twofold. The first is a method to solve an optimal income tax problem when income levels are not perfectly observable. We overcome this difficulty by modifying the households' preferences that account for their optimal non-compliance behaviors. Our strategy allows us to have a model with a continuum of heterogeneous households who choose both labor supply and reported income.

Our second contribution is providing insights on how to combine tax instruments, each limited in its own way, to yield better redistributive outcomes. As seen in the characterization of the optimal tax mix, the non-linearity of the income tax is crucial in its ability to complement the rigidity of the linear consumption tax. Policy makers should take into account the presence of linear taxes in an economy when calibrating the optimal income tax schedule, which yields lower marginal tax rates. Furthermore, if we consider the non-compliance behavior on income taxes, this adjustment is smaller than what first appears to

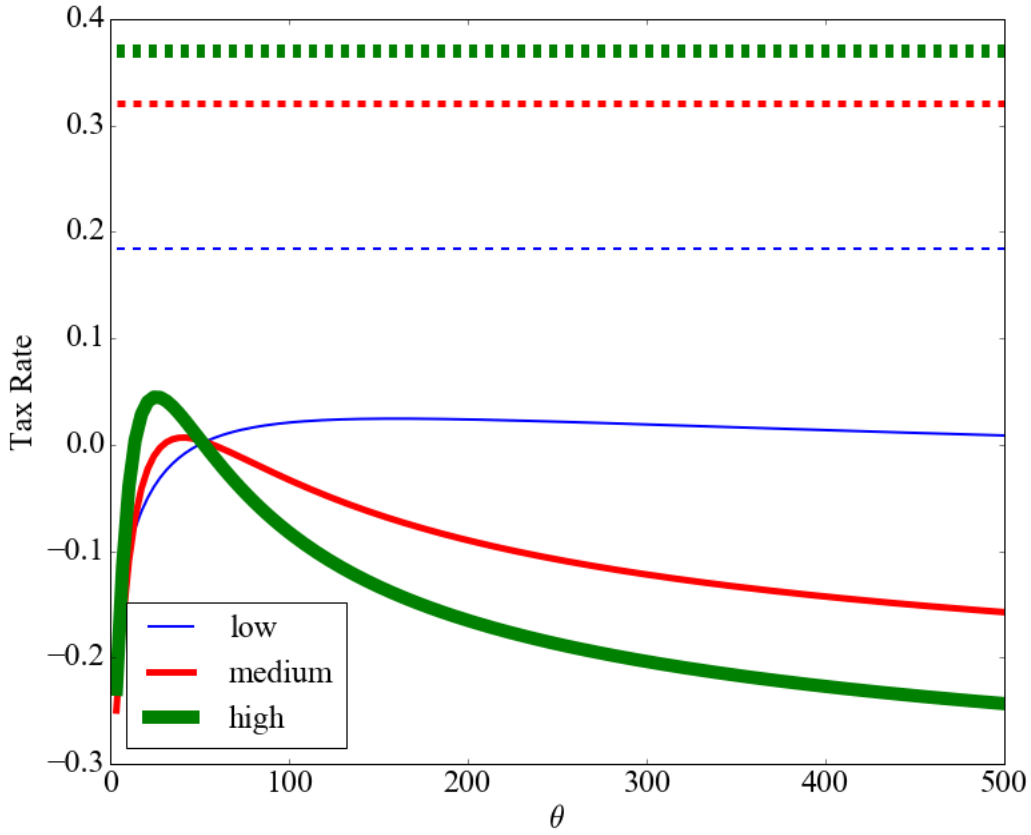


Figure 4: Optimal Tax Mix Under Different Pareto Weights

Note: The different colors represent social planners with different degrees of redistributive motives, with high meaning closer to Rawlsian and low being closer to utilitarian. We parametrize the Pareto weight, following Rothschild and Scheuer (2013), as $\tilde{F}(\theta) = (1 - (1 - F(\theta))^{\alpha})$, with $\alpha = 1.5, 1.3, 1.1$ for high, medium, and low.

be the case. This analysis also offers intuition for why developing economies with strong motives for redistribution but with weak income tax enforcement set higher consumption tax rates.

