This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 16-017

Optimal Tax Mix with Income Tax Non-compliance

By

Jason Huang and Juan Rios

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Stanford Institute for Economic Policy Research at Stanford University supports research bearing on economic and public policy issues. The SIEPR Discussion Paper Series reports on research and policy analysis conducted by researchers affiliated with the Institute. Working papers in this series reflect the views of the authors and not necessarily those of the Stanford Institute for Economic Policy Research or Stanford University

Optimal Tax Mix with Income Tax Non-compliance*

Jason Huang †
Stanford University

Juan Rios‡
Stanford University

March 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 respond to income tax not only through labor supply but also through mis-reporting their incomes. 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.
JEL: D31, D63, D82, H21, H23, H26, I30, O21

*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*

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 73.5% of the tax revenue comes from consumption taxes, this proportion is on average 32.6% among OECD countries (OECD, 2002) and it is below 7% in the United States (Office of Management and Budget, 2008). 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 individual income, while linear taxes are impersonal and therefore do not require from the government information on how much of each good was consumed by each individual. This milder information requirement makes it easier for the government to enforce linear taxes. The second reason is the self-enforcing mechanism of consumption 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 (10)). In theory, VAT and sales tax work the same way. But the former is collected at every transaction providing tax rebates for the input purchases, while the latter only collects taxes at the very last stage. The VAT's tax rebates create self-enforcing incentives that assist the governments in preventing tax evasion. This self-enforcement mechanism has been empirically documented in different settings (see for instance Naritomi 2013 (18) and Pomeranz 2013 (21)). Finally, consumption taxes could be easier to enforce since 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. A perturbation in the linear tax rate affects welfare by changing the reported and misreported income that tax payers choose in equilibrium. Nonetheless, the effect on the households' reporting behavior is second order when an optimal non-linear tax schedule is in place through two channels. First, a decrease (increase) in the consumption tax rate increases (decreases) social welfare by expanding (reducing) the set of implementable non-linear tax schedules, since the social planner is able to implement higher (lower) marginal tax rates for the same level of mis-reporting. Second, it ambiguously affects the marginal cost of non-compliance, since the direction of this effect depends on the distribution of mis-reporting behavior across the households. More specifically, since we assume the cost of mis-reporting increases with its absolute value, the increase (decrease) in the mis-reporting level would increase (decrease) this direct welfare cost for households over-reporting their income, but would decrease (increase) for households under-reporting it. We characterize the optimum linear tax rate by setting the net effect of these two channels to zero.

Finally, the optimal linear consumption tax rate is non-decreasing in the redistributive motives 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 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, (17)). Part of this literature has addressed the possibility of evasion. Sandmo (1981, (23)) 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 (1995, (5)) incorporates tax evasion to the general 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. Alternatevely, Schroyen (1997, (25)) model a two-class economy with an official and an unofficial labour 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 (2001, (12)) considers an optimal linear tax problem in which individuals differ not only in their labor productivity but also in their cost 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 (2004, (24)) reviews this literature. More recently, Piketty et al (2014, (20)) 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, (1)) showed that if a general income tax function may be chosen by the government, no commodity tax should be employed on commodities where the utility function is separable between labor and all commodities. Boadway and Jacob (2014, (9)) consider this problem, restricting the commodity taxes to be linear and writing the optimal tax formulae 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. It is important to note that 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, Marchand and Pestieau (1994, (3)). 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 hetereogeneity in the population. This richness allows us to derive a formula for the non-linear income tax schedule as a function of recoverable elasticities and moments of income distributions. Furthermore, we provide a precise characterization of this linear tax, rather than simply arguing that a positive consumption tax is optimal in the presence of tax non-

compliance. This precision 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 behavior, common in these countries, to recommend tax policies. Best et al (2013, (2)) consider the corporate tax evasion of firms in Pakistan. They note that the presence of evasion justifies taxes on turnover instead of on profits, sacrificing production efficiency but increasing tax revenue. Gordon and Li (2009, (6)) show that if a government needs to rely on the information available from bank records in order to enforce corporate taxes, the optimal tax structure includes capital income taxes, tariffs and inflation. Kleven and Kopczuk (2011, (11)) 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 (7)) analyze the welfare costs of income tax reform in Russia, taking into account the different margins of response (real vs mis-reporting). Kopczuk (2012, (14)) 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. And in section 6, we conclude.

