Theoretical Economics 18 (2023), 749–791 1555-7561/20230749

Optimal redistribution with a shadow economy

Paweł Doligalski
School of Economics, University of Bristol

Luis E. Rojas
Department of Economics and Economic History, Universitat Autónoma de Barcelona and Barcelona School of Economics

We extend the theory of optimal redistributive taxation to economies with an informal sector. In particular, in our model, workers can supply labor simultaneously to the formal and the informal sectors, which we call moonlighting. The optimal tax formula contains two novel terms capturing informality responses on an intensive and an extensive margin. Both terms decrease the optimal tax rates. We estimate the model with Colombian data and find that informality strongly reduces tax rates at all income levels. The possibility to migrate to entirely informal employment restricts tax rates at low and medium income levels, while the possibility of moonlighting is relevant at higher earnings.

Keywords. Informal sector, moonlighting, income taxation, redistribution.

JEL classification. H21, H26, J46.

1. Introduction

Informal activity, broadly defined as any economic endeavor that evades taxation, accounts for a large fraction of economic activity in both developing and developed economies. The share of informal production in gross domestic product is consistently estimated to be on average above 10% in high income countries and above 30% in developing and transition countries, in extreme cases reaching 70% (Schneider and Enste (2000), Schneider, Buehn, and Montenegro (2011)). Globally, 2 billion workers are employed informally (ILO (2018)). The shadow economy allows workers to earn income that is unobserved by the government. Intuitively, this additional margin of response

Paweł Doligalski: pawel.doligalski@bristol.ac.uk
Luis E. Rojas: lrojasdu@gmail.com

Rojas acknowledges financial support from the European Research Council under Grant Horizon2020 GA 788547 (APMPAL-HET), and from the Spanish Agencia Estatal de Investigacion (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S). Doligalski thanks the Central Bank of Hungary for the opportunity to work on this project during his stay. We are grateful for valuable comments by the anonymous referees as well as Árpád Ábrahám, Charles Brendon, Antoine Camous, Hal Cole, Mike Golosov, Piero Gottardi, Ramon Marimon, Wojciech Kopczuk, Claus Kreiner, Dirk Krueger, Etienne Lehmann, Humberto Moreira, Erwin Ooghe, Wojciech Paczos, Evi Pappa, Dominik Sachs, Julia Schmieder, Jon Temple, and Yanos Zylberberg. All errors remain our own.

 https://doi.org/10.3982/TE4569

to taxation makes income redistribution more difficult. Indeed, empirical studies document often large informality responses to tax reforms.[1] However, a theory of how the income tax schedule should depend on informality is missing.

Our aim is to fill this gap. We derive the optimal nonlinear income tax schedule with a shadow sector and characterize how informality determines its shape. To verify the significance of our theoretical results, we estimate the model with Colombian data. Accounting for informality turns out to be quantitatively very important for the optimal policy.

Building on the seminal work of Mirrlees (1971), we consider a framework with heterogeneous agents equipped with distinct formal and shadow productivities. Workers face an idiosyncratic fixed cost of working in the shadow economy, which may reflect either ethical or technological constraints. The government observes only formal incomes and introduces a nonlinear tax to maximize its redistributive welfare criterion. Importantly, we allow workers to supply labor simultaneously to the formal sector and the shadow sectors, which we call *moonlighting*. In this way we can study incentives of formal employees to have an informal side job. Informal secondary employment is common and accounts for a substantial fraction of informal workers in many countries.[2] Furthermore, evidence suggests that starting a tax-advantaged secondary job is an important margin of response to tax reforms (Tazhitdinova (2017)).

Our main theoretical result is a sufficient statistics formula for the optimal tax schedule in the economy with an informal sector. The formula contains two novel terms due to informality responses on the extensive margin (getting an informal job) and the intensive margin (shifting hours between a formal and an informal job). The extensive margin responses are typically modeled as binary: working or not working. In our setting, it would correspond to agents being able to work only formally or only informally, and, as a result, would rule out moonlighting. Instead, we allow workers to moonlight, which means that they can complement formal earnings with additional income from an informal job. Intuitively, these responses can be important for workers with well paid formal jobs who face high marginal tax rates and for whom transitioning to entirely informal employment is too costly. The possibility of moonlighting also gives rise to informality responses on the intensive margin—shifting hours between the formal main job and the informal secondary job. We find that moonlighting workers respond on the intensive margin differently than formal workers. First, the formal earnings of moonlighting workers are more elastic. Second, moonlighting workers would never choose formal earnings where the tax schedule is locally regressive, i.e., where the marginal tax rates are decreasing. If the tax schedule features regions of regressivity then, following

[1]A positive impact of income tax rates on tax evasion and informality has been documented in Brazil (Monteiro and Assunção (2012), Rocha, Ulyssea, and Rachter (2018)), Russia (Gorodnichenko, Martinez-Vazquez, and Peter (2009)), Pakistan (Waseem (2018)), and Denmark (Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011)).

[2]Out of all workers engaged in informal work, the share with a formal main job was more than 10% in Barbados, more than 20% in the Russian Federation and Lithuania (Hussmanns and Jeu (2002)), and more than 50% in Poland (Statistics Poland (2019)). In Brazil, 37% of secondary jobs are micro-enterprises and can be classified as informal (Henley, Arabsheibani, and Carneiro (2009)).

a tax reform, moonlighting workers may respond on the intensive margin by jumping over a regressivity region to a discretely lower level of formal earnings.

In contrast, formal workers rarely respond by jumping.[3] Even though informality responses may involve abrupt earnings changes, we summarize their impact on tax revenue with well defined elasticities.