We also see the complementarity between the two instruments in our analysis of the optimal tax structure as we change the social planer's redistributive motive. As the social planner puts more weight on the lower ability households, we see the income tax becoming more progressive as expected. But in reaction to the households' increased evasion behavior in response to the higher marginal tax rates, the optimal consumption tax rate also increases.

# References


Atkinson, A. B. and Stiglitz, J. E. (1976). The Design of Tax Structure: Direct Versus Indirect Taxation. *Journal of public Economics*, 6(1):55–75.

Boadway, R., Marchand, M., and Pestieau, P. (1994). Towards a theory of the direct-indirect tax mix. *Journal of Public Economics*, 55:71 – 88.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2):31–52.

Cremer, H. and Gahvari, F. (1996). Tax evasion and the optimum general income tax. *Journal of Public Economics*, 60(2):235–249.

Gordon, G. and Li, W. (2005). Tax structure in developing countries: Many puzzles and a possible explanation. *NBER*.

Gorodnichenko, Y., Martinez-Vazquez, J., and Peter, K. S. (2009). Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia. *Journal of political economy*, 117(3):504–554.

Grochulski, B. (2007). Optimal Nonlinear Income Taxation with Costly Tax Avoidance. *Economic Quarterly-Federal Reserve Bank of Richmond*, 93(1):77.

Jacobs, B. and Boadway, R. (2014). Optimal linear commodity taxation under optimal non-linear income taxation. *Journal of Public Economics*, 117:201–210.

Keen, M. and Lockwood, B. (2010). The value added tax: Its causes and consequences. *Journal of Development Economics*, pages 138 – 151.

Kleven, H. J. and Kopczuk, W. (2011). Transfer program complexity and the take-up of social benefits. *American Economic Journal: Economic Policy*, pages 54–90.

Kopczuk, W. (2000). Redistribution when avoidance behavior is heterogeneous. *Journal of Public Economics*, pages 51–71.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics*, 89(11):2093–2119.

Kopczuk, W. (2012). The polish business "flat" tax and its effect on reported incomes: a pareto improving tax reform. Technical report, Mimeo, Columbia University.

La Porta, R. and Shleifer, A. (2008). The unofficial economy and economic development. *Brooking Papers on Economic Activity*.

Lans Bovenberg, A. and Jacobs, B. (2005). Redistribution and education subsidies are siamese twins. *Journal of Public Economics*, 89(11).

Mankiw, G. N., Weinzierl, M., and Yagan, D. (2009). Optimal taxation in theory and practice. *Journal of Economic Perspective*.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, pages 175–208.

Naritomi, J. (2013). Consumers as tax auditors. *Job market paper, Harvard University*.

Piketty, T. (1997). La redistribution fiscale face au chômage. *Revue française d'économie*, 12(1):157–201.

Piketty, T., Saez, E., Stantcheva, S., et al. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–271.

Pomeranz, D. (2013). No taxation without information: Deterrence and self-enforcement in the value added tax. Technical report, National Bureau of Economic Research.

Rothschild, C. and Scheuer, F. (2013). Redistributive taxation in the roy model. *The Quarterly Journal of Economics*, 128(2):623–668.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, 68(1):205–229.

Sandmo, A. (1981). Income tax evasion, labour supply, and the equity - efficiency tradeoff. *Journal of Public Economics*, 16(3):265–288.

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, pages 643–663.

Schroyen, F. (1997). Pareto efficient income taxation under costly monitoring. *Journal of Public Economics*, 65(3):343–366.

# A Appendix

## A.1 The Case when $\tilde{y}$ is Constrained to be Non-negative

We discuss here why if we disallow negative $\tilde{y}$, the solution equals that of Mirrlees (1971). Let us call $y(\theta)^M$ as the optimal total income in which household cannot misreport its income. More precisely, $y(\theta)^M$ is defined as

$$y(\theta)^M = \arg\max_{y \geq 0} \left\{ y - \psi\left(\frac{y}{\theta}\right) + \frac{y}{\theta^2}\psi'\left(\frac{y}{\theta}\right)\left(\frac{F(\theta) - \tilde{F}(\theta)}{f(\theta)}\right) \right\}.$$

Now we are ready to present the following proposition.