2 Model

We consider a unit mass of heterogenous individuals who differ in their level of labor productivity, θ, 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 $\phi''(\tilde{y}) > 0$ 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 the Appendix.

Also note that θ enters only through the labor supply and does not impact the disutility of tax non-compliance. A justification for this assumption is that when a household is caught for misreporting its income, its penalty depends on the total amount of misreported income and not on the skill level of the household's breadwinner.

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, one could interpret the linear tax as a consumption tax considering the usual pattern of linear consumption tax versus a non-linear income tax. However, we denote t as the linear tax hereafter. 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 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 θ and $\bar{y}_{t,T}(\theta)$ and $\tilde{y}_{t,T}(\theta)$ as the optimal amount of income to report and mis-report, respectively, for a given income tax schedule $T()$ and a linear tax rate 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.

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) (22). 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 Linear 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 (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 $\bar{y}$.

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

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

i.e. each type chooses to report its true type. Let $v_t(\cdot)$ be the value function of type θ, 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 (6) reflects the marginal information rent a type θ household extracts for not emulating a household less productive by an arbitrarily small amount, this formula expresses 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 reported in any equilibrium of an incentive compatible tax schedule $\bar{y}_t(\theta) + \tilde{y}_t(\theta, \bar{y}_t(\theta))$ also decreases. This decreases $v'(\theta)$ and relaxes the local incentive constraint. To see this, consider the figure 1 that displays $v_t(\theta)$ for $t_L < t_M < t_H$. As the consumption taxes increases, the relationship between the value functions and the types become flatter so that the social planner needs to pay less informational rent for higher types.

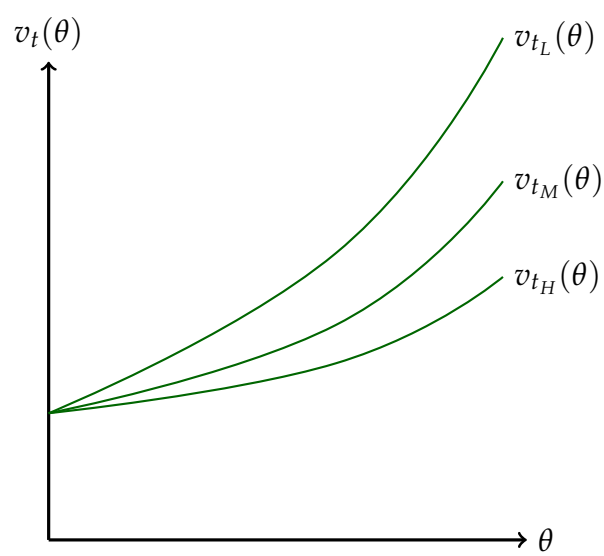

Figure 1: Value Function of Different Types

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* (4)*, an allocation is incentive compatible if and only if it satisfies equation* (6) *and* $\bar{y}(\theta)$ *is non-decreasing.*

Proof. See Appendix. □

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 θ. And since the preference is quasilinear, we have that the marginal substitution between $\bar{c}$ and $\bar{y}$ is decreasing in θ as well.

With Lemma 1, we proceed with our direct mechanism approach with the new preference (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 $d\tilde{F}(\theta)$[1] as the pareto weight, subject to (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 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*

[1] 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 $\underline{\theta}$.

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

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

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

Proof. See Appendix □

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 equation (8) is the same as Piketty (1997, (19)) and equation (9) is the same as equation (34) in Bovenberg and Jacobs (2005, (16)). However, the schedule is 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, (13)).

The optimal non-linear tax adjusts in the presence of the linear tax, since the linear tax 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 2008 (8), 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 the incomes from households in Grochulski (2008) (8) are exogenously endowed.

So far, we have characterized the optimal income tax for strictly positive reported income. We also have two comments for the case when $\bar{y}_t(\theta) = 0$. 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 quite 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 marginal tax schedules. When the marginal 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.[2]

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 (6) noted this fact when they argued 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.

[2]La Porta and Shleifer (2009) (15) 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.