We analytically examine how informality affects the optimal tax rates in two ways. First, we fix the distribution of formal income and examine what happens if informality responses were ignored, e.g., because of the erroneous beliefs of policymakers. We find that ignoring informality responses would result in higher tax rates. In other words, correctly accounting for work incentives in the presence of the informal sector leads to lower optimal tax rates. Second, we fix model primitives, such as the distribution of productivities in the two sectors, and compare the optimal top tax rate in the model with and without the shadow economy. This comparison is more challenging since the income distribution is allowed to freely adjust to tax policy. We find that the optimal top tax rate is weakly lower in the model with a shadow economy. Once the top tax rate exceeds a certain tipping point, a large fraction of top earners joins the informal sector with a large loss of tax revenue. We show that it is never optimal to cross this tipping point, which implies an upper bound on the optimal top tax rate with an informal sector.

We estimate the model with Colombian data. Colombia is an attractive case study for two reasons. First, it has a large informal sector: close to 60% of main jobs are informal. Second, the level of informality in Colombia is very close to the average for the whole of Latin America.[4] We extract the information on formal and shadow incomes from the household survey and estimate the model by maximum likelihood. The model replicates well the empirical sorting of workers between the formal sector and the informal sector.

In our quantitative exercise, we compare the optimal tax schedule with the tax schedules chosen when various informality responses are ignored. Importantly, in this comparison, we allow for the endogenous adjustment of the income distribution. We find that the possibility of workers to migrate to entirely informal employment restricts tax rates at low and medium income levels, while the possibility of moonlighting is relevant at higher levels of income. Specifically, if all informality responses are ignored, the marginal tax rates are overshot at all income levels and, in particular, at the bottom, where they are too high by 70 percentage points or more. As a result, the shadow economy doubles in size relative to the optimum, which has catastrophic welfare consequences. If, instead, it is acknowledged that workers can move to the shadow economy and only the moonlighting responses are ignored, the tax rates at the bottom are approximately optimal, but the rates above the median formal income are too high—by up to 20 percentage points—when preferences for redistribution are strong. That is because incentives for moonlighting are important higher in the income distribution compared to incentives for switching from entirely formal to entirely informal employment. When

[3] Bergstrom and Dodds (2021) study jump responses in a Mirrlees model with multidimensional heterogeneity.

[4] Based on ILO (2018), the national share of informal employment in total employment in Latin America has a mean of 58.3% and a median of 59%.

preferences for redistribution are strong, ignoring moonlighting responses substantially increases the incidence of moonlighting among the most productive workers. Thus, it leads to a large welfare loss, equivalent to a 2.4% drop in consumption.

Related literature. Kopczuk (2001) considers income taxation with tax avoidance, which can be reinterpreted as informality, and shows that the standard formula for the optimal linear tax is still valid. In contrast, our results imply that the standard formula for the optimal *nonlinear* tax is no longer valid.[5] Piketty and Saez (2013) and Piketty, Saez, and Stantcheva (2014) study linear and top income taxation with a possibility of shifting income between two tax bases, one of which could stand for an informal sector, yet they do not consider extensive margin responses. Selin and Simula (2020) derive the optimal nonlinear tax schedules with income shifting, but they effectively rule out partial shifting, which would correspond to moonlighting in our framework. Beaudry, Blackorby, and Szalay (2009) study redistribution in the model with the informal sector when both formal income and formal hours worked are observed. We, instead, maintain the Mirrleesian assumption of unobserved hours worked.

Another approach to study tax evasion, originating with Allingham and Sandmo (1972), uses a framework with probabilistic audits and penalties, taking a tax rate as given. Andreoni, Erard, and Feinstein (1998) and Slemrod and Yitzhaki (2002) review this strand of literature. We take a complementary approach and study the optimal nonlinear tax schedule conditional on the fixed tax evasion abilities of workers. Although we do not model tax audits and penalties explicitly, they are a possible justification for different productivities in the formal and the shadow sectors. Thus, our results on optimal taxation should be understood as taking the quality of tax enforcement as given. Some early results from merging both optimal taxation and optimal tax compliance policies were derived by Cremer and Gahvari (1996), Kopczuk (2001), and Slemrod and Kopczuk (2002). Leal Ordóñez (2014) and Di Nola, Kocharkov, Scholl, and Tkhir (2020) investigate tax and enforcement policies quantitatively in the dynamic incomplete markets models.

This paper is closely related to the literature on optimal taxation with multiple sectors. Rothschild and Scheuer (2014) consider uniform taxation of multiple sectors when agents can work in many sectors simultaneously. Kleven, Kreiner, and Saez (2009), Scheuer (2014), and Gomes, Lozachmeur, and Pavan (2017) study differential taxation of broadly understood sectors (e.g., individual tax filers and couples, employees and entrepreneurs) when agents can belong to one sector only. Jacobs (2015) studies a complementary problem when all agents work in all sectors at the same time. Our analysis differs in that we consider a particular case of differential taxation—only one sector is taxed—when agents face an idiosyncratic fixed cost of participating in one of the sectors. This structure implies that some agents can effectively work in one sector only, while others are unconstrained in supplying labor to two sectors simultaneously.

Emran and Stiglitz (2005) and Boadway and Sato (2009) study commodity taxation in the presence of informality. Both papers assume that commodity taxes affects only the

[5] Our settings is not identical to Kopczuk's, since we consider a fixed cost of shadow employment. In a previous working paper version (Doligalski and Rojas (2016)), we show that the standard formula for the optimal nonlinear tax is not valid even if we abstract from the fixed cost of shadow employment.

formal sector.[6] Hence, provided that formal and shadow goods are perfect substitutes, a consumption tax is equivalent to a proportional tax on formal income. Under these assumptions, our focus on nonlinear income tax is without loss of generality. Boadway, Marchand, and Pestieau (1994) and Huang and Rios (2016) study the optimal tax mix in the opposite case, when the consumption tax cannot be evaded. A related literature on the optimal commodity taxation with home production (Kleven, Richter, and Sørensen (2000), Olovsson (2015)) studies the case of non-perfect substitutability between market and home produced goods.

