Discussion Paper Series

RIEB

Kobe University

DP2024-18

Optimal Income Taxation and Formalization of the Informal Economy*

RIEB Junior Research Fellow

Hirofumi TAKIKAWA

June 19, 2024

* This Discussion Paper won the Kanematsu Prize (FY 2023).

Research Institute for Economics and Business Administration

Kobe University

2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

Optimal Income Taxation and Formalization of the Informal Economy

Hirofumi Takikawa*

Goethe University Frankfurt

June 18, 2024

Abstract

The United Nations' "2030 Agenda for Sustainable Development" highlights the importance of formalizing the informal economy, which could potentially increase tax revenues in developing countries. This paper investigates the impact of formalization on optimal tax schedules, emphasizing the need for redistributive incentives alongside formalization. Extending the Mirrlees model to incorporate government intervention against the informal economy, we propose an optimal tax formula. Quantitative analysis shows that aligning the tax schedule with formalization increases tax revenue and income transfers while maintaining social welfare. The result can be interpreted as an implicit cost of welfare-neutral formalization in terms of tax revenues and income transfers. Conversely, leaving the tax schedule unchanged undermines these benefits. This research provides insights into the design of optimal tax policies that incorporate formalization.

Keywords: informal economy, formalization, income tax, redistribution

JEL codes: E26, H21, H26, J46, O17

*Email: takikawa@wiwi.uni-frankfurt.de Address: Goethe University Frankfurt, Faculty of Economics and Business Administration, Theodor-W.-Adorno-Platz 4, 60629 Frankfurt, Germany. I am grateful to Alfons J. Weichenrieder for his supervision and continuous support. I would also like to thank Hitoshi Tsujiyama, Eren Gürer, Karl Schulz, Motohiro Sato, Enda Patrick Hargaden, Tim Lee, Gabriel Ulyssea, and Krzysztof Szczygielski for valuable and helpful comments. This paper was awarded the Best Paper Prize for the Warsaw International Economic Meeting 2023, and received the Kanematsu Prize and the Nishijima Prize in 2023 by the Research Institute for Economics and Business Administration at Kobe University. All errors remain my own.

1. Introduction

The United Nations emphasizes formalization of the informal economy in the "2030 Agenda for Sustainable Development" adopted in 2015. The informal economy[1] operates largely outside the tax system, allowing agents to earn untaxed income and thus playing a limited role in tax collection. This characteristic makes it an attractive target for increasing tax revenues, particularly in developing countries. With informal employment rates often exceeding 80% in these regions (Gaspar et al., 2019)[2] and personal income taxes accounting for only 12% of total tax revenue, compared to 25% in advanced economies (Benedek et al., 2022), there is significant potential to increase government revenue by formalizing the informal economy. However, based on the data, reducing the share of informal employment in total employment does not necessarily increase the share of personal income taxes[3], although there are studies that expect positive effects of formalization[4]. Since formalization offers the government an opportunity to broaden the tax base and intuitively increases tax revenues through personal income taxes, there are still unclear mechanisms of formalization that are not explained by the current framework.

To fill this gap and to show whether formalization contributes to an increase in the personal income tax share, this paper examines formalization within the optimal income tax framework. However, there is an issue that need to be considered when the government formalizes the informal economy. Since untaxed income from the informal economy serves as a safety net for low income people, appropriate redistribution during formalization is also needed to avoid concentrating the burden of formalization on the poor. Otherwise, there would be welfare losses due to institutional constraints of the government, which cannot distinguish between previously formal and informal workers, and formalization would simply impose its burden on those who need government support. ILO (2019) emphasizes the

[1]The terminology of "informality" was first introduced in the 1970s by Hart (1973) and ILO (1972), and "Recommendation No. 204" by ILO (2015) defines the informal economy as "all economic activities by workers and economic units that are – in law or in practice – not covered or insufficiently covered by formal arrangements; and does not cover illicit activities".

[2]From an economic size perspective, Medina and Schneider (2019) estimates that the size of the informal economy is, on average, about 20% of off official GDP in developed countries, but about 40% in developing countries, and in some cases it can reach more than 60%.

[3]In Appendix, Figure D.1 compares the change in the informal employment rate from 2011 to 2019 with the change in the personal income tax share of tax revenue over the same period. Despite the decline in the informal employment rate, Colombia, Bosnia and Herzegovina and Senegal experienced a fall in the personal income tax share.

[4]Schneider and Klinglmair (2004) and Schneider et al. (2010) anticipate that the informal economy could generate significant tax revenues, in contrast to the views of Keen (2012) and OECD/ILO (2019), which perceive its impact as limited.

need for sufficient redistributive incentives to facilitate a smooth transition to the formal economy, and redistribution for low income people must be considered simultaneously with formalization. For this purpose, I employ the framework pioneered by Mirrlees (1971) because it considers optimal income taxation and redistribution under minimal constraints, i.e., incentive compatibility constraints of workers and the budget constraint of the government, without assuming any functional form of income tax. Therefore, this paper characterizes how formalization affects an optimal tax schedule and, based on the theoretical framework, simulation studies quantitatively show the effectiveness of formalization on tax revenue and income transfers as redistributive incentives.

More specifically, this paper extends the Mirrlees model to incorporate the informal economy by considering both untaxed informal income and taxed formal income, as in Doligalski and Rojas (2023). However, since Doligalski and Rojas (2023) does not consider any government intervention against the informal economy and characterizes an optimal income tax schedule without any restrictions on it, my paper allows the government to impose a joint income tax on both formal and informal income when the informal economy is formalized. Thus, agents choose to work in the formal economy, the informal economy, or both, as their utilities are maximized, and combinations of their labor supply may differ when the informal economy is formalized and when it is not. Moreover, the the government treats both cases as equivalent and imposes a common income tax schedule on agents' incomes, as the highest social welfare is obtained.

My main theoretical result is an optimal tax formula with formalization of the informal economy, and it is expressed as a modified version of the well-known ABC formula. In other words, the optimal tax formula can be decomposed into the elasticity and efficiency term, the thickness of the right tale of the skill distribution, and the desire for redistribution. In the model, agents adhere to different optimality conditions of their income choices, depending on whether their informal income is taxed or untaxed. If their informal job is not formalized, they can earn from both economies and their informal income is not taxed. On the other hand, if their informal job is subject to formalization and informal income is taxed, they will choose either the formal economy or the informal economy, where they will earn a higher wage. I then consider an optimal tax formula based on the tax perturbation approach, assuming a perturbation in the marginal tax rate at an arbitrary income level and determining the tax schedule to offset the effects of the perturbation. Since the perturbation changes agents' income choices and labor supply, the optimal tax formula with formalization of the informal economy includes agents' behavior such as labor supply adjustment within the same

economy as intensive margin responses, labor supply shift to the other economy as extensive margin responses, and direct effects of marginal tax changes as mechanical and welfare effects. The overall impact of formalization appears through changes in these responses and effects, and is expressed as the modified ABC formula.

Fuerthermore, the optimal tax formula of this paper depends on a level of formalization because agents differ in their labor supply to the formal and informal economy when their informal job is subject to formalization or when it is not. This would change agents' responses to the perturbation in the marginal tax rate, and the difference becomes larger when agents can earn untaxed income from the informal economy. This means that formalization has a larger impact on marginal tax rates for low income people, who have a greater incentive to work in the informal economy and earn untaxed income, and for those who originally worked only in the formal economy, formalization is irrelevant.

My quantitative analysis uses the parameters of Doligalski and Rojas (2023), based on Colombian household data, and compares the optimal tax schedules when the informal economy is not formalized and when 30% of it is formalized[5]. In both cases, the optimal marginal tax rates start from a relatively low level and gradually increase to the highest level for the high income group. However, compared to the case of no formalization, 30% formalization requires a more U-shaped tax schedule and imposes a 15% higher marginal tax rate on the lowest income level. This is due to the fact that in the equilibrium with formalization, the lowest income level receives more than 1.5 times the amount of the income transfer, and the reduction in marginal tax rates has a phasing-out effect for those receiving income transfers, encouraging them to work more. The progressivity of the tax schedule is similar, and slightly higher tax rates are imposed at all but the lowest income levels when the informal economy is formalized. Moreover, if the government formalizes the informal economy and adjusts its tax schedule to be more U-shaped, formalization increases total tax revenue by 1.19% while slightly improving social welfare. The additional tax revenue can be used to increase income transfers to low income people and to mitigate the negative impact of formalization on welfare. Such an increase in the total tax revenue can be interpreted as an implicit cost of 30% formalization in order to maintain at least the same level of social welfare.

As counterfactual analyses, this paper examines the impacts of formalization (i) when marginal and average tax rates are fixed and not adjusted to the level of formalization, and (ii) when only marginal tax rates are fixed but average tax rates are optimized along with the

[5]Figure D.1 shows that Uruguay, Chile, and Bosnia and Herzegovina managed to reduce the informal employment rate by approximately 30% between 2011 and 2019. Thus, 30% is exogenously given, but it can be justified as a possible rate of reduction.

level of formalization. First, when the government fixes marginal and average tax rates at the level of no formalization, formalization increases total tax revenue by 0.90% but worsens social welfare by 1.22%. In this case, formalization simply reduces the number of transfer recipients and imposes income taxes on those who previously relied on untaxed income in the informal economy. Thus, the burden of formalization is concentrated on these low income individuals, resulting in lower social welfare. Second, if the government fixes marginal tax rates but adjusts the actual tax burden according to the level of formalization, it reduces total tax revenue by 0.88% while maintaining social welfare. Since lower average tax rates at all income levels are necessary to maintain social welfare, formalization reduces total tax revenue contrary to what is expected. Thus, to increase tax revenue without worsening social welfare, the government must optimize a tax schedule along with formalization. Since many countries use similar tax schedules and try to formalize the informal economy in terms of tax collection, it is an intuitive fact that formalization may lead to lower social welfare or lower tax revenue under some conditions, and that more redistribution is required with formalization in the equilibrium.

Related literature. With respect to the model of optimal income taxation with the informal economy, this paper could be considered as an extension of optimal income taxation with multidimensional heterogeneity. For example, Rothschild and Scheuer (2013) applies the Roy model in labor economics to optimal income taxation and considers sectoral choices in the Mirrlees framework. The subsequent work of Rothschild and Scheuer (2015) generalizes the model to a finite number of sectors, rather than just two, and allows simultaneous work across sectors. Both papers endogenize the wage rates by the total amount of labor input in each sector. However, in my work, I assume exogenously given wage functions for formal and informal wages, respectively. My paper focuses primarily on the behavior of agents between taxed and untaxed income, not on wage differences between two equivalent sectors. The wage is not the only determinant of the labor market, and the income tax is also a driver of sectoral choice, since one of the markets is an untaxed source of income, contrary to Rothschild and Scheuer (2013, 2015).

To the best of my knowledge, Doligalski and Rojas (2023) and da Costa and Lobel (2022) are the only papers that apply informality directly to the Mirrlees framework. Both can be considered as an extension along the literature of multidimensional heterogeneity and treat one sector as the untaxed informal economy. Doligalski and Rojas (2023), which I basically follow, assumes participation costs in the informal economy and allow agents to work in either or both economies simultaneously. On the other hand, da Costa and Lobel

(2022) assumes participation costs in the formal economy, which is opposite to Doligalski and Rojas (2023). Then, da Costa and Lobel (2022) adopts an inverse-optimum procedure to recover the Pareto weights that rationalize the Brazilian tax schedule, and analyzes the welfare effects when informality is fully eliminated. Compared to these studies, my paper assumes government intervention against the informal economy and partial formalization of it. Following Doligalski and Rojas (2023), agents are allowed to work in both economies simultaneously if it is optimal, and an implicit cost of formalization could be interpreted as the trade-off between tax revenue and social welfare under formalization.