4 Outer Problem: Characterizing Optimal Consumption Tax

In the previous section, we solved the inner problem by characterizing the optimal non-linear tax schedule for a given the linear tax rate. Now we solve the outer problem, i.e. we maximize the welfare with the optimal income tax in place with respect to the linear tax rate.

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

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\frac{\phi'(\tilde{y}_t(\theta, \bar{y})) + \lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta, \bar{y}))} \right) \Bigg|_{\bar{y}_t(\theta)} f(\theta) d\theta = 0, \tag{10}$$

with $\bar{y}_t(\theta)$ defined by (18) and with $\lambda_t(\theta)$ as the lagrange multiplier on the constraint $\bar{y}_t(\theta) \geq 0$.

Proof. See Appendix □

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, 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 changes the solution in two ways. First, an increase in $\tilde{y}$ increases the social cost of non-compliance by $\int \phi'(\tilde{y}(\theta))d\theta$. 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 2008, (4)). 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 α 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.

$$\begin{aligned}\bar{y}_t(\theta) = \arg\max_{\bar{y}\geq 0} \bar{y} + \tilde{y}_t(\theta,\bar{y}) - \psi\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right) - \alpha\phi(\tilde{y}_t(\theta,\bar{y})) \\ + \frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta^2}\psi'\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right)\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right). \end{aligned} \tag{11}$$

This problem is analogous to (18) found in the proof of proposition 1 with an α in front of $\phi()$ term. In the case that $\alpha = 0$, i.e. 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 productivity and the linear tax has no role. To see this more precisely, when $\alpha = 0$, the objective function of (11) 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 as long as $\bar{y} + \tilde{y}(\theta,\bar{y})$ is strictly increasing in $\bar{y}$. 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

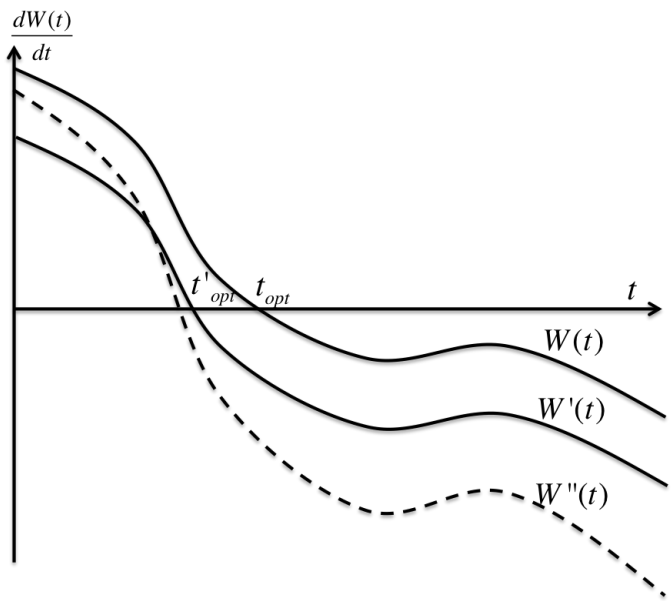

Figure 2

$\tilde{F}(\theta)$ 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 1). 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 1. *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 the appendix, 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

2. 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 $\tilde{F}'(\theta)$. Then $dW(t)/dt$ always lie above $dW'(t)/dt$, and the two curves do not cross. Furthermore, a social welfare function like $W''(t)$ cannot exist, since its derivative curve crosses the derivative curve of $W'(t)$.

Corollary 1 may seem surprising, as non-linear tax tend to be perceived as "progressive" and linear tax as "regressive." However, when social planner becomes more redistributive, the marginal 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 income will inflate less. The effect of a marginal increase in the linear tax rate on welfare, i.e. $\phi'()/\phi''()$ is positive. As a result, the social planner needs a higher linear tax rate to offset the rise in $\tilde{y}$.