Structure of the paper. In the following section, we introduce the framework and characterize the equilibrium for a given income tax. In Section 3, we derive the optimal tax formula and show that the informal sector reduces the optimal tax rates. Section 4 is devoted to the quantitative exploration of our theoretical results. The last section provides conclusions. Proofs are relegated to the Appendix.

2. Framework

There is a continuum of agents with heterogeneous labor productivities. Each agent can work in the formal sector (formal economy), in the informal sector (shadow economy), or in both simultaneously (which we call moonlighting). The fundamental difference between the two sectors is that formal earnings are observed by the tax authority and can be used to determine individual income tax payments, while informal earnings are hidden and cannot be used to determine taxes. In addition, individual labor productivity can differ between the sectors and participation in the informal sector is subject to a fixed cost, which we describe below.

Individuals are heterogeneous with respect to two privately observed characteristics: a productivity type θ and a cost type κ. The productivity type θ determines the labor productivity in the formal economy $w^f(\theta)$ and in the shadow economy $w^s(\theta)$. Earnings from each sector are a product of the sectoral productivity and the labor supplied to that sector. We assume that both productivity functions are nonnegative and continuously differentiable with respect to θ, and that the formal productivity is strictly increasing. The productivity type θ is drawn from the finite interval $[\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$ according to twice continuously differentiable cumulative distribution function $F(\theta)$ with density $f(\theta)$.[7] Note that the distribution of productivities is effectively one dimensional: conditional on the formal productivity, all agents have the same shadow productivity. In Appendix C, we extend our main result to the case of two-dimensional distribution of productivities.

The cost type κ is a fixed cost of engaging in informal employment. It can be interpreted either as a technological constraint on tax evasion or a utility cost of violating social norms.[8] Conditional on θ, the fixed cost is drawn from $[0, \infty)$ according to twice

[6]In principle, value added taxation covers informal firms indirectly if they purchase intermediate goods from the formal firms. De Paula and Scheinkman (2010) show that exactly for this reason, informal firms tend to make transactions with other informal firms. Bachas, Gadenne, and Jensen (2020) discuss more evidence that informal enterprises do not remit consumption taxes.

[7]We allow for an unbounded productivity distribution, in which case $\lim_{\theta \to \overline{\theta}} w^f(\theta) = \infty$.

[8]In principle, we could introduce a fixed cost of formal employment as well. This would correspond to what Magnac (1991) calls a segmentation approach to informal labor markets, according to which shadow

continuously differentiable cumulative distribution function $G_\theta(\kappa)$ with density $g_\theta(\kappa)$. For a model without the fixed cost of shadow employment, see the earlier working paper version (Doligalski and Rojas (2016)).

The agents' utility over consumption c and labor n, net of the fixed cost of shadow employment, is $c - v(n)$, where v is increasing, strictly convex, twice differentiable, and satisfies $v'(0) = 0$.[9] Using this quasi-linear preference structure, which follows Atkinson (1990) and Diamond (1998), we characterize the entire Pareto frontier, which is invariant to any increasing transformation of the utility function. Hence, our results are applicable also with utility functions $\mathbb{G}(c - v(n))$, where $\mathbb{G}$ is a strictly increasing and concave function. Nevertheless, this approach rules out the income effect. The impact of the income effect on the optimal tax schedules is well understood since Saez (2001) and the analysis can be easily extended in this direction.

Assumption 1 below ensures that the Spence–Mirrlees single-crossing condition holds for all workers. Consequently, formal income is increasing in productivity type θ even if agents are working informally, which helps us to keep track of workers' responses to tax reforms. The single-crossing condition holds when the comparative advantage in shadow labor $w^s(\theta)/w^f(\theta)$ is decreasing with formal productivity. In Section 4, we verify that this assumption holds in the data for Colombia.

Assumption 1 (Single-Crossing Condition). *The ratio $w^s(\theta)/w^f(\theta)$ is strictly decreasing with θ or $w^s(\theta) = 0$ for all θ.*

Lemma 1. *The formal earnings of formal workers and of moonlighting workers are increasing with productivity type θ.*

2.1 *Equilibrium income choices*

Consider a potentially nonlinear, twice differentiable income tax schedule T with tax rates strictly lower than 100%.[10] Denote the after-tax income schedule by $R(y) = y -$

---

workers are restricted from formal employment by various labor regulations. An alternative, competitive approach is that individuals sort between the two sectors according to their individual advantage, which corresponds more closely to our framework. Magnac (1991) shows that the data on married women in Colombia favor the latter, competitive approach. It has also been documented that informality is not driven by entry costs to the formal sector in other settings, e.g., in Argentina (Pratap and Quintin (2006)), Brazil (Rocha, Ulyssea, and Rachter (2018)), and Sri Lanka (De Mel, McKenzie, and Woodruff (2013)).

[9]We assume that disutility from working depends only on the total labor supply and not on the split of labor between sectors. This gives us tractability when describing the earnings responses of moonlighting agents, since the cost of substituting the formal and the informal labor depends only on their respective productivities and the tax schedule.

[10]We rule out tax kinks and, hence, bunching of different types along the productivity dimension alone. This kind of bunching is already well understood (Mussa and Rosen (1978), Ebert (1992)), it happens rarely, and is more important in the setting without the fixed cost of shadow employment (Doligalski and Rojas (2016)). We allow for all other bunching patterns, most importantly the bunching of agents with simultaneously different cost and productivity types, which happens when there are formal and moonlighting workers with the same formal earnings. Regarding the assumption of tax rates being below 100%, it is always satisfied in the optimum. Both assumptions combined imply that formal earnings of formal and moonlighting workers are strictly increasing with θ.