Furthermore, Saez and Piketty (2013) assumes linear taxes on two different tax bases and consider income shifting between the two, and Selin and Simula (2020) extends Saez and Piketty (2013) for nonlinear taxes. Both works could treat the second tax base as the informal economy by imposing a zero tax on it, but they do not consider simultaneous work across sectors. Rothschild and Scheuer (2016) also considers rent-seeking as an additional source of income, but this assumes top earners, not low income groups that correspond to those in the informal economy. In addition, Beaudry et al. (2009) considers the informal economy in the Mirrlees framework and relaxes the assumption of unobservability of formal labor supply to treat informal income as additional private information in the model. They assume that both formal hours and formal income are observable to the government and that informal income is added to the model as unobservable and untaxed income. This is more related to the discussion of work requirements and optimal income transfers.

From a different perspective, Allingham and Sandmo (1972) and subsequent work such as Chander and Wilde (1998) and Slemrod and Yitzhaki (2002) examine government interventions aimed at improving tax enforcement. They study the optimal combination of audit probability and penalty size, given a fixed tax rate. In these papers, taxpayers may not report the true value of their income, but they are penalized if the government detects such evasion with some probability. They focus more on an optimal policy structure against tax evasion and simplify a labor market structure and labor supply decisions. Although my paper does not consider a penalty for participating in the informal economy, it could be a policy instrument to discourage agents from working informally.

Structure of the paper. In the following section, I introduce the model setting and the optimal income choices of agents for a given income tax. In Section 3, I derive the optimal tax formula and show that the formalization of the informal economy affects the optimal tax rates. Section 4 is devoted to the quantitative analysis of my theoretical results. In the last section, I conclude the arguments. Proofs, detailed derivations, and additional

figures appear in the appendix.

2. Model

As in the typical Mirrlees model, there is asymmetric information between the government and agents. This means that in the model the government observes only income from the formal economy, but productivity and hours worked in the formal economy, as well as everything in the informal economy, are unobservable to the government. To study the impact of formalizing the informal economy, this paper assumes that the government formalizes a certain fraction $\pi \in [0, 1]$ of the informal economy, e.g., formalizes 30% of the informal economy and reduces it to 70% of its current level, and the income from both the formal and the informal economy is taxed by the government. In other words, π means a fraction of agents affected by formalization, and the rest of agents are irrelevant to formalization and only formal income is still taxed.

In this paper, the level of formalization π is set exogenously, and agents optimize their behavior based on the given level of formalization. In practice, when the government formalizes the informal economy, it is very difficult to formalize specific agents in the informal economy. Even if the government develops a regulatory framework or improves legal compliance to formalize the informal economy, it is not clear which agents will be subject to formalization when the government proposes formalization policies. This is consistent with the assumption of unobservable productivity and formalization status in the formal and informal economy.

In addition, this paper assumes that the informal economy is formalized at no cost, since the ILO and other institutions provide financial support for formalization. In other words, formalization is difficult to be implemented if there is no external financial source. According to ILO (2019), nearly 90% of their projects have budgets of over 1 million USD, specifically 62% between 1 million and 5 million USD and 27% over 5 million USD[6]. Additional funding is needed to continue the project for further formalization, which supports costless formalization in the model. Although the model can be extended to account for formalization costs, this only tightens the government budget constraint, but has no effect on an optimal tax formula itself.

[6]ILO (2019) collects as many reports as possible, including independent evaluation reports and internal reports. It then covers a total of 38 reports, some of which reflect multiple projects in one report. Small budget projects may not be included in the 38 reports because sometimes no reports are generated.

2.1 Environment

This paper follows and extends the setup of Doligalski and Rojas (2023), which allows agents to work in the formal economy, the informal economy, or both simultaneously, and proposes an optimal income tax schedule in the Mirrlees framework. To account for the impact of formalizing the informal economy on an optimal income tax schedule, I consider the behavior of agents separately when the informal economy is not formalized and when it is. I then present an optimal tax formula with the partially formalized informal economy.

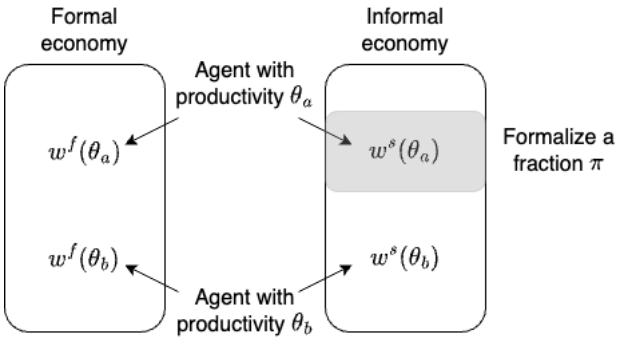
There is a continuum of agents with heterogeneity in productivity $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, which is privately observed by an agent and follows the twice continuously differentiable distribution function $F(\theta)$ with density $f(\theta)$. Given productivity type θ, the wage rate in the formal economy is determined by the function $w^f(\theta) : \Theta \to \mathbb{R}^+$ and that in the informal economy is determined by $w^s(\theta) : \Theta \to \mathbb{R}^+$. The wage rates in both economies are nonnegative, strictly increasing and continuously differentiable in θ, and satisfy the single crossing condition, i.e., $\frac{w^s(\theta)}{w^f(\theta)}$ is strictly decreasing in θ. That is, the comparative advantage in the informal economy declines with productivity. This assumption helps to track agents' responses to tax changes by allowing formal income to rise with productivity. Other empirical evidence also supports that informal wages can be higher at low productivity levels, as the single crossing condition holds in the Colombian data used for the quantitative analysis[7].

To explore the impacts of formalization, I examine how agents interact in the formal and informal economies. The model assumes that agents, each at their productivity level θ, receive fixed wages in both the formal and informal economies, $w^f(\theta)$ and $w^s(\theta)$, respectively. Job type and wage depend on a worker's productivity, not formalization status. Since the wage reflects an agent's productivity and suitability for a particular job, formal and informal wages are skill-specific and independent of formalization. Furthermore, in order to keep agent behavior tractable, this paper does not assume search frictions in the formal and informal economies. That is, an informal worker is not allowed to search for another informal job in order to remain in the informal economy, even if her informal job is formalized by the government.

Assuming asymmetric information, agents know their wage rates in both economies and determine their labor supply based on these wages, taxes, and formalization status. For example, as shown in Figure 1, an agent with productivity θ_a earns $w^f(\theta_a)$ from a formal

[7]Doligalski and Rojas (2023) uses Colombian household survey data to estimate formal and informal wage functions that satisfy the single crossing condition. Pratap and Quintin (2006) also show that low productivity workers earn higher wages in the informal economy in Argentina.

Figure 1: Image of formal and informal wages for agents with different productivity levels



Note: An agent with productivity θ_a is assigned $w^f(\theta_a)$ in the formal economy and $w^s(\theta_a)$ in the informal economy, and an agent with $\theta_b \neq \theta_a$ follows in the same way. The gray shaded area means that a fraction π of the informal economy is formalized and treated as formal.

job like working in a factory and $w^s(\theta_a)$ from an informal job like farming. If some fraction π of the informal economy is formalized and the informal job is subject to the formalization, the agent with θ_a adjusts her labor supply based on her formal wage and her formalized and taxable informal wage. In other words, if an informal job is excluded from the formalization and still treated as informal, like the agent with productivity θ_b in Figure 1, she determines her labor supply based on her formal wage and untaxable informal wage. The choice of labor supply is different when the informal economy is formalized than when it remains informal.

In addition to the productivity, agents are also heterogeneous in their privately observed participation cost in the informal economy, κ. I assume that an agent incurs idiosyncratic fixed cost $\kappa \in \mathbb{R}_+$ following $G_\theta(\kappa)$ with density $g_\theta(\kappa)$. The cost of participation causes the disutility of working in the informal economy when she works in the informal economy. Since κ can be interpreted as technological constraints on tax avoidance or the disutility of feeling guilty about informality[8], there is no cost to the agent if her formal job is subject to formalization and she chooses either her formal job or her formalized informal job. The distributions $G_\theta(\kappa)$ is conditional on θ, and twice continuously differentiable.

Agents have identical preferences over consumption c and total labor supply n by the strictly concave utility function $U(c, n)$. Total labor supply is decomposed into formal labor

[8]This paper does not consider a fixed cost of formal employment because the literature suggests that entry costs into the formal economy do not encourage informality. Informality in different situations, such as the studies by Pratap and Quintin (2006) for Argentina, Rocha et al. (2018) for Brazil, and De Mel et al. (2013) for Sri Lanka, have shown that the decision to work informally is not primarily motivated by the costs associated with joining the formal economy.

supply n^f and informal labor supply n^s. The formal and informal incomes of an agent with productivity θ are then given by $y^f = w^f(\theta)n^f$ and $y^s = w^s(\theta)n^s$, respectively. More specifically, I assume the following separable utility function:

$$U(c, n) = c - v(n) \tag{1}$$

where $v(n)$ is increasing and strictly convex in n.

The government imposes a single income tax $T(y)$ on a taxable income level y, and $T(\cdot)$ is nonlinear and continuously differentiable, but does not make any functional form assumptions. $T(\cdot) > 0$ means a positive income tax, but $T(\cdot) < 0$ denotes an income transfer. In particular, $T(0)$ is the income transfer to the unemployed.

2.2 Income choices without formalization

I first consider the case where the informal economy is not formalized and informal income is not taxed by the government. In this case, consumption c_1 is defined as $c_1 = y_1^f + y_1^s - T(y_1^f)$, and total labor supply n_1 can be expressed as $n_1 = n_1^f + n_1^s = \frac{y_1^f}{w^f(\theta)} + \frac{y_1^s}{w^s(\theta)}$. As long as $y_1^f = 0$, the agent is considered unemployed and entitled to receive the full amount of the income transfer $T(0)$. In other words, informal income y_1^s can be treated as additional income excluded from the tax base. This is in line with OECD/ILO (2019), which shows that public transfers constitute part of the total income of informal households and can be an important element of income sources. In the literature, informality is defined as the absence of contributions to the social security system, not social assistance financed by taxes.

Moreover, working in the informal economy causes the disutility κ. To distinguish agents in terms of their participation costs, I define high-cost workers as those with the costs above or equal to a threshold $\tilde{\kappa}(\theta)$ and low-cost workers as those with the costs below the threshold. When the informal economy is not formalized, high-cost workers always work in the formal economy because they face greater disutility from working in the informal economy. Low-cost workers can benefit from working in the informal economy, and they can choose to work in the formal economy, the informal economy, or both economies. Because of the costs of participating in the informal economy, some agents may choose to work in the formal economy, while others may choose to work in the informal economy, even if their productivity levels are the same.

Given an income tax $T(\cdot)$, an agent with (θ, κ) chooses her formal and informal income

that maximizes her utility.

$$V_1(\theta,\kappa) = \max_{y_1^f \geq 0,\, y_1^s \geq 0} c_1(\theta,\kappa) - v\left(\frac{y_1^f(\theta,\kappa)}{w^f(\theta)} + \frac{y_1^s(\theta,\kappa)}{w^s(\theta)}\right) - \kappa \mathbb{1}_{y_1^s>0} \tag{2}$$

where $V_1(\theta,\kappa)$ is the indirect utility if the informal economy is not formalized. The participation cost threshold $\tilde{\kappa}(\theta)$ is defined as $\tilde{\kappa}(\theta) = V_1(\theta,\kappa \mid \kappa = 0) - V_1(\theta,\kappa \mid \kappa = \infty)$.

I introduce the following assumptions in order to have a clear characterization of income choices.