6 Conclusion

We view the contribution of this paper as twofold. The first is providing a method to solve an optimal income tax problem when income levels are not perfectly observable. We overcome this difficulty by providing modified preferences of the households that account for their optimal non-compliance behavior. 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. The presence of linear taxes in an economy should be taken into account in calibrations of optimal income tax schedule, since even countries with high levels of compliance have linear taxes. This consideration requires an adjustment of the income tax schedule which yields, in general, lower marginal tax rates. However, if we take into account the non-compliance behavior on income taxes, this adjustment should be smaller than what first appears to be the case. 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 non-linear tax becoming more progressive as expected. But in reaction to the households' increased evasion behav-

ior in response to the higher marginal tax rates, the optimal linear tax rate also increases.

References

[1] Anthony Barnes Atkinson and Joseph E Stiglitz. The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, 6(1):55–75, 1976.

[2] Michael Carlos Best, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem. Production vs revenue efficiency with limited tax capacity: theory and evidence from pakistan. 2013.

[3] Marchand Maurice Boadway, Robin and Pierre Pestieau. Towards a theory of the direct-indirect tax mix. *Journal of Public Economics*, 55:71 – 88, 1994.

[4] Raj Chetty. Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance.

[5] Helmuth Cremer and Firouz Gahvari. Tax evasion and the optimum general income tax. *Journal of Public Economics*, 60(2):235–249, 1996.

[6] Gordan Gordon and We Li. Tax structure in developing countries: Many puzzles and a possible explanation. *NBER*, 2005.

[7] Yuriy Gorodnichenko, Jorge MARTINEZ-VAZQUEZ, and Klara Sabirianova PETER. 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, 2009.

[8] Borys Grochulski. Optimal nonlinear income taxation with costly tax avoidance. *Economic Quarterly-Federal Reserve Bank of Richmond*, 93(1):77, 2007.

[9] Bas Jacobs and Robin Boadway. Optimal linear commodity taxation under optimal non-linear income taxation. *Journal of Public Economics*, 117:201–210, 2014.

[10] Michael Keen and Ben Lockwood. The value added tax: Its causes and consequences. *Journal of Development Economics*, pages 138 – 151, 2010.

[11] Henrik Jacobsen Kleven and Wojciech Kopczuk. Transfer program complexity and the take-up of social benefits. *American Economic Journal: Economic Policy*, pages 54–90, 2011.

[12] Wojciech Kopczuk. Redistribution when avoidance behavior is heterogeneous. *Journal of Public Economics*, pages 51–71, 2000.

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

[14] Wojciech Kopczuk. The polish business "flat" tax and its effect on reported incomes: a pareto improving tax reform. Technical report, Mimeo, Columbia University, 2012.

[15] Rafael La Porta and Andrei Shleifer. The unofficial economy and economic development. *Brooking Papers on Economic Activity*, 2008.

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

[17] James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, pages 175–208, 1971.

[18] Joana Naritomi. Consumers as tax auditors. *Job market paper, Harvard University*, 2013.

[19] Thomas Piketty. La redistribution fiscale face au chômage. *Revue française d'économie*, 12(1):157–201, 1997.

[20] Thomas Piketty, Emmanuel Saez, Stefanie Stantcheva, et al. Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–271, 2014.

[21] Dina Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. Technical report, National Bureau of Economic Research, 2013.

[22] Casey Rothschild and Florian Scheuer. Redistributive taxation in the roy model. *The Quarterly Journal of Economics*, 128(2):623–668, 2013.

[23] Agnar Sandmo. Income tax evasion, labour supply, and the equity - efficiency trade-off. *Journal of Public Economics*, 16(3):265–288, 1981.

[24] Agnar Sandmo. The theory of tax evasion: A retrospective view. *National Tax Journal*, pages 643–663, 2005.

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

Appendix

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

$$y(\theta)^M = \arg\max_{y \geq 0} \left\{ y - \psi\left(\frac{y}{\theta}\right) + \frac{y}{\theta^2}\psi'\left(\frac{y}{\theta}\right)\left(\frac{F(\theta) - \tilde{F}(\theta)}{f(\theta)}\right) \right\}. \tag{12}$$

Now we are ready to present the following proposition.