$T(y)$. A worker of type (θ, κ) chooses formal earnings y^f and informal earnings y^s by solving the maximization problem

$$V(\theta, \kappa) = \max_{y^f \geq 0, y^s \geq 0} R\big(y^f\big) + y^s - v\left(\frac{y^f}{w^f(\theta)} + \frac{y^s}{w^s(\theta)}\right) - \kappa \cdot \mathbb{1}\big(y^s > 0\big), \tag{1}$$

where $V(\theta, \kappa)$ is the indirect utility function and $y^i/w^i(\theta)$ stands for the labor supplied to sector $i \in \{f, s\}$. The optimal income choices are not necessarily unique, for instance, a worker with $w^f = w^s$ and $\kappa = 0$ who faces no income tax is always indifferent between supplying formal and informal labor. To have a clearcut characterization of income choices, we introduce the following tie-breaking rule.

Assumption 2. *A worker who is indifferent between multiple formal income levels chooses the highest one.*

We denote the income choices that solve the worker's problem (1) under Assumption 2 by $\mathrm{y}^f(\theta, \kappa)$ and $\mathrm{y}^s(\theta, \kappa)$.

The fixed cost κ affects the worker's decision whether to participate in the shadow economy. Beyond this decision, income choices are unaffected by κ. This allows us to summarize the income choices of θ workers in the following way. Suppose that a θ worker with fixed cost κ finds it optimal to participate in the shadow economy. Naturally, all θ workers with lower fixed cost will also choose to supply informal labor. It follows that there exists a threshold $\tilde{\kappa}(\theta)$ such that workers with fixed costs below the threshold join the shadow economy, while workers with fixed costs above the threshold remain fully formal. Define $\overline{y}^f(\theta)$ as the formal income of θ workers who choose to remain entirely formal. Define $\underline{y}^f(\theta)$ and $y^s(\theta)$ as the formal and the informal earnings, respectively, of θ workers who earn some informal income. Naturally, $\overline{y}^f(\theta) > \underline{y}^f(\theta)$. The value of threshold $\tilde{\kappa}(\theta)$ then follows from the indifference between participating or not in the shadow economy of a θ worker with the fixed cost equal exactly $\tilde{\kappa}(\theta)$:

$$R\big(\underline{y}^f(\theta)\big) + y^s(\theta) - v\left(\frac{\underline{y}^f(\theta)}{w^f(\theta)} + \frac{y^s(\theta)}{w^s(\theta)}\right) - \tilde{\kappa}(\theta) = R\big(\overline{y}^f(\theta)\big) - v\left(\frac{\overline{y}^f(\theta)}{w^f(\theta)}\right). \tag{2}$$

It is possible that for some θ, none of the workers chooses to earn informal income, which may happen if the shadow productivity or the marginal tax rates are sufficiently low. In this case, we set $\underline{y}^f(\theta) = \overline{y}^f(\theta)$, $y^s(\theta) = 0$, and $\tilde{\kappa}(\theta) = 0$.[11]

The income choices of any type (θ, κ) then follow

$$\mathrm{y}^f(\theta, \kappa) = \begin{cases} \overline{y}^f(\theta) & \text{if } \kappa \geq \tilde{\kappa}(\theta) \\ \underline{y}^f(\theta) & \text{otherwise,} \end{cases} \qquad \mathrm{y}^s(\theta, \kappa) = \begin{cases} 0 & \text{if } \kappa \geq \tilde{\kappa}(\theta) \\ y^s(\theta) & \text{otherwise.} \end{cases} \tag{3}$$

[11] Equivalently, we can define these objects as $\overline{y}^f(\theta) = \lim_{\kappa \to \infty} \mathrm{y}^f(\theta, \kappa)$, $\underline{y}^f(\theta) = \mathrm{y}^f(\theta, 0)$, $y^s(\theta) = \mathrm{y}^s(\theta, 0)$, and $\tilde{\kappa}(\theta) = V(\theta, 0) - \lim_{\kappa \to \infty} V(\theta, \kappa)$ for all θ.

Now let us characterize the income schedules $\overline{y}^f$, $\underline{y}^f$, and y^s. The first-order condition of the θ worker who works only in the formal sector pins down $\overline{y}^f(\theta)$:

$$\left(1 - T'\left(\overline{y}^f(\theta)\right)\right) \cdot w^f(\theta) = v'\left(\frac{\overline{y}^f(\theta)}{w^f(\theta)}\right). \tag{4}$$

According to this condition, the marginal return to formal labor—the product of the formal productivity and the net-of-tax rate—is equal to the marginal disutility from labor. Thus, the worker cannot gain by marginally adjusting formal labor.

Suppose that some θ workers are working only in the shadow economy. Then $\underline{y}^f(\theta) = 0$ and $y^s(\theta) > 0$, and it must be the case that

$$\left(1 - T'(0)\right) \cdot w^f(\theta) < v'\left(\frac{y^s(\theta)}{w^s(\theta)}\right) = w^s(\theta). \tag{5}$$

By the equality on the right-hand side, the return to informal labor, given by $w^s(\theta)$, is equal to the marginal disutility from labor. Thus, there are no gains to be made from marginally adjusting informal labor supply. The inequality on the left ensures that the worker also cannot benefit on the margin from starting to work formally.