**Proposition 3.** *Suppose that constraint of the problem (Eq. (2)) is $\tilde{y} \geq 0$ rather than $\tilde{y} \geq -\bar{y}$. The optimal solution involve any $t \geq t^*$ with $t^*$ satisfying*

$$(1 - t^*) = \min_{\theta} \frac{\psi'(y(\theta)^M/\theta)}{\theta}.$$

*And the optimal allocation involves $\tilde{y}_t(\theta, \bar{y}_t(\theta)) = 0$ and $\bar{y}_t = y(\theta)^M$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.*

*Proof.* Suppose that in our optimal allocation, there exists a $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ such that $\tilde{y}_t(\hat{\theta}, \bar{y}_t(\hat{\theta})) > 0$. We can increase social welfare by first setting $t = t^*$. Note that $\tilde{y}_{t^*}(\theta, \bar{y}) = 0$ for $\bar{y} \geq 0$ and for all $\theta$. Then set new allocation as $\bar{y}_{t^*}(\theta) = \bar{y}_t(\theta) + \tilde{y}_t(\theta, \bar{y}_t(\theta))$ for all $\theta$. This new allocation increases social welfare by at least $\phi(\tilde{y}_t(\hat{\theta}, \bar{y}_t(\hat{\theta})))$. Hence the optimal allocation must involve zero income tax evasion, and therefore it must coincide with that of Mirrlees. □

## A.2 Proof of Lemma 1

*Proof.* Let $b(\bar{y}, \theta) \equiv (1 - t)\tilde{y}_t(\theta, \bar{y}) - \psi\left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\right) - \phi(\tilde{y}_t(\theta, \bar{y}))$. Using the definition of $\tilde{y}_t(\theta, \bar{y})$, we can rewrite $b(\bar{y}, \theta) = \max_{\tilde{y}}(1 - t)\tilde{y} - \psi\left(\frac{\bar{y} + \tilde{y}}{\theta}\right) - \phi(\tilde{y})$. Applying envelope theorem when differentiating with respect to $\theta$ gives us $b_\theta(\theta, \bar{y}) = \frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta^2}\psi'\left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\right)$.
($\Rightarrow$): We rewrite IC as

$$\hat{\theta} \in \operatorname*{argmax}_{\theta} \{u_t(\bar{c}_t(\hat{\theta}), \bar{y}_t(\hat{\theta}); \theta) - v_t(\theta)\},$$

where $v_t(\theta)$ is defined in Eq. (5). Since the objective function of the above expression is differentiable with respect to $\theta$, IC implies that the first order condition evaluated at $\theta = \hat{\theta}$ is:

$$v_t'(\hat{\theta}) = b_\theta(\bar{y}_t(\hat{\theta}), \hat{\theta}),$$

which is equivalent to Eq. (6).

We now show that IC implies monotonicity of $\bar{y}_t(\theta)$. For any $\theta_1 \geq \theta_0$, $v(\theta_1) - v(\theta_0) = [\bar{c}(\theta_1) - \bar{c}(\theta_0)] + [b(\bar{y}_t(\theta_1), \theta_1) - b(\bar{y}_t(\theta_0), \theta_0)]$. Hence note that:

$$b(\bar{y}_t(\theta_0), \theta_1) - b(\bar{y}_t(\theta_0), \theta_0) \leq v(\theta_1) - v(\theta_0) \leq b(\bar{y}_t(\theta_1), \theta_1) - b(\bar{y}_t(\theta_1), \theta_0)$$

where the first inequality comes from the IC of $\theta_1$ and the second from the IC of $\theta_0$. We can hence simplify this expression as follows:

$$b(\bar{y}_t(\theta_0), \theta_1) - b(\bar{y}_t(\theta_1), \theta_1) \leq \bar{c}(\theta_1) - \bar{c}(\theta_0) \leq b(\bar{y}_t(\theta_0), \theta_0) - b(\bar{y}_t(\theta_1), \theta_0)$$