Assumption 1. *If an agent with (θ,κ) is indifferent between the formal economy and the informal economy, she will choose to work in the formal economy. Furthermore, if she is indifferent between more than one level of income, she will choose the highest level.*

To simplify the notation, $\overline{y}_1^f(\theta)$ denotes the formal income of an agent with (θ,κ) if she participates only in the formal economy regardless of her cost type, and $\underline{y}_1^f(\theta)$ denotes the formal income of an agent with (θ,κ) if she is a low-cost worker and participates in both economies simultaneously. The notation for informal income follows the same rule, so $\overline{y}_1^s(\theta)$ is the informal income if an agent with (θ,κ) works only in the informal economy, and $\underline{y}_1^s(\theta)$ if she works in both economies. This allows us to consider that income choices are independent of participation costs κ, depending only on productivity θ.

Agents' income choices follow the first-order condition of the maximization problem (2). If an agent with (θ,κ) is a high-cost worker and works only in the formal economy, i.e. $y_1^f(\theta,\kappa) = \overline{y}_1^f(\theta) > 0$ and $y_1^s(\theta,\kappa) = 0$, then the optimality condition for $\overline{y}_1^f(\theta)$ is defined as follows.

$$\left[1 - T'\left(\overline{y}_1^f(\theta)\right)\right] w^f(\theta) = v'\left(\frac{\overline{y}_1^f(\theta)}{w^f(\theta)}\right) \tag{3}$$

In the optimality condition (3), the marginal return to formal labor on the left hand side is equal to the marginal disutility of labor supply on the right hand side. Thus, given the marginal tax rate $T'(\cdot)$, the formal income is determined by the formal wage and the marginal disutility of labor supply. When an agent with productivity θ is a low-cost worker and works only in the formal economy, the agent has the same formal wage rate $w^f(\theta)$ as the high-cost worker and her formal income follows the same condition (3).

Similarly, if an agent with (θ,κ) is a low-cost worker and works only in the informal economy, i.e. $y_1^f(\theta,\kappa) = 0$ and $y_1^s(\theta,\kappa) = \overline{y}_1^s(\theta) > 0$, the optimal $\overline{y}_1^s(\theta)$ is determined by the

following equation.

$$w^s(\theta) = v'\left(\frac{\overline{y}_1^s(\theta)}{w^s(\theta)}\right) \tag{4}$$

Since the marginal disutility of labor supply is fixed by the informal wage, the level of informal income is independent of other variables.

Finally, if a low-cost worker with (θ, κ) works simultaneously in the formal and informal economies, i.e. $y_1^f(\theta,\kappa) = \underline{y}_1^f(\theta) > 0$ and $y_1^s(\theta,\kappa) = \underline{y}_1^s(\theta) > 0$, the first-order condition shows the following equation.

$$\left[1 - T'\left(\underline{y}_1^f(\theta)\right)\right] w^f(\theta) = v'\left(\frac{\underline{y}_1^f(\theta)}{w^f(\theta)} + \frac{\underline{y}_1^s(\theta)}{w^s(\theta)}\right) = w^s(\theta) \tag{5}$$

The total labor supply is fixed by the informal wage, as in the fully informal case. Thus, an income tax only affects the labor market choice between the formal and the informal economy. Moreover, the optimality condition (5) implies $T'\left(\underline{y}_1^f(\theta)\right) = 1 - \dfrac{w^s(\theta)}{w^f(\theta)}$, where the right-hand side is strictly increasing in θ by the single crossing condition. Thus, a low-cost worker participates in both economies only if a tax function is progressive[9]. That is, a marginal increase in the income tax reduces the incentive for a low-cost worker to remain in the formal economy and encourages her to increase her labor supply to the informal economy.

Following the optimality conditions (3) to (5), the income choice of an agent with (θ, κ) can be summarized as follows:

$$\left(y_2^f(\theta,\kappa), y_2^s(\theta,\kappa)\right) = \begin{cases} \left(\overline{y}_2^f(\theta), 0\right) & \text{if} \quad \kappa \geq \tilde{\kappa}(\theta) \\ \left(y_2^f(\theta), y_2^s(\theta)\right) & \text{otherwise} \end{cases} \tag{6}$$

$$\text{where} \quad \left(y_2^f(\theta), y_2^s(\theta)\right) = \begin{cases} \left(\overline{y}_2^f(\theta), 0\right) & \text{if} \quad y_2^s(\theta) = 0 \\ (0, \overline{y}_2^s(\theta)) & \text{if} \quad y_2^f(\theta) = 0 \\ \left(\underline{y}_2^f(\theta), \underline{y}_2^s(\theta)\right) & \text{otherwise} \end{cases}$$

Technically, an agent with (θ, κ) chooses her utility maximizing income choice, or combination of formal and informal income, based on a given income tax schedule and her formal and informal wage rates. In the following section, the government assumes the behavior of such agents and determines an optimal tax schedule that maximizes social welfare.

[9]For more details on the discontinuity of the formal income schedule, see Doligalski and Rojas (2023).

2.3 Income choices with formalization

I follow the same steps as in the previous section. When an agent's informal job is formalized, informal income is taxed and consumption c_2 is defined as $c_2 = y_2^f + y_2^s - T(y_2^f + y_2^s)$. Since the informal economy can be considered as the second sector of the labor market in this case, there is no cost to participate in the formalized informal economy. Given an income tax $T(\cdot)$, an agent with (θ, κ) chooses her formal and informal income that maximizes her utility.

$$V_2(\theta, \kappa) = \max_{y_2^f \geq 0,\, y_2^s \geq 0} c_2(\theta, \kappa) - v\left(\frac{y_2^f(\theta, \kappa)}{w^f(\theta} + \frac{y_2^s(\theta, \kappa)}{w^s(\theta}\right) \tag{7}$$

where $V_2(\theta, \kappa)$ is the indirect utility function. Following the same notation rule as in the previous section, the formal income of an agent with (θ, κ) is denoted by $\overline{y}_2^f(\theta)$ if she works only in the formal economy, and by $\underline{y}_2^f(\theta)$ if she works in both the formal and informal economies. Informal income $y_2^s(\theta, \kappa)$ is denoted by $\overline{y}_2^s(\theta)$ if the agent works only in the formalized informal economy, and $\underline{y}_2^s(\theta)$ if she works in both economies.

An agent's optimal income choices follow the first-order condition of the maximization problem (7) when her informal job is formalized. If an agent with (θ, κ) is a high-cost worker and works only in the formal economy, i.e. $y_2^f(\theta, \kappa) = \overline{y}_2^f(\theta) > 0$ and $y_2^s(\theta, \kappa) = 0$, the optimal $\overline{y}_1^f(\theta)$ is pinned down by the following equation.

$$\left[1 - T'\left(\overline{y}_2^f(\theta)\right)\right] w^f(\theta) = v'\left(\frac{\overline{y}_2^f(\theta, \kappa)}{w^f(\theta}\right) \tag{8}$$

When an agent with productivity θ is a low-cost worker and works only in the formal economy, the agent has the same formal wage rate $w^f(\theta)$ as the high-cost worker and her formal income follows the same condition (8). As long as agents work only in the formal economy, the first-order condition has the same form as condition (3).

However, if a low-cost worker chooses to work in the informal economy, the optimality condition becomes differentr from condition (4) because informal income is also taxed. If an agent with (θ, κ) works only in the informal economy, i.e., $y_2^f(\theta, \kappa) = 0$ and $y_2^s(\theta, \kappa) = \overline{y}_2^s(\theta) > 0$, the optimal $\overline{y}_2^s(\theta)$ is determined by the following equation.

$$\left[1 - T'\left(\overline{y}_2^s(\theta)\right)\right] w^s(\theta) = v'\left(\frac{\overline{y}_2^s(\theta, \kappa)}{w^f s\theta}\right) \tag{9}$$

The optimality condition (9) is similar to equation (8) because the formal and informal

incomes are taxed and follow the same tax function.

For the case that an agent with (θ, κ) works in both economies simultaneously, i.e. $y_2^f(\theta, \kappa) = \underline{y}_2^f(\theta) > 0$ and $y_2^s(\theta, \kappa) = \underline{y}_2^s(\theta) > 0$, the first-order condition shows the following relationship.

$$\left[1 - T'\left(\underline{y}_2^f(\theta) + \underline{y}_2^s(\theta)\right)\right] w^f(\theta) = v'\left(\frac{\underline{y}_2^f(\theta)}{w^f(\theta)} + \frac{\underline{y}_2^s(\theta)}{w^s(\theta)}\right) = \left[1 - T'\left(\underline{y}_2^f(\theta) + \underline{y}_2^s(\theta)\right)\right] w^s(\theta) \tag{10}$$

Since the agent faces the same marginal tax rate, equation (10) can be simplified to $w^f(\theta) = w^s(\theta)$. In other words, the agent will work in both economies only if the wage rates in both economies are equivalent. For simplicity, I assume that agents always work in the formal economy when they face the same wages in both economies. Then, if the informal economy is formalized and both formal and informal incomes are taxable, agents will always work in one of the two economies and the case for equation (10) can be ignorable.

Assumption 2. *When the informal economy is formalized and an agent with (θ, κ) faces the same wages in the formal and informal economy, she always chooses to work in the formal economy.*

Following the optimality conditions (8) to (10) and Assumption 2, the income choices of low-cost workers can also be simplified as in Lemma 1.

Lemma 1. *If the informal economy is formalized, an agent will work either in the formal economy or in the informal economy, where she will earn the higher wage. Moreover, if an agent faces the same wage rate in both economies, she will always work in the formal economy.*

According to Lemma 1, an agent with productivity θ chooses to work in the formal economy when $w^f(\theta) \geq w^s(\theta)$ and in the informal economy when $w^s(\theta) > w^f(\theta)$. Since both economies are treated as formal and there is no difference between them in terms of income taxes, the income choice in this case can be considered as a simple sectoral choice between the two sectors, as in Rothschild and Scheuer (2013). Thus, there is a threshold $\tilde{\theta}$ defined by $\tilde{\theta} = \min_{\theta}\{\theta \mid w^f(\theta) \geq w^s(\theta)\}$ and the income choice of an agent with (θ, κ) can be summarized as follows.

$$\left(y_2^f(\theta, \kappa), y_2^s(\theta, \kappa)\right) = \begin{cases} \left(\overline{y}_2^f(\theta), 0\right) & \text{if} \quad \theta \geq \tilde{\theta} \\ (0, \overline{y}_2^s(\theta)) & \text{otherwise} \end{cases} \tag{11}$$

Furthermore, since agents face the uniform income tax schedule and have the identical utility function, high-cost workers follow the same formal income schedule when the informal economy is formalized and when it is not. Lemma 2 describes that there is no effect of formalizing the informal economy on high-cost workers. Thus, a certain amount of agents with sufficiently high participation costs follow the same optimality condition regardless of the formalization status.

Lemma 2. *If an agent is a high-cost worker with productivity $\theta \geq \tilde{\theta}$, the formal incomes are identical regardless of formalization of the informal economy, i.e. $\overline{y}_1^f(\theta) = \overline{y}_2^f(\theta) = \overline{y}^f(\theta)$.*

3. Optimal tax schedule

In this paper, the government formalizes a fraction $\pi \in [0, 1]$ of the informal economy, but does not know which agent is affected by the formalization. Thus, for the government, the social welfare function becomes a weighted sum of $V_1(\theta, \kappa)$ and $V_2(\theta, \kappa)$ defined in equations (2) and (7), respectively. On the other hand, since an agent with productivity θ earns the fixed wages $w^f(\theta)$ and $w^s(\theta)$ in each economy and knows whether she is subject to formalization, she chooses different sizes of labor supply to the formal and informal economies if her informal job is formalized and if it is not. Therefore, the government wants to propose an income tax schedule based on observable incomes and tries to find an optimal level of income tax or transfer for each agent with productivity θ.