Finally, suppose that some θ workers are moonlighting, i.e., working in the two sectors simultaneously: $\underline{y}^f(\theta) > 0$ and $y^s(\theta) > 0$. Their income choices satisfy

$$\left(1 - T'\left(\underline{y}^f(\theta)\right)\right) \cdot w^f(\theta) = v'\left(\frac{\underline{y}^f(\theta)}{w^f(\theta)} + \frac{y^s(\theta)}{w^s(\theta)}\right) = w^s(\theta). \tag{6}$$

Intuitively, in this case, the worker cannot gain by either (i) adjusting only formal labor, (ii) adjusting only informal labor, or (iii) shifting labor between the two sectors while keeping total labor supply fixed.[12]

The first-order condition of moonlighting workers has two important implications. First, the right-hand equality means that the total labor supply of a moonlighting worker is fully determined by the shadow productivity and, hence, cannot be affected by taxes. What taxes affect is only the sectoral split of labor. Second, moonlighting is closely related to tax progressivity. Condition (6) implies that $T'(\underline{y}^f(\theta)) = 1 - w^s(\theta)/w^f(\theta)$, where the right-hand side is strictly increasing with θ by Assumption 1. Thus, the marginal tax rates faced by moonlighting workers are strictly increasing with their productivity type. The proposition below explores the implications of this result. We show that moonlighting happens only where the tax is strictly progressive, i.e., has strictly increasing marginal tax rates.

Proposition 1. *If the tax schedule is weakly regressive locally at some $y > 0$, i.e., $T''(y) \leq 0$, then there are no moonlighting workers with formal earnings y.*

[12] Conditions (5) and (6) bound the marginal rate of substitution from below by $w^s(\theta)$. Similar constraints were found in other settings with hidden side trades, e.g., hidden saving; see Ábrahám and Pavoni (2005) and Golosov and Tsyvinski (2007).

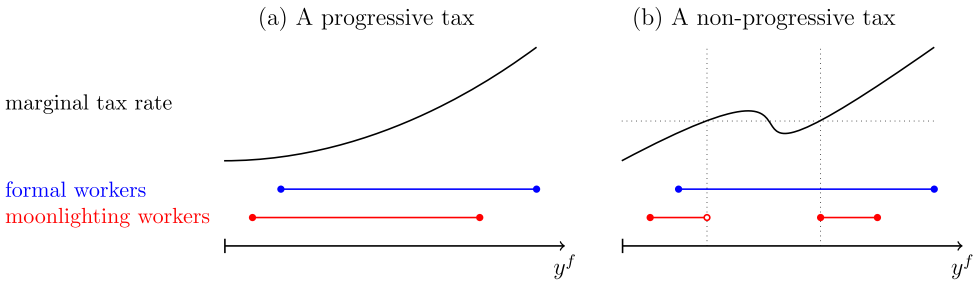

FIGURE 1. Tax progressivity and continuity of $\underline{y}^f(\cdot)$. The horizontal lines indicate that there are workers of a given kind at a given formal income level.

The intuition is that workers will be moonlighting if the marginal benefit to supplying formal labor relative to informal labor is decreasing. In that case, workers supply formal labor at first, but as the marginal benefit decreases sufficiently, they switch to the informal labor. That is exactly what happens when the tax schedule is progressive: low marginal tax rates at low income levels encourage formal labor at first, but high tax rates at higher levels discourage it.

What happens with moonlighting when the tax schedule is neither progressive nor regressive everywhere, but has regions of local progressivity and regressivity? Empirical income tax and transfer schedules, which typically have increasing statutory income tax rates, often become locally regressive where transfers are phased out. By Proposition 1, no moonlighting worker will be found in the regions of local regressivity. If such regions are surrounded by regions of local progressivity, then the formal income schedule of moonlighting workers can become discontinuous, as depicted in Figure 1.

Although tax regressivity may lead to a discontinuity in the formal income schedule of moonlighting workers, their indirect utility function must remain continuous, which yields an additional equilibrium condition.[13] Suppose that $\underline{y}^f(\cdot)$ increases discontinuously at θ_d and denote the left limit at θ_d by $\underline{y}^f(\theta_d^-)$, so that we have $\underline{y}^f(\theta_d^-) < \underline{y}^f(\theta_d)$.[14] In equilibrium it must be the case that

$$\lim_{\theta \uparrow \theta_d} V(\theta, \kappa) = V(\theta_d, \kappa) \quad \text{for all } \kappa < \tilde{\kappa}(\theta), \tag{7}$$

since otherwise some types in the neighborhood of θ_d could improve by jumping across the income discontinuity. It is useful to rewrite this condition as

$$\frac{T\big(\underline{y}^f(\theta_d)\big) - T\big(\underline{y}^f(\theta_d^-)\big)}{\underline{y}^f(\theta_d) - \underline{y}^f\big(\theta_d^-\big)} = T'\big(\underline{y}^f(\theta_d)\big). \tag{8}$$

Thus, the average tax rate on incomes between $\underline{y}^f(\theta_d^-)$ and $\underline{y}^f(\theta_d)$ is equal to the marginal tax rate $T'(\underline{y}^f(\theta_d))$. Note that while the sectoral split of labor changes discontinuously at θ_d, the total labor supply remains continuous, as it is pinned down by $w^s(\theta)$:

[13] By Corollary 1 in Milgrom and Segal (2002), the value function is (absolutely) continuous.

[14] By Assumption 2, $\underline{y}^f(\cdot)$ is right continuous and, hence, $\underline{y}^f(\theta_d^+) = \underline{y}^f(\theta_d)$.

recall condition (6). Condition (8) ensures then that the increase of formal after-tax income is exactly offset by the reduction of informal income such that consumption—and, hence, utility—remains continuous as well.[15]

In principle, the income schedule of formal workers also can be discontinuous. This happens when tax regressivity is strong enough such that the second-order condition of formal workers ceases to hold as a strict inequality; see Bergstrom and Dodds (2021) for a detailed analysis of such a case. For simplicity of exposition, we abstract from such a possibility in the theoretical analysis with the following assumption. We verify that it holds in all our quantitative exercises.