Therefore:

$$\begin{aligned} b(\bar{y}_t(\theta_0), \theta_0) - b(\bar{y}_t(\theta_1), \theta_0) - [b(\bar{y}_t(\theta_0), \theta_1) - b(\bar{y}_t(\theta_1), \theta_1)] = \\ \int_{\theta_0}^{\theta_1} b_\theta(\bar{y}_t(\theta_1), x) - b_\theta(\bar{y}_t(\theta_0), x) dx \geq 0 \end{aligned}$$

By Lemma 2, we have that $\bar{y}_t(\theta_1) \geq \bar{y}_t(\theta_0)$.
($\Leftarrow$) For $\theta_0 \leq \theta_1$, we have

$$v(\theta_1) - v(\theta_0) = \int_{\theta_0}^{\theta_1} b_\theta(\bar{y}_t(x), x) dx \geq \int_{\theta_0}^{\theta_1} b_\theta(\bar{y}_t(\theta_0), x) dx = b(\bar{y}_t(\theta_0), \theta_1) - b(\bar{y}_t(\theta_0), \theta_0),$$

where the first equality follows from local incentive compatibility constraint and the inequality follows from Lemma 2 and the monotonicity of $\bar{y}_t(\theta)$. Hence,

$$\bar{c}(\theta_1) + b(\bar{y}_t(\theta_1), \theta_1) - [\bar{c}(\theta_0) + b(\bar{y}_t(\theta_0), \theta_0)] \geq b(\bar{y}_t(\theta_0), \theta_1) - b(\bar{y}_t(\theta_0), \theta_0),$$

which implies,

$$v(\theta_1) = \bar{c}(\theta_1) + b(\bar{y}_t(\theta_1), \theta_1) \geq \bar{c}(\theta_0) + b(\bar{y}_t(\theta_0), \theta_1) = u_t(\bar{c}_t(\theta_0), \bar{y}_t(\theta_0), \theta_1).$$

Similarly, we have

$$v_t(\theta_1) - v_t(\theta_0) = \int_{\theta_0}^{\theta_1} b_\theta(\bar{y}_t(x), x)dx \leq \int_{\theta_0}^{\theta_1} b_\theta(\bar{y}_t(\theta_1), x)dx = b(\bar{y}_t(\theta_1), \theta_1) - b(\bar{y}_t(\theta_1), \theta_0),$$

which implies

$$v_t(\theta_0) = \bar{c}(\theta_0) + b(\bar{y}_t(\theta_0), \theta_0) \geq \bar{c}(\theta_1) + b(\bar{y}_t(\theta_1), \theta_0) = u_t(\bar{c}_t(\theta_1), \bar{y}(\theta_1), \theta_0).$$

□

**Lemma 2.** *The preference defined in Eq.* (4) *exhibit single crossing property,* i.e. $\frac{\partial u}{\partial \theta}$ *is non-decreasing in* $\bar{y}$.

*Proof.* By the envelope theorem, we have

$$\frac{\partial u}{\partial \theta} = \frac{\tilde{y}(\theta, \bar{y}) + \bar{y}}{\theta^2} \psi' \left( \frac{\tilde{y}(\theta, \bar{y}) + \bar{y}}{\theta} \right)$$

Applying the implicit function theorem on the FOC of Eq. (3), we see that $\tilde{y}(\theta, \bar{y}) + \bar{y}$ is increasing in $\bar{y}$. Since both $\psi'()$ and total income are non-negative and $\psi'()$ is increasing, $\frac{\partial u}{\partial \theta}$ is non-decreasing in $\bar{y}$. □

## A.3 Proof of Proposition 1

*Proof.* We start by rewriting the constraint Eq. (7)

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[ \bar{y}(\theta) + \tilde{y}_t(\theta, \bar{y}(\theta)) - \psi \left( \frac{\tilde{y}_t(\theta, \bar{y}_t(\theta)) + \bar{y}_t(\theta)}{\theta} \right) - \phi \left( \tilde{y}_t(\theta, \bar{y}(\theta)) \right) - v_t(\theta) \right] f(\theta) d\theta \geq 0.$$

using Eq. (5).