3.1 Tax perturbation approach

I consider a social planner problem with the following general social welfare function and the government budget constraint.

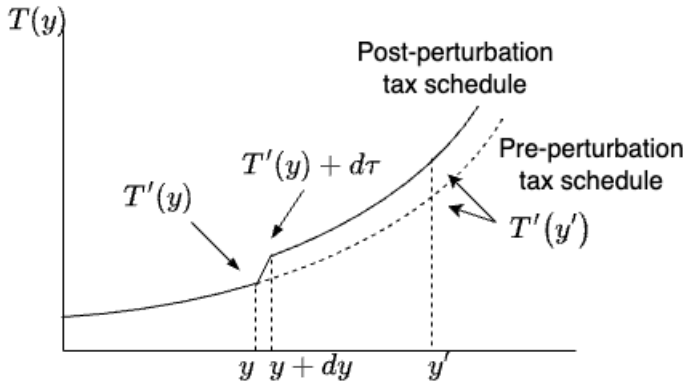
$$\int_{\underline{\theta}}^{\overline{\theta}} \int_0^{\infty} \lambda(\theta, \kappa) W(\theta, \kappa) dG_\theta(\kappa) dF(\theta) \tag{12}$$

where $W(\theta, \kappa) = (1 - \pi)V_1(\theta, \kappa) + \pi V_2(\theta, \kappa)$.

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_0^{\infty} \left[(1 - \pi) T \left(y_1^f(\theta, \kappa) \right) + \pi T \left(y_2^f(\theta, \kappa) + y_2^s(\theta, \kappa) \right) \right] dG_\theta(\kappa) dF(\theta) \geq E \tag{13}$$

The social welfare function (12) consists of two different indirect utilities, $V_1(\theta, \kappa)$ and $V_2(\theta, \kappa)$, and $\lambda(\theta, \kappa)$ is a Pareto weight. The expectation of $\lambda(\theta, \kappa)$, or $\mathbb{E}[\lambda(\theta, \kappa)]$, is nor-

Figure 2: Image of the tax perturbation approach



Note: The solid line curve is the status quo tax schedule before a perturbation, and the dashed part is the tax schedule after the perturbation. If the marginal tax rate is perturbed by $d\tau$ over an arbitrary income range $[y, y + dy]$, the agents whose taxable income corresponds to this income range will face slightly different marginal tax rates by $d\tau$, while others will not. I assume that $d\tau$ and dy are sufficiently small and that the perturbation occurs over the limited number of agents.

malized to 1, and this implies that the Pareto weights coincide with the marginal social welfare weights[10]. The government budget constraint is expressed by condition (13) and E stands for exogenous government expenditure. The tax schedule is optimal if it maximizes the social welfare (12), satisfying the income optimality conditions (3) to (5) and (8) to (10), and the government budget constraint (13).

Using the tax perturbation approach, I derive an optimal tax formula with a perturbation in the marginal tax rate[11]. In this approach, as in Figure 2, I consider an increase in the marginal tax rate $d\tau$ over an arbitrary income range $[y, y + dy]$. Both $d\tau$ and dy are assumed to be sufficiently small. If the marginal tax rate is increasing in its argument, as in Figure 2, an agent with taxable income y faces a higher marginal tax rate by $d\tau$, i.e. she has less incentive to supply the same amount of labor and reduces her labor supply at her current job. This is an intensive margin response to the tax perturbation. In addition, she may have the option of shifting some of her labor supply to the other economy, which is an extensive margin response. Then the tax perturbation simply changes social welfare through changes in net income and tax revenues. This is the mechanical and welfare effect. To balance these positive and negative effects of the perturbation, the optimal tax formula is determined to

[10]For more details, see Doligalski and Rojas (2023).

[11]The tax perturbation approach originated with Saez (2001) and was further refined by Golosov et al. (2014) and Jacquet and Lehmann (2021).

equalize the total impact of the perturbation to 0.

This paper does not assume a guaranteed minimum income and does not consider an extensive margin response for labor market participation. I focus on how the formalization of the informal economy affects labor supply decisions in the formal and informal economies, not the decision to be employed or unemployed. The incentive compatibility constraint guarantees that an appropriate level of labor supply is better than being completely unemployed, and the results of the quantitative analyses support this assumption.

According to Corollary 1 from Milgrom and Segal (2002), the indirect utility functions, $V_1(\theta, \kappa)$ and $V_2(\theta, \kappa)$, are differentiable by θ almost everywhere. Hence, the slope of the weighted sum of the indirect utility functions $W(\theta, \kappa)$ is

$$\begin{aligned}\frac{dW(\theta,\kappa)}{d\theta} &= (1-\pi)V_{1,\theta}(\theta,\kappa) + \pi V_{2,\theta}(\theta,\kappa) \\ &= -(1-\pi)\left(\rho^f(\theta)\frac{y_1^f(\theta,\kappa)}{w^f(\theta)} + \rho^s(\theta)\frac{y_1^s(\theta,\kappa)}{w^s(\theta)}\right)v'(n_1(\theta,\kappa)) \\ &\quad - \pi\left(\rho^f(\theta)\frac{y_2^f(\theta,\kappa)}{w^f(\theta)} + \rho^s(\theta)\frac{y_2^s(\theta,\kappa)}{w^s(\theta)}\right)v'(n_2(\theta,\kappa)) \equiv W_\theta(\theta,\kappa) \qquad (14)\end{aligned}$$

where $\rho^i(\theta) = \dfrac{w_\theta^i(\theta)}{w^i(\theta)}$, the growth rate of productivity in economy $i = \{f, s\}$. Then, $W(\theta, \kappa)$ can be expressed as it follows.

$$W(\theta,\kappa) = W(0,\kappa) + \int_0^\theta W_\theta(\theta',\kappa)d\theta' \qquad (15)$$

Equation (15) implies that an allocation perturbation of dy shifts utility levels along the weighted sum of the slopes of the indirect utility functions, $(1-\pi)V_{1,\theta}(\theta,\kappa) + \pi V_{2,\theta}(\theta,\kappa)$. Thus, an optimal tax formula determines an optimal tax schedule and offsets positive and negative effects on the slopes of the indirect utility functions by equalizing the impacts to 0.

3.2 Intensive margin responses

Agents with productivity above the threshold. My starting point for thinking about the intensive margin responses is the case where high-cost workers with productivity $\theta \geq \tilde{\theta}$ are subject to the allocation perturbation. In this case, they work only in the formal economy, regardless of their formalization status, and their formal income is $\overline{y}^f(\theta)$. If the tax perturbation of $d\tau$ occurs over an arbitrary income range $[y^f, y^f + dy]$ where $y^f = \overline{y}^f(\theta)$,

the total fiscal impact of the perturbation due to changes in the labor supply of the high-cost workers is defined as follows.

$$T'(y^f)\left(1-G_\theta\left(\tilde{\kappa}(\theta)\right)\right)f(\theta)\frac{dy^f}{dT'}dyd\tau \tag{16}$$

where $\left(1-G_\theta\left(\tilde{\kappa}(\theta)\right)\right)f(\theta)$ expresses the total number of high-cost workers affected by the perturbation, and $\frac{dy^f}{dT'}$ is an change in taxable income due to the perturbation. Thus, the fiscal impact (16) means the total marginal change in tax payment due to the perturbation when an agent's income choice is irrelevant to formalization.

Then, since the slopes of $V_1(\theta,\kappa)$ and $V_2(\theta,\kappa)$ have the same derivative, the impact of the perturbation on the slope of $W(\theta,\kappa)$ is summarized as

$$\begin{aligned}\frac{dW_\theta(\theta,\kappa)}{dT'} &= \frac{dW_\theta(\theta,\kappa)}{dy^f}\frac{dy^f}{dT'} \\ &= -\left\{(1-\pi)\frac{dV_{1,\theta}(\theta,\kappa)}{dy^f}+\pi\frac{dV_{2,\theta}(\theta,\kappa)}{dy^f}\right\}\frac{dy^f}{dT'} \\ &= \left(1-T'(y^f)\right)\rho^f(\theta)\left(1+\frac{1}{\varepsilon}\right)\frac{dy^f}{dT'}\end{aligned} \tag{17}$$

where $\varepsilon=\frac{v'(n)}{nv''(n)}$, the elasticity of labor supply. $\left(1+\frac{1}{\varepsilon}\right)$ can be interpreted as a scaling factor that adjusts for the effect of net wage changes on labor supply. By multiplying wage growth $\rho^f(\theta)$ by this factor, we emphasize how sensitive labor supply is to changes in net wages. In this case, I consider only the change in formal income due to the perturbation, not that in informal income. This is because high-cost workers with productivity $\theta\geq\tilde{\theta}$ work only in the formal economy, regardless of formalization status.

Combining equations (16) and (17), the intensive margin responses of high-cost workers with productivity $\theta\geq\tilde{\theta}$ are defined as the total fiscal impact of the perturbation, normalized by a marginal change in welfare, and have the following impact on the tax revenue.

$$IM^h=\frac{T'(y^f)}{1-T'(y^f-)}\frac{\left(1-G_\theta\left(\tilde{\kappa}(\theta)\right)\right)f(\theta)}{\rho^f(\theta)\left(1+\frac{1}{\varepsilon}\right)}dW_\theta(\theta,\kappa)dy \quad \text{for } \theta\geq\tilde{\theta} \tag{18}$$

As long as $d\tau$ is sufficiently small, dT' can be canceled out by $d\tau$ in the equation. For high-cost workers with productivity $\theta\geq\tilde{\theta}$, there is no impact of formalization on their intensive margin responses.

Low-cost workers respond to the tax perturbation in the same way. If their informal jobs are formalized and treated as taxable, they work in the formal economy regardless of formalization status, and there are no changes in the intensive margin responses from when they were high-cost workers. However, if the informal economy is not formalized, low-cost workers have the option of working in the informal economy and earning untaxed income from it. In other words, the intensive margin responses in this case reflect the distortion of formal income due to the perturbation.

Suppose that the mapping $s(\theta) = \min\left\{\theta' \in [\underline{\theta}, \overline{\theta}] \text{ s.t. } \underline{y}_1^f(\theta') \geq \overline{y}^f(\theta)\right\}$ indicates whether low-cost workers with productivity θ are distorted by the perturbation when their informal jobs are not formalized, i.e., agents are distorted if $s(\theta) \neq \theta$ but not distorted if $s(\theta) = \theta$. Thus, when the informal income is not taxable, the perturbation of $d\tau$ at the income level $y^f = \overline{y}^f(\theta)$ corresponds to the low-cost workers with productivity $s(\theta)$ such that $y^f = \underline{y}_1^f(s(\theta))$. Since the perturbation affects the low-cost workers, who face the same marginal tax rate due to the definition of $s(\theta)$, i.e, $T'(y^f) = T'\left(\underline{y}_1^f(s(\theta))\right)$, the overall fiscal impact of the perturbation is defined as follows.

$$T'(y^f)\Big[(1-\pi)G_{s(\theta)}\left(\tilde{\kappa}(s(\theta))\right)f(s(\theta)) + \pi G_\theta\left(\tilde{\kappa}(\theta)\right)f(\theta)\Big]\frac{dy^f}{dT'}dyd\tau \tag{19}$$

where $\Big[(1-\pi)G_{s(\theta)}\left(\tilde{\kappa}(s(\theta))\right)f(s(\theta)) + \pi G_\theta\left(\tilde{\kappa}(\theta)\right)f(\theta)\Big]$ expresses the total number of agents corresponding to the perturbation.