Assumption 3. *The income schedule of formal workers $\overline{y}^f(\cdot)$ is continuous and the utility-maximizing earnings level of formal $\underline{\theta}$ workers is unique.*

3. Optimal tax schedule

In this section we will derive and characterize the optimal tax schedule. We consider a general social welfare function

$$\mathcal{W} = \int_{\underline{\theta}}^{\overline{\theta}} \int_0^\infty \lambda(\theta, \kappa) \cdot V(\theta, \kappa)\, dG_\theta(\kappa)\, dF(\theta). \tag{9}$$

We normalize the Pareto weights λ such that their population average is 1, which implies that they coincide with the marginal social welfare weights.[16] The tax schedule is optimal if it maximizes the social welfare function subject to the government budget constraint

$$\mathcal{TR} = \int_{\underline{\theta}}^{\overline{\theta}} \int_0^\infty T\big(y^f(\theta, \kappa)\big)\, dG_\theta(\kappa)\, dF(\theta) \geq E, \tag{10}$$

where E stands for exogenous government expenditures. By finding the optimal tax schedule for arbitrary welfare weights, we recover the entire Pareto frontier of the model without income effects.[17]

From now on we will focus on the endogenous distribution of formal income. Denote the cumulative distribution function (cdf) and the density of formal income by H

[15]As formal earnings increase from $y^f(\theta_d^-)$ to $y^f(\theta_d)$, formal after-tax income increases by $R(y^f(\theta_d)) - R(y^f(\theta_d^-))$ and the shadow income decreases by $(y^f(\theta_d) - y^f(\theta_d^-)) \cdot w^s(\theta_d)/w^f(\theta_d)$. Requiring that the two are equal and using (6) yields (8).

[16]The marginal social welfare weight describes the welfare impact of marginally increasing consumption of a given agent, expressed in the units of tax revenue (see, e.g., Piketty and Saez (2013)). In our environment, it is equal to $\lambda(\theta, \kappa)/\eta$, where η is the multiplier of the government budget constraint. It is easy to show that at the optimum, η is equal to the average Pareto weight.

[17]Suppose that the social welfare function is $\int_{\underline{\theta}}^{\overline{\theta}} \int_0^\infty \mathbb{G}(V(\theta, \kappa))\, dG_\theta(\kappa)\, dF(\theta)$, where $\mathbb{G}$ is an increasing and differentiable function. The function $\mathbb{G}$ is typically assumed to be strictly concave and it can represent either the decreasing marginal utility of consumption or the social taste for equality. In this case, we find the optimal tax schedule iteratively. Start with an initial guess of the Pareto weights. In each step, find the optimal tax schedule and the indirect utility function V given the Pareto weights, and set the new Pareto weights—to be used in the next step—according to $\lambda(\theta, \kappa) = \mathbb{G}'(V(\theta, \kappa))$.

and h, respectively. We can decompose it into the (scaled) cdf of earnings of formal workers H^f and the (scaled) cdf of formal earnings of workers with some shadow income H^s, such that $H = H^f + H^s$, with the corresponding (scaled) densities h^f and h^s.[18] Denote the average welfare weight at formal earnings y by $\overline{\lambda}(y)$.

We derive the optimal tax schedule with the tax perturbation approach that originate from Saez (2001) and was further refined by Golosov, Tsyvinski, and Werquin (2014) and Jacquet and Lehmann (2021). Consider a status quo tax schedule T and a new tax schedule $T + \mu \cdot \mathrm{d}T$, where the schedule $\mathrm{d}T$ indicates the *direction* of the tax reform and the scalar μ controls the *size* of the reform. We describe the impact of arbitrary tax reforms on the equilibrium outcomes with a Gateaux derivative. The Gateaux derivative of some functional $T \mapsto Z[T]$ in the direction $\mathrm{d}T$ is defined as

$$\mathrm{d}Z[T, \mathrm{d}T] = \lim_{\mu \to 0} \frac{Z[T + \mu \cdot \mathrm{d}T] - Z[T]}{\mu}. \tag{11}$$

For instance, consider formal earnings $\mathrm{y}^f(\theta, \kappa)$ as a functional of the tax schedule. The expression $\mathrm{d}\mathrm{y}^f(\theta, \kappa)[T, \mathrm{d}T]$ informs us about the first-order impact of a small reform in the direction $\mathrm{d}T$ on the formal earnings of workers with type (θ, κ). Typically we omit the arguments and write it simply as $\mathrm{d}\mathrm{y}^f(\theta, \kappa)$.

To use Gateaux derivatives, we need to ensure that such derivatives exist. There are two potential issues. First, the formality threshold $\tilde{\kappa}(\theta)$ is bounded from below by 0 and, thus, can be non-differentiable with respect to tax reforms at the bound. Intuitively, if nobody works in the shadow economy to start with, then increasing taxes can increase informality, but decreasing taxes cannot reduce it. Second, some moonlighting workers respond by jumping to a discretely different formal income level and their formal earnings are not differentiable. Lemma 2 ensures that the Gateaux derivative of the aggregate tax revenue $\mathcal{TR}$ exists nonetheless. To address the first issue, we show that almost everywhere threshold $\tilde{\kappa}(\theta)$ either has a Gateaux derivative or it does not affect tax revenue to the first order. To address the second issue, we express $\mathcal{TR}$ as the sum of integrals over the regions where income responses are differentiable, with jump responses accounted for by the endogenous edges of the integration regions. We then apply Leibniz integral rule to show that each of the integrals and, hence, $\mathcal{TR}$ as a whole, has a Gateaux derivative.[19]

Lemma 2. *The Gateaux derivative of the aggregate tax revenue $\mathcal{TR}$ in an arbitrary direction $\mathrm{d}T \in \mathcal{C}^2$ exists.*

[18]The functions $H^f(\cdot)$ and $H^s(\cdot)$ are *scaled* cdfs as they do not converge to 1 as $y \to \infty$, but rather to the shares of formal and non-formal workers in total employment, respectively.