Using integration by parts, we have

$$\int_{\underline{\theta}}^{\bar{\theta}} v_t(\theta) \tilde{f}(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \left[ \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left( \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left( \frac{1 - \tilde{F}(\theta)}{f(\theta)} \right) \right] f(\theta) d\theta + v(\underline{\theta}),$$

and

$$\int_{\underline{\theta}}^{\bar{\theta}} v_t(\theta) f(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \left[ \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left( \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left( \frac{1 - F(\theta)}{f(\theta)} \right) \right] f(\theta) d\theta + v(\underline{\theta}).$$

Using the above expressions, we rewrite the integrand of the objective function as[9]:

$$\begin{aligned} &\bar{y}(\theta) + \tilde{y}_t(\theta, \bar{y}(\theta)) - \psi \left( \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}}{\theta} \right) - \phi(\tilde{y}_t(\theta, \bar{y}(\theta))) \\ &+ \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left( \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left( \frac{F(\theta) - \tilde{F}(\theta)}{f(\theta)} \right). \end{aligned} \tag{14}$$

Hence, the social planner's solution involves a point-wise maximization of the above objective function. The objective function of (14) often arise in screening problems, with the first line as the household first best utility and the second line as the information rent the social planner gives to each household in order for it to reveal its type.

The maximization problem Eq. (14) also suggests that optimal allocation may involve households not complying in their tax reports, i.e. $\tilde{y}_t(\theta, \bar{y}_t(\theta)) \neq 0$.

Note that the FOC of the household problem with respect to $\tilde{y}$ implies that $(1-t) - \frac{1}{\theta}\psi'\left(\frac{\bar{y}+\tilde{y}}{\theta}\right) - \phi'(\tilde{y}) = 0$. Thus the FOC of Eq. (14) with respect to $\bar{y}$ is

$$\begin{aligned} &1 - \frac{1}{\theta}\psi'\left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\right) + \frac{\partial \tilde{y}_t}{\partial \bar{y}} t + \lambda_t(\theta) = \\ &\frac{1}{\theta^2}\left(\frac{\partial \tilde{y}_t}{\partial \bar{y}} + 1\right)\left(\frac{\tilde{F}(\theta) - F(\theta)}{f(\theta)}\right)\left(\psi'\left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\right) + \frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\psi''\left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta}\right)\right), \end{aligned} \tag{15}$$

where $\lambda_t(\theta)$ represents the Lagrange multiplier on the constraint that $\bar{y} \geq 0$.

In order to determine the optimal income tax schedule, remember that the household problem is,

$$\max_{\bar{y} \geq 0, \tilde{y} + \bar{y} \geq 0} \left[ (1-t)\left[\tilde{y} + \bar{y} - T(\bar{y})\right] - \psi\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) - \phi(\tilde{y}) \right].$$

[9]Note that the Lagrangian multiplier on the resource constraint is one because of the quasi-linear specification of the utility function.

Hence the first order condition with respect to $\bar{y}$ in the interior is

$$(1 - T'(\bar{y}))(1 - t) - \frac{1}{\theta}\psi'\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) = 0, \tag{16}$$

and the first order condition with respect to $\tilde{y}$ is

$$1 - t - \frac{1}{\theta}\psi'\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) = \phi'(\tilde{y}). \tag{17}$$

The above two first order conditions implicitly define $\bar{y}(\theta)$ and $\tilde{y}(\theta)$. Using the fact that $\frac{\partial \tilde{y}_t}{\partial \bar{y}} + 1 = 1 + \frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T'}}$ and $\frac{(1-t)(1-T')\theta^2}{y\psi''(\frac{y}{\theta})} = \mathcal{E}_{y,1-T'}$, rearranging Eq. (15) gives us the first expression.