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

$$(1 - t^*) = \min_{\theta} \frac{\psi'(y(\theta)^M/\theta)}{\theta}. \tag{13}$$

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 θ. Then set new allocation as $\bar{y}_{t^*}(\theta) = \bar{y}_t(\theta) + \tilde{y}_t(\theta, \bar{y}_t(\theta))$ for all θ. 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. □

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 θ 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

$$\hat{\theta} \in \operatorname*{argmax}_{\theta} \{u_t(\bar{c}_t(\hat{\theta}), \bar{y}_t(\hat{\theta}); \theta) - v_t(\theta)\}, \tag{14}$$

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

$$v_t'(\hat{\theta}) = b_\theta(\bar{y}_t(\hat{\theta}), \hat{\theta}), \tag{15}$$

which is equivalent to (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 θ_1 and the second from the IC of θ_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* (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 (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 $\bar{y}$. □

Proof of Proposition 1

Proof. We start by rewriting the constraint (7)

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

using (5).

Using integration by parts, we have

$$\int_{\underline{\theta}}^{\bar{\theta}} v_t(\theta) \tilde{f}(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \left[\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left(\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left(\frac{1 - \tilde{F}(\theta)}{f(\theta)} \right) \right] f(\theta) d\theta + v(\underline{\theta}), \tag{16}$$

and

$$\int_{\underline{\theta}}^{\bar{\theta}} v_t(\theta) f(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \left[\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left(\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left(\frac{1 - F(\theta)}{f(\theta)} \right) \right] f(\theta) d\theta. \\ + v(\underline{\theta}). \tag{17}$$

Using the above expressions, we rewrite the integrand of the objective function as[3]:

$$\bar{y}(\theta) + \tilde{y}_t(\theta, \bar{y}(\theta)) - \psi \left(\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}}{\theta} \right) - \phi(\tilde{y}_t(\theta, \bar{y}(\theta))) \\ + \frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta^2} \psi' \left(\frac{\tilde{y}_t(\theta, \bar{y}(\theta)) + \bar{y}(\theta)}{\theta} \right) \left(\frac{F(\theta) - \tilde{F}(\theta)}{f(\theta)} \right). \tag{18}$$

Hence, the social planner's solution involves a point-wise maximization of the above objective funciton. The objective function of (18) 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 should give to each household in order for it to reveal its type.

The maximization problem (18) 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 (18) with respect to $\bar{y}$ is

$$1 - \frac{1}{\theta}\psi' \left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta} \right) + \frac{\partial \tilde{y}_t}{\partial \bar{y}} t + \lambda_t(\theta) = \\ \frac{1}{\theta^2} \left(\frac{\partial \tilde{y}_t}{\partial \bar{y}} + 1 \right) \left(\frac{\tilde{F}(\theta) - F(\theta)}{f(\theta)} \right) \left(\psi' \left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta} \right) + \frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta} \psi'' \left(\frac{\tilde{y}_t(\theta, \bar{y}) + \bar{y}}{\theta} \right) \right), \tag{19}$$

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,

$$\max_{\bar{y} \geq 0, \tilde{y} + \bar{y} \geq 0} \left[(1-t) \left[\tilde{y} + \bar{y} - T(\bar{y}) \right] - \psi \left(\frac{\tilde{y} + \bar{y}}{\theta} \right) - \phi(\tilde{y}) \right]. \tag{20}$$

[3]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 $\bar{y}$ is

$$(1 - T'(\bar{y}))(1 - t) - \frac{1}{\theta}\psi'\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) = 0, \tag{21}$$

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

$$1 - t - \frac{1}{\theta}\psi'\left(\frac{\tilde{y} + \bar{y}}{\theta}\right) = \phi'(\tilde{y}). \tag{22}$$

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 (19) gives us the first expression.

To get the second expression, we need to transform the skill distribution into the reported income distribution. Differentiating (21) with respect to θ gives us,

$$\bar{y}'(\theta) = \frac{\frac{1}{\theta^2}\left[\psi'\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right) + \frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\psi''\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right)\right]}{(1-t)T''(\bar{y}(\theta)) + \frac{1}{\theta^2}\psi''\left(\frac{\tilde{y}(\theta)+\bar{y}(\theta)}{\theta}\right)\left(\frac{(1-t)T''(\bar{y}(\theta))}{\phi''(\tilde{y}(\theta))} + 1\right)}, \tag{23}$$

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 (19) as

$$T'(\bar{y})(1-t) + t\frac{dy}{d\bar{y}} =$$
$$\frac{dy}{d\bar{y}}\frac{\tilde{H}(\bar{y}) - H(\bar{y})}{h(\bar{y})}\left((1-t)T''(\bar{y}) + \frac{1}{\theta^2}\psi''\left(\frac{\tilde{y}+\bar{y}}{\theta}\right)\left(\frac{(1-t)T''(\bar{y})}{\phi''(\tilde{y})} + 1\right)\right). \tag{24}$$

Note that differentiating (22) 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 (24), we get the result.