[19]Golosov, Tsyvinski, and Werquin (2014) ensures that $\mathrm{d}\mathcal{TR}$ exists by assuming sufficient smoothness of income choices with respect to tax reforms. Jacquet and Lehmann (2021) use implicit function theorem to show that income choices are differentiable when jump responses are ruled out. Hendren (2019) speculates that $\mathrm{d}\mathcal{TR}$ may exist even with jump responses, but does not prove it. To the best of our knowledge, Bergstrom and Dodds (2021) are the first to use the Leibniz integral rule to obtain $\mathrm{d}\mathcal{TR}$ with jump responses.

Consider a tax reform in direction $\mathrm{d}T$. The reform affects the social welfare and the tax revenue. The latter impact can be decomposed into the mechanical effect, as well as the behavioral effects due to (i) intensive margin responses of formal and moonlighting workers, and (ii) extensive margin responses due to workers changing their informality status. We describe these effects below and then collect them into the optimal tax formula. The detailed derivations are available in Appendix B.

Mechanical and welfare effects. The tax reform increases tax level at formal earnings y by $\mathrm{d}T(y)$, which mechanically increases tax revenue. The impact on the utility of each agent earning y is exactly $-\mathrm{d}T(y)$, since behavioral responses have no first-order utility impact by the envelope theorem. The social welfare impact is then obtained by multiplying the utility impact with the average marginal social welfare weight $\overline{\lambda}(y)$. Integrating over the entire income distribution yields the mechanical (ME) and welfare (WE) impacts of the reform:

$$\mathrm{ME} = \int_0^\infty \mathrm{d}T(y)\, dH(y), \qquad \mathrm{WE} = -\int_0^\infty \overline{\lambda}(y) \cdot \mathrm{d}T(y)\, dH(y). \tag{12}$$

Intensive margin responses. Formal workers adjust their earnings on the intensive margin in response to changes of the marginal tax rates $\mathrm{d}T'$. The tax revenue loss at earnings level y is standard and equal to

$$T'(y) \cdot \tilde{\varepsilon}^f(y) \cdot y \cdot \frac{\mathrm{d}T'(y)}{1-T'(y)}, \quad \text{where } \tilde{\varepsilon}^f(y) = \left(\frac{1}{\varepsilon(y)} + \frac{T''(y) \cdot y}{1-T'(y)}\right)^{-1}. \tag{13}$$

The term $\tilde{\varepsilon}^f(y)$ is the elasticity of earnings of formal workers with respect to the net-of-tax rate $1 - T'(y)$. It depends both on $\varepsilon(y)$—the elasticity along the *linear* tax schedule or the Frisch elasticity—and the local tax curvature. With a locally progressive tax ($T''(y) > 0$), income increase in response to a tax rate cut is dampened, as higher income leads to a higher tax rate. Hence, local tax progressivity (resp. regressivity) reduces (resp. increases) the elasticity of income.

Suppose that there are some moonlighting workers with formal income y and that their formal earnings schedule is locally continuous. The tax revenue loss due to the reduction of formal earnings of moonlighting workers is equal to

$$T'(y) \cdot \tilde{\varepsilon}^s(y) \cdot y \cdot \frac{\mathrm{d}T'(y)}{1-T'(y)}, \quad \text{where } \tilde{\varepsilon}^s(y) = \frac{1-T'(y)}{T''(y) \cdot y}. \tag{14}$$

The formal earnings responses of moonlighting workers are summarized by elasticity $\tilde{\varepsilon}^s(y)$.[20] We find that formal earnings of moonlighting workers are more elastic than those of exclusively formal workers: $\tilde{\varepsilon}^s(y) > \tilde{\varepsilon}^f(y)$. The intuition behind this result is tightly related to the first-order conditions (4) and (6). An increase of the tax rate reduces the marginal benefit from supplying formal labor for formal and moonlighting workers in a symmetric manner. Both formal and moonlighting workers will reduce formal labor supply until the marginal benefit increases up to the level of the marginal cost. The

[20]By Proposition 1, we know that $T''(y) > 0$. Otherwise, there would be no moonlighting workers with such earnings.

difference between them is in the determination of the marginal cost of formal labor. For the formal worker, the marginal cost is the marginal disutility of labor $v'(\cdot)$, which decreases as the total labor supply is reduced. For the moonlighting worker, however, the total labor supply is fixed and the tax reform affects only the sectoral split of labor. The marginal cost for these workers is the forgone informal income, which is equal to the shadow productivity $w^s(\theta)$. Given that the marginal cost of the moonlighting workers is constant in formal labor, rather than decreasing as in the case of the formal workers, they will adjust formal labor more than formal workers.

As we discussed, the formal income schedule of the moonlighting workers can become discontinuous when the tax schedule is not fully progressive. Suppose that formal earnings of moonlighting workers increase discontinuously between levels $\underline{s}$ and $\overline{s}$. By condition (8), the workers at these income levels are indifferent between earning $\underline{s}$ or $\overline{s}$. Thus, a tax reform that changes the relative tax burden at these earnings, e.g., an increase of the marginal tax rate at some earnings $y \in (\underline{s}, \overline{s})$, will imply a discrete jump of workers between $\underline{s}$ and $\overline{s}$, as depicted in Figure 2. Suppose that a tax reform increases the relative tax burden by $\mathrm{d}[T(\overline{s}) - T(\underline{s})]$, which affects the higher level of earnings $\overline{s}$ by $\mathrm{d}\overline{s}$. As a result, the measure $\mathrm{d}\overline{s} \cdot h^s(\overline{s})$ of moonlighting workers reduces formal earnings by a discrete amount $\overline{s} - \underline{s}$. In Appendix B, we show that formal earnings loss due to these jump responses satisfies

$$-(\overline{s} - \underline{s}) \cdot \mathrm{d}\overline{s} \cdot h^s(\overline{s}) = \tilde{\varepsilon}^s(\overline{s}) \cdot \overline{s} \cdot h^s(\overline{s}) \cdot \frac{\mathrm{d}\big[T(\overline{s}) - T(\underline{s})\big]}{1 - T'(\overline{s})}. \tag{15}$$