To get the second expression, we transform the skill distribution into the reported income distribution. Differentiating Eq. (16) with respect to $\theta$ gives us,

$$\bar{y}'(\theta) = \frac{\frac{1}{\theta^2}\left[\psi'\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right) + \frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\psi''\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right)\right]}{(1-t)T''(\bar{y}(\theta)) + \frac{1}{\theta^2}\psi''\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right)\left(\frac{(1-t)T''(\bar{y}(\theta))}{\phi''(\tilde{y}(\theta))} + 1\right)},$$

which we use to transform the type pdf $f(\theta)$ into reported income pdf $h(\bar{y})$. Noting that $1 - \frac{1}{\theta}\psi'\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right) = t + T'(\bar{y}(\theta))(1-t)$, we rewrite the interior case of Eq. (15) as

$$T'(\bar{y})(1-t) + t\frac{dy}{d\bar{y}} =$$
$$\frac{dy}{d\bar{y}}\frac{\tilde{H}(\bar{y}) - H(\bar{y})}{h(\bar{y})}\left((1-t)T''(\bar{y}) + \frac{1}{\theta^2}\psi''\left(\frac{\tilde{y}+\bar{y}}{\theta}\right)\left(\frac{(1-t)T''(\bar{y})}{\phi''(\tilde{y})} + 1\right)\right). \tag{18}$$

Note that differentiating Eq. (17) with respect to $\bar{y}$ gives us $\frac{dy}{d\bar{y}} = \frac{\phi''()}{\frac{\psi''()}{\theta^2} + \phi''()} = \frac{(1-t)(1-T'(\bar{y}))}{\frac{1}{\theta^2}\psi''()\mathcal{E}_{\bar{y},1-T'}\bar{y}}$. Finally, from the equalities above and rearranging Eq. (18), we get the result.

□

## A.4 Proof of Proposition 2

*Proof.* First, we rewrite the optimal social welfare as

$$\max_t W(t),$$

or

$$\max_t \left\{ \int_{\underline{\theta}}^{\bar{\theta}} \max_{\bar{y}\geq 0} \left\{ \bar{y} + \tilde{y}_t(\theta,\bar{y}) - \psi\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right) - \phi(\tilde{y}_t(\theta,\bar{y})) \right.\right.$$
$$\left.\left. + \frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta^2}\psi'\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right)\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right) \right\} f(\theta)d\theta \right\}.$$

We introduce a new function $\tilde{x}_t(\theta,y)$, implicitly defined as the solution to $(1-t)-\frac{1}{\theta}\psi'(\frac{y}{\theta}) = \phi'(\tilde{x})$. This function $\tilde{x}_t(\theta,y)$ represents the optimal non-compliance when the household must produce total output $y$ facing commodity tax $t$. Then, we can rewrite the integrand of the objective function as

$$\max_y \left\{ y - \psi\left(\frac{y}{\theta}\right) - \phi(\tilde{x}_t(\theta,y)) + \frac{y}{\theta^2}\psi'\left(\frac{y}{\theta}\right)\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right) + \lambda_t(\theta)[y-\tilde{x}_t(\theta,y)] \right\}, \quad (19)$$

with $y_t(\theta)$ as the argmax. Using the envelope theorem and interchanging differentiation and integration, we get

$$\frac{dW(t)}{dt} = -\int_{\underline{\theta}}^{\bar{\theta}} [\phi'(\tilde{x}_t(\theta,y_t(\theta))) + \lambda_t^y(\theta)]\frac{\partial \tilde{x}_t}{\partial t}(\theta,y_t(\theta))dF(\theta), \quad (20)$$

where $\lambda_t^y(\theta)$ is the Lagrange multiplier on the constraint that $y \geq \tilde{x}_t(\theta,y)$. Notice from the FOC of Eq. (19) that:

$$-\lambda_t^y(\theta)\left(1-\frac{\partial \tilde{x}_t}{\partial y}\right) = t + \left(1-t-\frac{\psi'(y_t(\theta)/\theta)}{\theta}\right)\left(1-\frac{\partial \tilde{x}_t}{\partial y}\right) +$$
$$\frac{1}{\theta^2}\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right)\left(\psi'(y_t(\theta)/\theta) + \frac{y_t(\theta)\psi''(y_t(\theta)/\theta)}{\theta}\right),$$

and that $\left(1-\frac{\partial \tilde{x}_t}{\partial y}\right) = \frac{1}{\partial y/\partial \bar{y}}$, we see that $\lambda_t^y(\theta) = \lambda_t(\theta)$. Finally, noting that $\frac{\partial \tilde{x}_t}{\partial t} = \frac{-1}{\phi''(\tilde{x}_t)}$ by implicitly differentiating $\tilde{x}_t(\theta,y)$ with respect to $t$, setting Eq. (20) equal to 0 gives us equation Eq. (11).