If the informal jobs are not subject to formalization, low-cost workers may adjust their informal incomes as well as their formal incomes to the perturbation. Thus, there is an indirect effect of the perturbation on informal incomes. If the the perturbation affects low-cost workers with productivity $s(\theta)$ and the informal economy is still untaxable, the impact of the perturbation on the slope of $W(\theta, \kappa)$ can be expressed as follows.

$$\begin{aligned}
\frac{dW_\theta(\theta,\kappa)}{dT'} &= \frac{dW_\theta(\theta,\kappa)}{dy^f}\frac{dy^f}{dT'} \\
&= \left\{(1-\pi)\left[\frac{dV_{2,\theta}(s(\theta),\kappa)}{dy^f} + \frac{dV_{2,\theta}(s(\theta),\kappa)}{dy^s}\frac{dy^s}{dy^f}\right] + \pi\frac{dV_{1,\theta}(\theta,\kappa)}{dy^f}\right\}\frac{dy^f}{dT'} \\
&= \left(1 - T'(y^f)\right)\left[(1-\pi)\Delta\rho(s(\theta)) + \pi\rho^f(\theta)\left(1+\frac{1}{\varepsilon}\right)\right]\frac{dy^f}{dT'}
\end{aligned} \tag{20}$$

where $\Delta\rho(s(\theta)) = \rho^f(s(\theta)) - \rho^s(s(\theta))$ is the difference of the wage growth rates in the formal economy and the informal economy. When the informal jobs are not subject to formalization

and agents are allowed to change where they work, the slope of $W(\theta, \kappa)$ depends only on the difference in wage growth, not the responsiveness of labor supply.

Combining equations (19) and (20), the intensive margin responses of low-cost workers are defined as the total fiscal impact of the perturbation, normalized by a marginal change in welfare.

$$IM^{\ell} = \frac{T'(y^f)}{1 - T'(y^f)} \frac{(1-\pi) G_{s(\theta)} \left(\tilde{\kappa}(s(\theta))\right) f(s(\theta)) + \pi G_{\theta} \left(\tilde{\kappa}(\theta)\right) f(\theta)}{(1-\pi)\Delta\rho(s(\theta)) + \pi\rho^f(\theta)\left(1 + \frac{1}{\varepsilon}\right)} dW_{\theta}(\theta, \kappa) dy \quad \text{for } \theta \geq \tilde{\theta} \tag{21}$$

Equation (21) is similar to equation (18). However, comparing $\Delta\rho(s(\theta))$ and $\rho^f(\theta)\left(1 + \frac{1}{\varepsilon}\right)$, the intensive margin responses tend to be smaller when the informal economy is partially formalized[12] in the equation (21) unless there is a large difference in the number of agents affected by the perturbation.

Agents with productivity below the threshold. If the informal economy is not formalized, agents with productivity $\theta < \tilde{\theta}$ have the option of working in the informal economy. Their behavior depends on their cost type, not their productivity, and the intensive margin responses remain the same as before.

However, when the informal economy is formalized, the behavior of agents depends on productivity, and those with productivity $\theta < \tilde{\theta}$ work in the formalized informal economy regardless of their cost type. Since I assume that the tax perturbation of $d\tau$ at the income level $y^f = \overline{y}^f(\theta)$ occurs to the high-cost workers with productivity θ as the reference, the perturbation can be considered to affect agents with productivity $r(\theta)$ satisfying $r(\theta) = \min\left\{\theta' \in [\underline{\theta}, \tilde{\theta}] \text{ s.t. } \overline{y}_2^s(\theta') = \overline{y}^f(\theta)\right\}$[13] when their informal jobs are subject to formalization. Therefore, the agents with income $\overline{y}_2^s(r(\theta))$ are affected by the tax perturbation of $d\tau$ because of the definition of $r(\theta)$, and the agents face the same marginal tax rate, i.e. $T'(y^f) = T'\left(\overline{y}_2^s(r(\theta))\right)$.

Intensive margin responses. The intensive margin responses can be summarized to one equation as follows. Equation (22) combines the case with productivity $\theta \geq \tilde{\theta}$ and that

[12] This analysis is based on the parameter values used in the quantitative part: $\rho^f(\theta) = 4.29$, $\rho^s(\theta) = 2.68$, $\Delta\rho(\theta) = 1.61$, and $\varepsilon = 0.33$. However, the result is not limited, this combination of parameters and a wide range of parameter values are applicable.

[13] $w^s(\theta) > w^f(\theta)$ for $\theta < \tilde{\theta}$ because of strictly increasing wage rates in θ and the single crossing condition. Hence, $r(\theta)$ is always smaller than a given θ corresponding to $\overline{y}^f(\theta)$. This assumption holds in the data for Colombia, which is used for the parameters in my quantitative analyses.

with $\theta \leq \tilde{\theta}$, and high-cost and low-cost workers behave diferently in each case.

$$IM = \frac{T'(y^f)}{1-T'(y^f)} \left[\frac{(1-\pi)\left[1-G_\theta\left(\tilde{\kappa}(\theta)\right)\right] f(\theta) + \pi\left[1-G_{t(\theta)}\left(\tilde{\kappa}(t(\theta))\right)\right] f(t(\theta))}{\left[(1-\pi)\rho^f(\theta) + \pi\varrho(t(\theta))\right]\left(1+\frac{1}{\varepsilon}\right)} \right.$$
$$\left. + \frac{(1-\pi)G_{s(\theta)}\left(\tilde{\kappa}(s(\theta))\right) f(s(\theta)) + \pi G_{t(\theta)}\left(\tilde{\kappa}(t(\theta))\right) f(t(\theta))}{(1-\pi)\Delta\rho(s(\theta)) + \pi\varrho(t(\theta))\left(1+\frac{1}{\varepsilon}\right)} \right] dW_\theta(\theta,\kappa)dy \quad (22)$$

where

$$t(\theta) = \begin{cases} \theta & \text{if } \theta \geq \tilde{\theta} \\ r(\theta) & \text{otherwise} \end{cases} \qquad \text{and} \qquad \varrho(t(\theta)) = \begin{cases} \rho^f(\theta) & \text{if } \theta \geq \tilde{\theta} \\ \rho^s(r(\theta)) & \text{otherwise} \end{cases}$$

3.3 Mechanical and welfare effects and extensive margin responses

The tax perturbation has mechanical effects on tax revenue and net income, because if the perturbation marginally increases the income tax, an increase in tax revenue has a positive effect on social welfare, but a decrease in net income hurts it. In the case of a marginal decrease in income tax, there are opposite effects on social welfare and net income. Thus, when the informal economy is formalized, the mechanical and welfare effects simply consider the total change in social welfare due to the perturbation.

When the informal economy is not formalized, the mechanical and welfare effects of the perturbation are more complicated than when the informal economy is formalized, because there are extensive margin responses by agents who shift their labor supply to the other economy. In other words, if a low-cost worker only has income from the formal economy, she may have an incentive to start working in the informal economy for untaxed income. The perturbation changes the threshold $\kappa(\theta)$ in the range $[\theta, s(\theta)]$, where a high-cost worker with productivity $\theta' \in [\theta, s(\theta)]$ is affected by the perturbation, but a low-cost worker with the same productivity is not. Let $\Delta T(\theta) = T(y^f) - T\left(\underline{y}_2^f(\theta)\right)$ be the tax loss to an agent with productivity θ who shifts some of her labor supply to the informal economy.

Hence, the total mechanical and welfare effects, including the extensive margin responses

of low-cost workers under no formalization, are

$$ME = \Bigg\{ \int_{s(\theta)}^{\overline{\theta}} \int_0^{\infty} \mu(\theta', \kappa) dG_{\theta'}(\kappa) dF(\theta') + \pi \int_{\theta}^{s(\theta)} \int_0^{\infty} \mu(\theta', \kappa) dG_{\theta'}(\kappa) dF(\theta') \\ + (1-\pi) \int_{\theta}^{s(\theta)} \left[\int_{\tilde{\kappa}(\theta')}^{\infty} \mu(\theta', \kappa) dG_{\theta'}(\kappa) - g_{\theta'}(\tilde{\kappa}(\theta')) \Delta T(\theta') \right] dF(\theta') \Bigg\} dW_{\theta}(\theta, \kappa) dy. \quad (23)$$

where $\mu(\theta', \kappa) = 1 - \lambda(\theta', \kappa)$. The first term shows the mechanical and welfare effects of the perturbation on agents regardless of the formalization status. The second term multiplied by π expresses the effects of the perturbation when the informal economy is formalized and there are no extensive margin responses of low-cost workers. Thus, the higher the level of formalization, the greater the positive impact of the perturbation on social welfare. The second line of equation (23) includes a negative welfare impact due to the extensive margin responses of low-cost workers. The level of formalization itself has no direct effect on agents' behavior, but it changes the welfare effects of the perturbation and indirectly affects income choices through changes in marginal tax rates defined by the optimal tax formula. When a smaller fraction π of the informal economy is formalized, a larger $1 - \pi$ is multiplied by the extensive margin responses, which leads to smaller mechanical and welfare effects. This effect is limited because extensive margin responses are only considered for those at the participation cost threshold and at a productivity level where high-cost workers are affected by the perturbation but low-cost workers are not.

3.4 Optimal tax formula

Equating the intensive margin responses with the mechanical and welfare effects including the extensive margin responses, i.e. $IM = ME$, means that a small allocation perturbation has no fiscal impact along the slope of $W(\theta, \kappa)$, i.e. no fiscal impact along the weighted sum of the indirect utilities $V_1(\theta, \kappa)$ and $V_2(\theta, \kappa)$. Theorem 1 expresses this condition based on the well-known ABC formula.

Theorem 1. *The optimal tax rate at income y corresponding to $\overline{y}^f(\theta)$ satisfies*

$$\frac{T'(y)}{1 - T'(y)} = \Big[A^h(\theta) B^h(\theta) + A^{\ell}(\theta) B^{\ell}(\theta) \Big] C(\theta) \quad (24)$$

where

$$A^h(\theta) = \left[(1-\pi)\rho^f(\theta) + \pi\varrho(t(\theta))\right]\left(1+\frac{1}{\varepsilon}\right) \tag{25}$$

$$A^\ell(\theta) = \left[(1-\pi)\Delta\rho(s(\theta)) + \pi\varrho(t(\theta))\right]\left(1+\frac{1}{\varepsilon}\right) \tag{26}$$

$$B^h(\theta) = \frac{1-F(\theta)}{(1-\pi)\left[1-G_\theta\left(\tilde{\kappa}(\theta)\right)\right]f(\theta) + \pi\left[1-G_{t(\theta)}\left(\tilde{\kappa}(t(\theta))\right)\right]f(t(\theta))} \tag{27}$$

$$B^\ell(\theta) = \frac{1-F(\theta)}{(1-\pi)G_{s(\theta)}\left(\tilde{\kappa}(s(\theta))\right)f(s(\theta)) + \pi G_{t(\theta)}\left(\tilde{\kappa}(t(\theta))\right)f(t(\theta))} \tag{28}$$

$$C(\theta) = \frac{ME}{1-F(\theta)} \tag{29}$$

The first two terms, $A^h(\theta)$ and $A^\ell(\theta)$, capture the standard elasticity and efficiency arguments for high-cost and low-cost types, respectively. Compared to the typical ABC formula in the literature, these terms are modified by the wage growth rates due to the choice between the formal and informal economy. A lower elasticity of labor supply, which could be amplified by wage growth rates, allows the government to set higher marginal tax rates. The second two termms, $B^h(\theta)$ and $B^\ell(\theta)$, measure the thickness of the right tail of the skill distribution for each cost type. A thicker tail is associated with higher tax rates. The last term, $C(\theta)$, measures the desire for redistribution, including extensive margin responses.