□

Proof of Proposition 2

Proof. First, we rewrite the optimal social welfare as

$$\max_t W(t),$$

or

$$\max_t \left\{ \int_{\underline{\theta}}^{\bar{\theta}} \max_{\bar{y}\geq 0} \left\{ \bar{y} + \tilde{y}_t(\theta,\bar{y}) - \psi\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right) - \phi(\tilde{y}_t(\theta,\bar{y})) \right.\right.$$
$$\left.\left. + \frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta^2}\psi'\left(\frac{\tilde{y}_t(\theta,\bar{y})+\bar{y}}{\theta}\right)\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right) \right\} f(\theta)d\theta \right\}. \tag{25}$$

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

$$\max_y \left\{ y - \psi\left(\frac{y}{\theta}\right) - \phi(\tilde{x}_t(\theta,y)) + \frac{y}{\theta^2}\psi'\left(\frac{y}{\theta}\right)\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right) + \lambda_t(\theta)[y - \tilde{x}_t(\theta,y)] \right\}, \tag{26}$$

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

$$\frac{dW(t)}{dt} = -\int_{\underline{\theta}}^{\bar{\theta}} [\phi'(\tilde{x}_t(\theta, y_t(\theta))) + \lambda_t^y(\theta)]\frac{\partial \tilde{x}_t}{\partial t}(\theta, y_t(\theta))dF(\theta), \tag{27}$$

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 (26) that:

$$-\lambda_t^y(\theta)\left(1 - \frac{\partial \tilde{x}_t}{\partial y}\right) = t + \left(1 - t - \frac{\psi'(y_t(\theta)/\theta)}{\theta}\right)\left(1 - \frac{\partial \tilde{x}_t}{\partial y}\right) +$$
$$\frac{1}{\theta^2}\left(\frac{F(\theta)-\tilde{F}(\theta)}{f(\theta)}\right)\left(\psi'(y_t(\theta)/\theta) + \frac{y_t(\theta)\psi''(y_t(\theta)/\theta)}{\theta}\right), \tag{28}$$

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 (27) equal to 0 gives us the result.

□

Proof of Corollary 1

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 (18) 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 θ 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 (19), we have

$$\lambda_t(\theta) - \lambda'_t(\theta) = \frac{1}{\theta^2}\left(\frac{\tilde{F}(\theta) - \tilde{F}'(\theta)}{f(\theta)}\right)\left(\psi'(\tilde{y}_t(\theta,0)/\theta) + \frac{\tilde{y}_t(\theta,0)\psi''(\tilde{y}_t(\theta,0)/\theta)}{\theta}\right)\frac{\partial y}{\partial \bar{y}} \geq 0, \quad (29)$$

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

$$\frac{dW(t)}{dt} = \int_{\underline{\theta}}^{\bar{\theta}} \left(\frac{\phi'(\tilde{y}_t(\theta, \bar{y}_t(\theta))) + \lambda_t(\theta)}{\phi''(\tilde{y}_t(\theta, \bar{y}_t(\theta)))}\right) f(\theta) d\theta \quad (30)$$

and

$$\frac{dW'(t)}{dt} = \int_{\underline{\theta}}^{\bar{\theta}} \left(\frac{\phi'(\tilde{y}_t(\theta, \bar{y}'_t(\theta))) + \lambda'_t(\theta)}{\phi''(\tilde{y}_t(\theta, \bar{y}'_t(\theta)))}\right) f(\theta) d\theta \text{ for all } t, \quad (31)$$

we have

$$dW(t)/dt \geq dW'(t)/dt \text{ for all } t, \quad (32)$$

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 θ 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 θ 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 (32).