To derive equation Eq. (12), note that, when the household problem has an interior solution, $\phi'(\tilde{y}_t(\theta)) = (1-t)T'(\bar{y}_t(\theta))$ from the FOC of the household problem and $\phi''(\tilde{y}_t(\theta)) = \frac{(1-t)(1-T'(\bar{y}_t(\theta)))}{y_t\theta\mathcal{E}_{y,1-T'}}$ from implicitly differentiating the FOC with respect to $\tilde{y}$. We then write

the marginal change in welfare with respect to $t$ as

$$\frac{dW'(t)}{dt} = \int \frac{\phi'(\tilde{y}(\theta,\bar{y})) + \lambda_t(\theta)}{\phi''(\tilde{y}(\theta,\bar{y}))}\Big|_{\bar{y}_t(\theta)} dF(\theta)$$
$$= \int_{\hat{\theta}}^{\bar{\theta}} \frac{T'(\bar{y}_t(\theta))}{1-T'(\bar{y}_t(\theta)} y_t(\theta)\mathcal{E}_{y,1-T'} dF(\theta) + \int_{\underline{\theta}}^{\hat{\theta}} \frac{\phi'(\tilde{y}(\theta,0)) + \lambda_t(\theta)}{\phi''(\tilde{y}(\theta,0))} dF(\theta),$$

with $\hat{\theta} = \max\{\theta : \bar{y}_t(\theta) = 0\}$. Also, let us denote

$$A_t(\theta) \equiv \left(\frac{\tilde{y}\mathcal{E}_{\tilde{y},1-T'}}{\bar{y}\mathcal{E}_{\bar{y},1-T'}} + 1\right)$$
$$B_t(\theta) \equiv \frac{\tilde{F}(\theta) - F(\theta)}{f(\theta)\theta}\left(1 + \frac{1}{\mathcal{E}_{y,1-T'}}\right).$$

Rearranging Eq. (8), we have

$$\frac{T'(\bar{y}_t(\theta))}{1-T'(\bar{y}_t(\theta))} = \frac{A_t(\theta)(B_t(\theta)(1-t)-t)}{1-t(1-A_t(\theta))}$$

Combine this with the above and setting $dW(t)/dt = 0$ gets us Eq. (12). □

## A.5 Proof of Corollary 2

*Proof.* Suppose that $\tilde{F}(\theta) \geq \tilde{F}'(\theta)$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$, so that $\tilde{F}(\theta)$ is more redistributive than $\tilde{F}'(\theta)$. Since Eq. (14) exhibit increasing difference in (-$\tilde{F}(\theta)$,$\bar{y}$), when the pareto weights change from $\tilde{F}(\theta)$ to $\tilde{F}'(\theta)$, the corresponding optimal reported income allocation does not decrease, i.e $\bar{y}_t(\theta) \leq \bar{y}'_t(\theta)$ for every type. Since misreported income is decreasing in reported income, $\tilde{y}_t(\theta, \bar{y}_t(\theta)) \geq \tilde{y}_t(\theta, \bar{y}'_t(\theta))$ for all $\theta$ and $t$. Also, if we denote $\lambda_t(\theta)$ and $\lambda'_t(\theta)$ as the Lagrange multiplier for the pareto weights of $\tilde{F}(\theta)$ and $\tilde{F}'(\theta)$, we have that $\lambda_t(\theta) \geq \lambda'_t(\theta)$. To see this, suppose that $\lambda_t < \lambda'_t(\theta)$. Since $\lambda_t(\theta) \geq 0$, we have $\lambda'_t(\theta) > 0$, which implies that both $\bar{y}'_t(\theta)$ and $\bar{y}_t(\theta)$ are 0, i.e. the optimal allocations for the two pareto weights are the same. Then, from Eq. (15), we have