More specifically, for productivity $\theta < \tilde{\theta}$, higher π leads to lower $A^h(\theta)$ but has little effect on $A^\ell(\theta)$. Thus, when the informal economy is partially formalized, high-cost workers with productivity $\theta < \tilde{\theta}$ have a downward pressure on the marginal tax rate through $A^h(\theta)$. However, the effects of the formalization on $B^h(\theta)$ and $B^\ell(\theta)$ depend on the distribution and are unclear, and $C(\theta)$ always increases in π. Thus, the downward pressure of $A^h(\theta)$ and the upward pressure of $C(\theta)$ are offset. Thus, if the formalization changes the marginal tax rates, the thickness of the right tail of the distribution is a key driver in this case.

Similarly, for productivity $\theta \geq \tilde{\theta}$, the formalization has little effect on $A^h(\theta)$ but increases $A^\ell(\theta)$. Since $C(\theta)$ also increases in π, formalization is likely to put upward pressure on the marginal tax rate. The changes in $B^h(\theta)$ and $B^\ell(\theta)$ due to formalization are unclear, but formalization may increase marginal tax rates if $B^h(\theta)$ and $B^\ell(\theta)$ do not decrease much.

4. Quantitative analysis

To examine the impact of formalization on the tax schedule and tax revenue, this paper uses the parameters of Doligalski and Rojas (2023)[14], which uses the data from the Colombian Household Survey in 2013 by the National Administration of Statistics of Colombia (DANE). The data is restricted to those aged 24 to 50 without children, and the total number of observations is 34,000. In the quantitative analysis, formal workers are defined as those who report that they are affiliated to all three components of the social security system: the pension system, the health insurance system, and the occupational accident insurance system[15]. A summary of the parameters can be found in the Appendix.

Based on the observed hourly wages in the data, the wage functions for the formal and informal economies are defined as the following equations, and the parameters are estimated by maximum likelihood.

$$\log\left(w^f(\theta)\right) = \log\left(w^f(0)\right) + \rho^f\theta + u \tag{30}$$

$$\log\left(w^s(\theta)\right) = \log\left(w^s(0)\right) + \rho^s\theta + u. \tag{31}$$

where $\theta = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $u \sim N(0, \sigma_u^2)$. X consists of individual characteristics, job characteristics, and worker-firm relationship. Then the distribution of productivity θ is obtained by kernel density estimation as shown in Figure D.2(a). In the simulation of optimal tax schedules in the following part, the distribution of θ is complemented with a Pareto tail for the top 1% of wages and the support of θ is normalized to [0,1].

Furthermore, κ follows a generalized Pareto distribution $G_\theta(\kappa) = 1 - \left(1 + \frac{\kappa}{\tilde{\sigma}}\right)^{-1}$, where $\tilde{\sigma} = \sigma_\kappa\left(w^f(\theta) - w_\kappa\right)^{\alpha_\kappa}$ is the productivity-dependent scale parameter. Since $w^f(\theta)$ increases with θ, $G_\theta(\kappa)$ decreases and shifts to the right as θ increases. This means that an agent with a higher productivity θ tends is more likely to have a higher participation cost κ in the informal economy and to be a high-cost worker. In addition to that, the disutility of labor supply $v(n)$ is defined as $v(n) = \Gamma\frac{n^{1+1/\varepsilon}}{1+1/\varepsilon}$, where $\varepsilon = 0.33$ is the Frisch elasticity of labor supply from Chetty (2012) and $\Gamma = 0.032$ is a weight on labor supply in the utility. As in Rothschild and Scheuer (2013), the Pareto weights follow $\lambda(\theta) = r(1 - F(\theta))^{r-1}$. The parameter $r \geq 1$ reflects the strength of preferences for redistribution and is equal to the Pareto weight given to the lowest productivity level. In the following part, I consider $r = 1.4$

[14]I appreciate that Paweł Doligalski shared with me the original code for Doligalski and Rojas (2003).

[15]For more details on the data, estimation methods, and the fitness of the parameters, see Doligalski and Rojas (2023).

as relatively strong preferences for redistribution, with a downward slope from the 0th to the 100th percentile. Figure D.2(b) in the appendix shows how the Pareto weights are allocated across productivity.

In this section, I use $\pi = 0.3$ as the level of formalization. That is, I characterize an optimal tax schedule when 30% of the informal economy is formalized, and then compare the effects of formalizing the informal economy on optimal tax schedules, tax revenues, income transfers, and social welfare. First, I show the main results when the informal economy is partially formalized and the tax schedule is also optimized along with formalization. Then, as counterfactual analyses, I show the results (i) when both marginal and average tax rates are fixed regardless of formalization, and (ii) when only marginal tax rates are fixed, i.e., average tax rates can be modified by changing the actual amount of tax paid. As shown in Figure D.1, successful countries have had a reduction in their rates of informal employment by about 20-30% in 5-10 years. Thus, 30% formalization is a possible policy target, and the most successful case shows a reduction of almost 50% through a combination of measures[16].

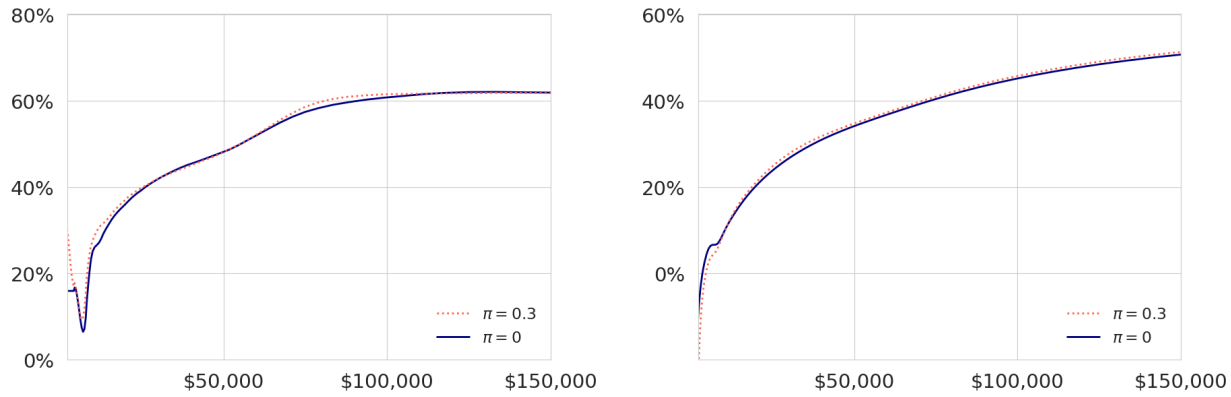
4.1 Optimal tax schedule

Figure 3(a) simply compares the optimal marginal tax rates when the informal economy is not formalized, $\pi = 0$, and when 30% of it is formalized, $\pi = 0.3$. If the government does not formalize the informal economy at all, the optimal marginal tax rates start at about 15% and fall sharply to less than 10% at the very low income level for phasing-out. The marginal tax rates then rise progressively to 60% and remain stable at the high income level. If 30% of the informal economy is formalized, the marginal tax rate at the lowest income level is about 30% and falls below 10% as in the case of no formalization. Then the marginal tax rates increase similarly and stable around 60% at the high income level.

By converting the marginal tax rates to the average tax rates for each income level, Figure 3(b) shows the tax burden more intuitively. Since the two lines start with negative tax rates at income level 0 and intersect with the horizontal line at some positive points, the tax schedules in the equilibrium require the government to provide income transfers to the lowest income group. Moreover, when 30% of the informal economy is formalized, lower average tax rates are imposed on the lowest income level, but higher tax rates are imposed on all income levels above about USD 20,000. Thus, 30% formalization increases the amount of income transfer for the lowest income level, $T(0)$, by 62%, and such a larger transfer requires

[16]In Uruguay, informal employment fell from 44.0% in 2007 to 23.6% in 2014, a reduction of about 50%.

Figure 3: Equilibrium tax rates with formalization



(a) Marginal tax rates **(b)** Average tax rates

Note: The straight blue line shows the optimal tax schedule without formalization, $\pi = 0$, and the dashed orange line shows that when 30% of the informal economy is formalized, $\pi = 0.3$. (a) shows the optimal marginal tax rates. For (b), the average tax rates dip into negative territory on the left, indicating that these people are receiving income transfers rather than paying income taxes. Income levels in Colombian pesos are adjusted to 2013 USD PPP.

larger tax burden by taxpayers. The 15% higher marginal tax rate for the lowest income people pushes them more into the informal economy if they are allowed to work informally, but even larger income transfers are optimal for the government to mitigate the negative effects of 30% formalization.

However, if the government keeps the tax schedule with the blue line in Figure 3(a), but formalizes 30% of the informal economy without adjusting the tax schedule, the maximum income transfer $T(0)$ increases by only about 17% and the average tax rates become slightly lower for almost all income levels as shown in Figure D.3. Therefore, implementing a formalization policy without considering the adjustment of tax schedules leads to a lower amount of income transfers to low income people and a slightly smaller tax burden for taxpayers. Since the formalized low-skilled workers face the marginal tax rate of 15%, not 30%, a smaller amount of income transfers is optimal for the government.

4.2 Tax revenues, income transfers and social welfare

Formalizing the informal economy is expected to broaden the tax base and increase the size of the government budget, but does not necessarily increase the total amount of tax revenue and income transfers. If the government formalizes 30% of the informal economy

and optimizes the tax schedule along with formalization, total tax revenue increases by 1.19%. Since Gaspar et al. (2019) estimates that developing countries will need additional government spending equal to about 4% of current GDP in 2030, i.e., increasing tax revenue by more than 20%, the 1.19% increase in tax revenue is a large contribution to tax revenue. Moreover, as the amount of income transfers in the model is defined as the difference between total tax revenue and the fixed amount of government spending for general purposes, the additional part of tax revenue fully contributes to the increase in income transfers, and 30% formalization raises the total amount of income transfers by 23.4%. As shown in Figure 3(b) and the discussion in the previous part, 30% formalization increases the income transfer for the lowest income level, $T(0)$, by 62%, and the government imposes higher average tax rates on middle and high income people to make such an increase in income transfers. Thus, formalization with optimal tax adjustment allows the government to improve tax revenues and income transfers without worsening social welfare. This can be interpreted as an implicit cost of 30% formalization in order to maintain the same level of social welfare. In other words, to mitigate the negative effects of 30% formalization on social welfare, particularly for low income people, 23.4% larger income transfers are needed overall, about 1.6 times the transfer for the lowest income level.

On the other hand, if the government formalizes the informal economy but fixes the marginal and average tax rates at the $\pi = 0$ level, 30% formalization increases total tax revenue by 0.90%, which is smaller than the increase when the tax schedule is optimized along with formalization. In addition, the total amount of income transfer is reduced by 29.7%, reflecting the fact that 30% of the informal economy is formalized and the corresponding number of agents are forced to earn taxable income instead of untaxed income from the informal economy. Formalizing the informal economy increases tax revenue as expected, but it worsens social welfare by 1.22%, which is not observed when the tax schedule is also optimized. In this case, the government formalizes the informal economy, but does not provide additional compensation to those who previously worked in the informal economy. Thus, this shows that when the government implements a formalization policy, it must provide sufficient redistributive incentives to agents who are forced to switch from the informal to the formal economy.

Furthermore, if the government fixes the marginal tax rates but adjusts the amount of tax paid, 30% formalization reduces total tax revenue by 0.88%. Since formalization is expected to be one of the effective measures to increase tax revenue and improve the low contribution of personal income tax to tax revenue, it is important to note that formalization can reduce

Table 1: Percentage change due to 30% formalization

(a) Formalization with the tax schedule adjustment

Tax revenue	**Income transfer**	**Social welfare**
1.19%	23.4%	0.02%

(b) Formalization without any adjustment in income tax

Tax revenue	**Income transfer**	**Social welfare**
0.90%	-29.7%	-1.22%

(c) Formalization with the fixed marginal tax rates, but average tax rates adjusted

Tax revenue	**Income transfer**	**Social welfare**
-0.88%	-17.3%	-0.03%

Note: Each ratio shows the percentage change when 30% of the informal economy is formalized. The first table (a) shows the results when the tax schedule is optimized along with formalization $\pi = 0.3$ and how much tax revenue and other variables increase with formalization. The second table (b) shows the results when the marginal and average tax rates are fixed at $\pi = 0$ and how much tax revenue and other variables change with formalization. The bottom table (c) shows the results when marginal tax rates are fixed but average tax rates are allowed to adjust to 30% formalization.

tax revenue in a certain situation. Moreover, the reduction in tax revenue directly leads to a reduction in the total amount of income transfers by 17.3%. While the government maintains the same marginal tax rates and agents follow the same optimality conditions for income choices, this does not necessarily mean that the actual tax burden remains unchanged for all agents. To keep social welfare at the same level, the government reduces the actual amount of income tax paid instead of formalizing part of the informal economy, which prevents a sharp decline in social welfare. As shown in Figure D.3 and the discussion in the previous part, 30% formalization slightly increases the income transfer for the lowest income level, $T(0)$, and lowers the average tax rates at almost all income levels. Thus, a small increase in the income transfer may not be sufficient to mitigate the effects of formalization, and formalizing the informal economy requires a comprehensive adjustment of the tax schedule, including changes in marginal tax rates.

5. Conclusion

This paper studies optimal income taxation when agents can choose between the formal and the informal economy as their place of work, and the informal economy is formalized. The main technical contribution is to allow agents to behave differently when the informal economy is formalized and when it is not, and to impose a common income tax on their income in both cases. Moreover, the optimal tax formula is expressed as the well-known ABC formula with a slight modification for labor market choices. Based on the quantitative results, the impact of formalizing the informal economy depends on whether the tax schedule is optimized along with formalization. When the tax schedule is also optimized, formalizing the informal economy has positive effects on tax revenues, income transfers, and social welfare. On the other hand, when the tax schedule is fixed, formalization worsens social welfare instead of increasing tax revenues, or reduces the target variables, contrary to what is expected. Thus, formalization requires a corresponding change in the tax schedule for more efficient redistribution.

The theoretical tools developed in this paper could be used in other settings. In my framework, an income tax depends only on formal income when the informal economy is not formalized, but is a joint tax of formal and informal income when the informal economy is formalized. Since the model considers different income compositions in a joint tax formula, it could be applied to any government intervention problem that changes the composition of taxable income, such as probabilistic tax audits, taxes for couples, changes in tax rules, and so on. Furthermore, this model assumes that agents know their states and behave quite differently when the informal economy is formalized than when it is not. However, if it is assumed that agents do not know their states and respond partially to formalization, this could be another extension in a different direction.

Appendix

A. Proof of Lemma 1

Proof. Suppose $w^f(\theta) > w^s(\theta)$. If the informal economy is formalized and an agent with productivity (θ, κ) chooses her formalized informal job, this means that the utility of working in the formalized informal economy is greater than or equal to that of working in the formal economy. If the agent chooses an optimal informal income according to the condition (9), her utility is maximal and denoted by $U\left(c_1(\theta,\kappa), \frac{\underline{y}_1^s(\theta)}{w^s(\theta)}\right)$. If the agent keeps the same income level $\overline{y} = \underline{y}_1^s(\theta)$ and switches to the formal economy, she keeps the same utility from consumption but has a lower disutility from labor supply due to $w^f(\theta) > w^s(\theta)$. Thus, the agent improves her utility in this case, $U\left(c_1(\theta,\kappa), \frac{\overline{y}}{w^f(\theta)},\right) > U\left(c_1(\theta,\kappa), \frac{\underline{y}_1^s(\theta)}{w^s(\theta)}\right)$. This is a contradiction. Therefore, if $w^f(\theta) > w^s(\theta)$, it is optimal for the low-cost worker to choose the formal economy. The case $w^f(\theta) < w^s(\theta)$ is also proved in the same way.

Suppose an agent with productivity θ follows the optimality conditions for her formal and informal incomes (8) and (9), and earns $\underline{y}_1^f(\theta)$ and $\underline{y}_1^s(\theta)$ when working in either the formal or formalized informal economy. Then, if she starts participating in both economies and maintains the same total income level by reducing her formal income by Δ instead of increasing her informal income by Δ, her formal and informal incomes are defined as $\hat{y}_1^f(\theta) = \underline{y}_1^f(\theta) - \Delta$ and $\hat{y}_1^s(\theta) = \Delta$. Compare the utilities,

$$U\left(c_1(\theta,\kappa), \frac{\underline{y}_1^f(\theta)}{w^f(\theta)}\right) > U\left(c_1(\theta,\kappa), \frac{\hat{y}_1^f(\theta)}{w^f(\theta)} + \frac{\hat{y}_1^s}{w^s(\theta)}\right) \tag{32}$$

$$= U\left(c_1(\theta,\kappa), \frac{\underline{y}_1^f(\theta) - \Delta}{w^f(\theta)} + \frac{\Delta}{w^s(\theta)}\right) \tag{33}$$

Given that $w^f(\theta) > w^s(\theta)$ and $U_n < 0$, the marginal disutility increases as the agent begins to participate in both economies. Since the agent has the same consumption as long as the total income level remains unchanged, it is optimal for her to work in one of the two economies where she earns the higher wage when the informal economy is formalized. □

B. Proof of Lemma 2

Proof. As long as an agent with productivity θ has sufficiently high participation costs $\kappa \geq \tilde{\kappa}(\theta)$, she works only in the formal economy, and by definition there is no incentive to work in the informal economy. By equations (3) and (8), the optimal formal income depends only on the marginal tax rate and the marginal disutility of labor supply, which are identical for both states. Thus, the agent with productivity θ chooses the same formal income level given the marginal tax rate. □

C. Detailed derivations of intensive margin responses

The fiscal effects of a tax perturbation can be obtained directly from the intuitions. However, the effects of the perturbation on the slope of $W(\theta, \kappa)$ in equation (17) require $\frac{dV_{1,\theta}(\theta,\kappa)}{dy^f}$ and $\frac{dV_{2,\theta}(\theta,\kappa)}{dy^f}$ to be considered separately. Since the high-cost workers with productivity $\theta \geq \tilde{\theta}$ have no informal income regardless of formalization status, I assume $y_1^f(\theta) = \overline{y}^f(\theta)$. By equation (14) and $\overline{y}^f(\theta) = w^f(\theta) n^f(\theta, \kappa)$,

$$\begin{aligned}
\frac{dV_{1,\theta}(\theta,\kappa)}{dy^f} &= \frac{d}{dy^f}\left[\rho^f(\theta)\frac{\overline{y}^f(\theta)}{w^f(\theta)} v'(n^f(\theta,\kappa))\right] \\
&= \frac{\rho^f(\theta)}{w^f(\theta)} v'(n^f(\theta,\kappa)) + \rho^f(\theta)\frac{\overline{y}^f(\theta)}{w^f(\theta)}\frac{v''(n^f(\theta,\kappa))}{w^f(\theta)} \\
&= \frac{\rho^f(\theta)}{w^f(\theta)} v'(n^f(\theta,\kappa))\left(1 + \frac{n^f(\theta,\kappa) v''(n^f(\theta,\kappa))}{v'(n^f(\theta,\kappa))}\right) \qquad (34)
\end{aligned}$$

Substituting $\varepsilon = \frac{v'(n)}{n v''(n)}$ and $\left(1 - T'\left(\overline{y}^f(\theta)\right)\right) = \frac{v'(n^f(\theta,\kappa))}{w^f(\theta)}$ into the last equation above,

$$\begin{aligned}
\frac{dV_{1,\theta}(\theta,\kappa)}{dy^f} &= \frac{\rho^f(\theta)}{w^f(\theta)} v'(n^f(\theta,\kappa))\left(1 + 1/\varepsilon\right) \\
&= -\rho^f(\theta)\left(1 - T'\left(\overline{y}^f(\theta)\right)\right)\left(1 + 1/\varepsilon\right). \qquad (35)
\end{aligned}$$

Then, following the same steps allow me to obtain

$$\frac{dV_{2,\theta}(\theta,\kappa)}{dy^f} = -\rho^f(\theta)\left(1 - T'\left(\overline{y}^f(\theta)\right)\right)\left(1 + 1/\varepsilon\right). \qquad (36)$$

Therefore, the slopes of $V_1(\theta,\kappa)$ and $V_2(\theta,\kappa)$ have the same derivative with respect to y^f, and the total effect of the perturbation on $W(\theta,\kappa)$ follows equation (17).

For the intensive margin responses of low-cost workers with productivity $\theta > \tilde{\theta}$, the first term of equation (20) is obtained by the same steps as for those of high-cost workers. However, the second term includes the interaction between formal and informal income and follows the different optimality condition for income choices. An indirect effect of the perturbation on informal income is $\dfrac{dy_1^s(\theta)}{dy^f} = -\dfrac{w^s(\theta)}{w^f(\theta)}$ because the total labor supply of a low-cost worker is fixed by $w^s(\theta)$ if the informal economy is not formalized according to equation (5) and $y_2^s(\theta)$ can be expressed as a decreasing function of formal income.

$$\begin{aligned} y_1^s(\theta) &= w^s(\theta) n_1^s(\theta,\kappa) \\ &= w^s(\theta)(n_1(\theta,\kappa) - n_1^f(\theta,\kappa)) \\ &= w^s(\theta)\left(n_1(\theta,\kappa) - \frac{\underline{y}_1^f(\theta)}{w^f(\theta)}\right) \end{aligned} \tag{37}$$

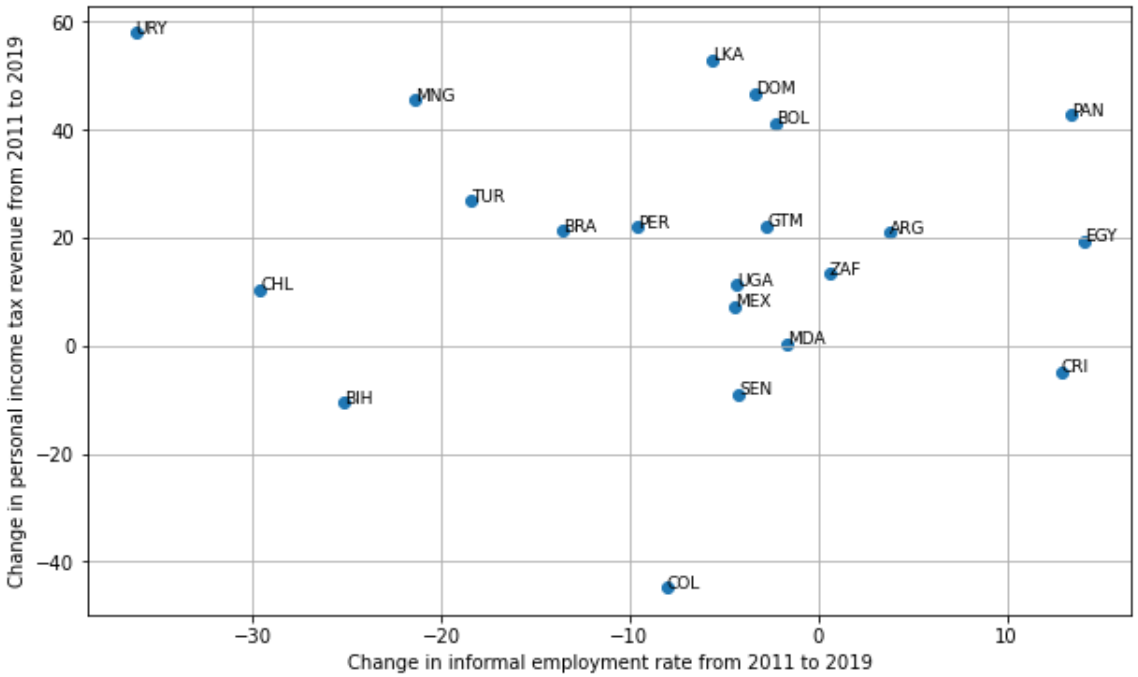
Then, the impact of the perturbation on the slope of the indirect utility function becomes as it follows.