$$\lambda_t(\theta) - \lambda'_t(\theta) = \frac{1}{\theta^2}\left(\frac{\tilde{F}(\theta) - \tilde{F}'(\theta)}{f(\theta)}\right)\left(\psi'(\tilde{y}_t(\theta,0)/\theta) + \frac{\tilde{y}_t(\theta,0)\psi''(\tilde{y}_t(\theta,0)/\theta)}{\theta}\right)\frac{\partial y}{\partial \bar{y}}$$
$$\geq 0,$$

which implies that $\lambda_t(\theta) \geq \lambda'_t(\theta)$, a contradiction.

Let $W(t)$ and $W'(t)$ be the social welfare and $t_{opt}$ and $t'_{opt}$ be the optimal linear tax for pareto weights $\tilde{F}(\theta)$ and $\tilde{F}'(\theta)$, respectively. Suppose by contradiction that $t_{opt} < t'_{opt}$. Since

$$\frac{dW(t)}{dt} = \int_{\underline{\theta}}^{\bar{\theta}} \left( \frac{\phi'(\tilde{y}_t(\theta, \bar{y}_t(\theta))) + \lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta, \bar{y}_t(\theta)))} \right) f(\theta) d\theta$$

and

$$\frac{dW'(t)}{dt} = \int_{\underline{\theta}}^{\bar{\theta}} \left( \frac{\phi'(\tilde{y}_t(\theta, \bar{y}'_t(\theta))) + \lambda'_t(\theta)}{\phi''(\tilde{y}_t(\theta, \bar{y}'_t(\theta)))} \right) f(\theta) d\theta \text{ for all } t,$$

we have

$$dW(t)/dt \geq dW'(t)/dt \text{ for all } t, \tag{21}$$

To see this inequality, first note from our assumption that $\phi'()/\phi''()$ is non-decreasing, and that $\tilde{y}_t(\theta, \bar{y}_t(\theta)) \geq \tilde{y}_t(\theta, \bar{y}'_t(\theta))$, so we have $\frac{\phi'(\tilde{y}_t(\theta,\bar{y}_t(\theta)))}{\phi''(\tilde{y}_t(\theta,\bar{y}_t(\theta)))} \geq \frac{\phi'(\tilde{y}_t(\theta,\bar{y}'_t(\theta)))}{\phi''(\tilde{y}_t(\theta,\bar{y}'_t(\theta)))}$ for all $\theta$ and $t$. Then notice that $\frac{\lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta,\bar{y}_t(\theta)))} \geq \frac{\lambda'_t(\theta)}{\phi''(\tilde{y}_t(\theta,\bar{y}'_t(\theta)))}$ for all $\theta$ and $t$. To see this inequality, consider the three possible cases: $\lambda_t(\theta) = \lambda'_t(\theta) = 0$, $\lambda_t(\theta) > 0$ and $\lambda'_t(\theta) = 0$, and $\lambda_t(\theta) \geq \lambda'_t(\theta) > 0$. Because we assume that $\phi''() > 0$, the first two cases satisfy the inequality trivially. And in the third case, $\bar{y}_t(\theta) = \bar{y}'_t(\theta) = 0$, so the denominators are the same.

Since we assume that $W(t)$ is single-peaked and differentiable in $t$, i.e. $dW(t)/dt$ crosses 0 once, which implies there exists $\epsilon > 0$, where for $t \in (t_{opt} - \epsilon, t_{opt})$, $dW(t)/dt > 0$ and for $t \in (t_{opt}, t + \epsilon)$, $dW(t)/dt < 0$. Also, for all $t \leq t'_{opt}$, $dW'(t) \geq 0$. However, there exists a $t'' \in (t_{opt}, t'_{opt})$ such that $dW(t'')/dt < 0$ and $dW'(t'')/dt \geq 0$, which contradicts Eq. (21). □