$$\begin{aligned} \frac{dV_{1,\theta}(\theta,\kappa)}{dy^f} &= \frac{d}{dy^f}\left[\left(\rho^f(\theta)\frac{\underline{y}_1^f(\theta)}{w^f(\theta)} + \rho^s(\theta)\frac{y_1^s(\theta)}{w^s(\theta)}\right) v'(n(\theta,\kappa))\right] \\ &= \left(\frac{\rho^f(\theta)}{w^f(\theta)} + \frac{\rho^s(\theta)}{w^s(\theta)}\frac{dy_1^s(\theta)}{dy^f}\right) v'(n(\theta,\kappa)) \\ &\quad + \left(\rho^f(\theta)\frac{\underline{y}_2^f(\theta)}{w^f(\theta)} + \rho^s(\theta)\frac{y_1^s(\theta)}{w^s(\theta)}\right) v''(n(\theta,\kappa))\left(\frac{1}{w^f(\theta)} + \frac{1}{w^s(\theta)}\frac{dy_1^s(\theta)}{dy^f}\right) \\ &= (\rho^f(\theta) - \rho^s(\theta))\frac{v'(n(\theta,\kappa))}{w^f(\theta)} \\ &= (\rho^f(\theta) - \rho^s(\theta))\left(1 - T'\left(\underline{y}_1^f(\theta)\right)\right) \end{aligned} \tag{38}$$

In the last part, the optimality condition of income choices (5) is substituted to drop the marginal disutility from labor supply. Combining these results shows equation (20).

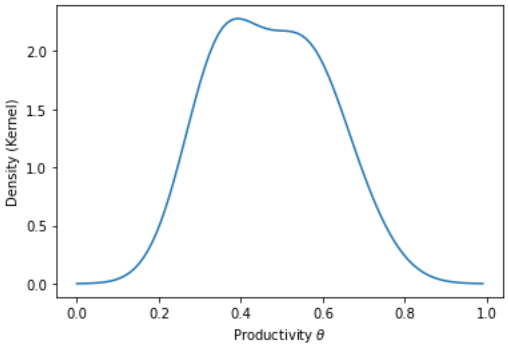
D. Additional figures and tables

Figure D.1: Change in informal employment rate compared to change in personal income tax revenue

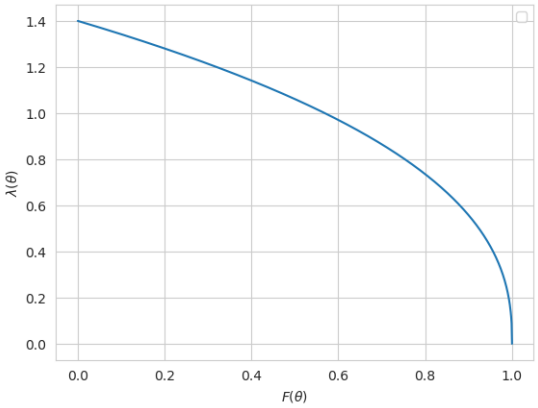


Note: The graph compares the change in the informal employment rate from 2011 to 2019 with the change in the personal income tax share of tax revenue over the same period. The informal employment rate is the share of all employment that is classified as informal out of the total employment. Negative values indicate that a country reduced the informal employment rate or the personal income tax share during this period. Colombia, Bosnia and Herzegovina, Senegal and Moldova experienced a decrease in the informal employment rate, but the personal income tax share did not increase or even decrease in these countries. Due to data availability, particularly for the informal employment rate, only 21 countries appear in the graph. Source: ILO for informal employment rate, IMF for tax revenue.

Figure D.2: Distribution of productivity level and Pareto weights



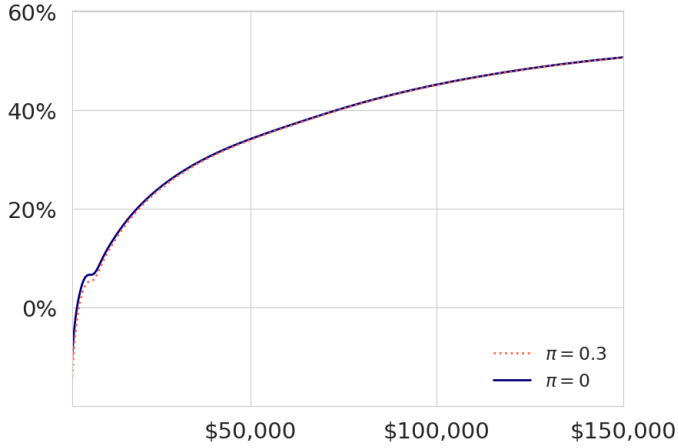
(a) Distribution of productivity level



(b) Pareto weights with weakly and strongly redistributive preferences

Note: (a): The distribution of productivity level θ is obtained by kernel density estimation. (b): Pareto weights follow $\lambda(\theta) = r(1 - F(\theta))^{r-1}$ with $r = 1.4$.

Figure D.3: Average tax rates with 30% formalization when the marginal tax rates are fixed, the actual amount of tax paid is adjusted



Note: The straight blue line shows the average tax without formalization, $\pi = 0$, and the dashed orange line shows that when 30% of the informal economy is formalized without adjusting the tax schedule along with formalization, $\pi = 0.3$. The graph dips into negative territory on the left, indicating that these people are receiving income transfers rather than paying income taxes. Income levels in Colombian pesos are adjusted to 2013 USD PPP.

Table D.1: Summary of parameters (from Doligalski and Rojas (2023))

Preference				
ε	Γ			
0.33	0.032			
(-)	(8e-4)			
Productivity				
$\log(w^f(0))$	$\log(w^s(0))$	ρ^f	ρ^s	α_w
0.0038	0.0068	4.29	2.68	2.25
(1e-4)	(1e-4)	(0.06)	(0.06)	(0.03)
Distribution of θ and κ				
σ_ε	σ_u	σ_κ	α_κ	w_κ
0.09	0.53	1.38	0.88	0.018
(2e-3)	(3e-3)	(0.03)	(0.01)	(2e-4)

Note: Standard errors are reported in brackets. Standard errors are obtained by Case Resampling Bootstrap using 150 draws. For the estimate of β, see Doligalski and Rojas (2023).

References

[1] **Allingham, Michael G., and Agnar Sandmo.** 1972. "Income tax evasion: A theoretical analysis." *Journal of Public Economics* 1 (3): 323–338.

[2] **Beaudry, Paul, Charles Blackorby, and Dezsö Szalay.** 2009. "Taxes and employment subsidies in optimal redistribution programs." *American Economic Review* 99 (1): 216–242.

[3] **Benedek, Dora, Juan Carlos Ben´itez, and Charles Vellutini.** 2022. "Progress of the personal income tax in emerging and developing countries." *IMF Working Paper* WP/22/20.

[4] **Chander, Parkash, and Louis L. Wilde.** 1998. "A general characterization of optimal income tax enforcement." *The Review of Economic Studies* 65 (1): 165–183.

[5] **Chetty, Raj.** 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica* 80 (3): 969–1018.

[6] **da Costa, Carlos E., and Felipe Lobel.** 2022. "The cost of informality: An optimal taxation approach." Available at SSRN: https://ssrn.com/abstract=4231413.

[7] **Diamond, Peter A.** 1998. "Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates." *American Economic Review* 88 (1): 83–95.

[8] **Doligalsk, Paweł, and Luis E. Rojas.** 2023. "Optimal redistribution with a shadow economy." *Theoretical Economics* 18 (2): 749–791.

[9] **Gaspar, Vitor, David Amaglobeli, Mercedes Garcia-Escribano, Delphine Prady, and Mauricio Soto.** 2019. "Fiscal policy and development: Human, social, and physical investments for the SDGs." *IMF Staff Discussion Note 19/03*.

[10] **Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin.** 2014. "A variational approach to the analysis of tax systems." *NBER Working Paper 20780*.

[11] **Hart, Keith.** 1973. "Informal income opportunities and urban employment in Ghana." *The Journal of Modern African Studies* 11 (1): 61–89.

[12] **ILO.** 1972. "Employment, incomes and equality. A strategy for increasing productive employment in Kenya."

[13] **ILO.** 2015b. "Recommendation No. 204 concerning the transition from the informal to the formal economy."

[14] **ILO.** 2017. "Way out of informality: Facilitating formalization of informal economy in South Asia."

[15] **ILO.** 2018. "Women and men in the informal economy: A statistical picture (third edition)."

[16] **ILO.** 2019. "Formalization of the informal economy, 2014–18: Synthesis review of ILO project evaluations." ILO, Evaluation Office.

[17] **Jacquet, Laurence, and Etienne Lehmann.** 2021. "Optimal income taxation with composition effects." *Journal of the European Economic Association* 19 (2): 1299–1341.

[18] **Keen, Michael.** 2012. "Taxation and development - again." *IMF Working Paper* WP/12/220.

[19] **Medina, Leandro, and Friedrich G. Schneider.** 2019. "Shedding light on the shadow economy: A global database and the interaction with the official one." *CESifo Working Paper No.7981.*

[20] **de Mel, Suresh, David McKenzie, and Christopher Woodruff.** 2013. "The semand for, and consequences of, formalization among informal firms in Sri Lanka." *American Economic Journal: Applied Economics* 5 (2): 122–50.

[21] **Milgrom, Paul, and Ilya Segal.** 2002. "Envelope theorems for arbitrary choice sets." *Econometrica* 70 (2): 583–601.

[22] **Mirrlees, James A.** 1971. "An exploration in the theory of optimum income taxation." *Review of Economic Studies* 38 (2): 175–208.

[23] **OECD/ILO.** 2019. "Tackling vulnerability in the informal economy." Development Centre Studies, OECD.

[24] **Piketty, Thomas, and Emmanuel Saez.** 2013. "Optimal labor income taxation." In *Handbook of Public Economics, Vol. 5*, edited by Auerbach, Alan J., Raj Chetty, Martin Feldstein, and Emmanuel Saez: 391–474, Elsevier.

[25] **Pratap, Sangeeta, and Erwan Quintin.** 2006. "Are labor markets segmented in developing countries? A semiparametric approach." *European Economic Review* 50 (7): 1817–1841.

[26] **Rocha, Rudi, Gabriel Ulyssea, and Laísa Rachter.** 2018. "Do lower taxes reduce informality? Evidence from Brazil." *Journal of Development Economics* 134 28–49.

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

[28] **Rothschild, Casey, and Florian Scheuer.** 2015. "A theory of income taxation under multi-dimensional skill Heterogeneity." *CESifo Working Paper No.5165.*

[29] **Rothschild, Casey, and Florian Scheuer.** 2016. "Optimal taxation with rent-seeking." *The Review of Economic Studies* 83 (3): 1225–1262.

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

[31] **Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro.** 2010. "Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007." *Policy Research Working Paper* 5356.

[32] **Schneider, Friedrich, and Robert Klinglmair.** 2004. "Shadow economies around the world: What do we know?" *IZA Discussion Paper* No. 1043.

[33] **Selin, Hakan, and Laurent Simula.** 2020. "Income shifting as income creation?" *Journal of Public Economics* 182 104081.

[34] **Slemrod, Joel, and Shlomo Yitzhaki.** 2002. "Tax avoidance, evasion, and administration." In *Handbook of Public Economics, Vol. 3*, edited by Auerbach, Alan J., and Martin Feldstein: 1423–1470, Elsevier.