We find that although each individual earnings response is discrete, the sum of responses is described with a well behaved elasticity. The intuition is that although each jumping individual reduces formal earnings by $\overline{s} - \underline{s}$, the measure of jumping individuals is inversely proportional to $\overline{s} - \underline{s}$. Hence, the total formal income reduction is independent of the size of the jump. Furthermore, the elasticity describing the jumping responses is exactly the same as the elasticity of moonlighting workers responding marginally, $\tilde{\varepsilon}^s$. It follows from the indifference condition (8) that states that the average tax rate over earnings interval $[\overline{s}, \underline{s}]$ is equal to the marginal tax rate at $\overline{s}$. Using this

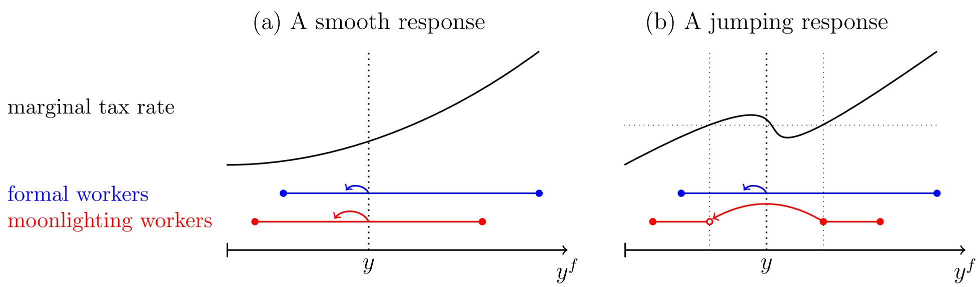

Figure 2. Intensive margin responses of moonlighting workers. The horizontal lines indicate that there are workers of a given kind at a given formal income level. The arrows represent the formal income responses to an increase of $T'(y)$.

equality of the average and the marginal tax rates, we can represent the tax revenue loss from these jump responses as

$$-\big(T(\overline{s})-T(\underline{s})\big)\cdot \mathrm{d}\overline{s}\cdot h^s(\overline{s}) = T'(\overline{s})\cdot \tilde{\varepsilon}^s(\overline{s})\cdot \overline{s}\cdot h^s(\overline{s})\cdot \frac{\mathrm{d}\big[T(\overline{s})-T(\underline{s})\big]}{1-T'(\overline{s})}. \tag{16}$$

Collecting the terms, intensive margin responses (either smooth responses or jumps) have the impact on the aggregate tax revenue

$$\begin{aligned}\mathrm{BE}_{\mathrm{int}} = &-\int_0^\infty \frac{T'(y)}{1-T'(y)}\cdot y\cdot \big(\tilde{\varepsilon}^f(y)\cdot h^f(y)+\tilde{\varepsilon}^s(y)\cdot h^s(y)\big)\cdot \mathrm{d}T'(y)\,dy \\ &-\sum_{(\overline{s},\underline{s})\in D}\frac{T'(\overline{s})}{1-T'(\overline{s})}\cdot \tilde{\varepsilon}^s(\overline{s})\cdot \overline{s}\cdot h^s(\overline{s})\cdot \mathrm{d}\big[T(\overline{s})-T(\underline{s})\big],\end{aligned} \tag{17}$$

where set D contains pairs of income levels between which the moonlighting workers jump.[21] The first term describes the tax revenue impact of the marginal responses of the formal and the moonlighting workers, respectively. The second term captures the impact of the jumping responses of moonlighting workers.

Extensive margin responses. These responses consist of switching from working exclusively formally to either moonlighting or working exclusively informally. The possibility of moonlighting means that the extensive margin responses are not equivalent to responses on the formal participation margin. In particular, a worker who switches from exclusively formal employment to moonlighting continues to work in the formal sector and retains a fraction of formal earnings. This has important implications for the incidence of the extensive margin responses, as depicted in Figure 3. The depicted tax reform increases the tax burden for workers with incomes above y. Consequently, incentives for informality increase for formal agents with earnings above y who, conditional on joining the shadow economy, would earn less than y in the formal sector. On the other hand, incentives for informality are unaffected for formal workers who, even

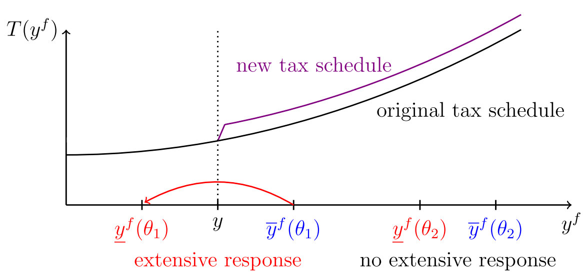

FIGURE 3. The incidence of extensive margin responses. An increase of $T'(y)$ triggers an extensive margin response for workers with productivity type θ_1, but not for workers with productivity type θ_2.

[21] Formally, $D=\{(\overline{s},\underline{s})\in\mathbb{R}^2_+ : \underline{s}=\lim_{\theta\uparrow\theta_d}\underline{y}^f(\theta)<\lim_{\theta\downarrow\theta_d}\underline{y}^f(\theta)=\overline{s}$ for some $\theta_d\in[\underline{\theta},\overline{\theta}]\}$. Since $\underline{y}^f$ is increasing, it has countably many discontinuity points and, hence, D is countable.

if they moonlighted, would have formal income above y; they would pay a higher tax either way.

To capture the tax revenue impact of the extensive margin responses, consider formal workers with earnings z. Denote by $\rho(z) = \underline{y}^f(\overline{y}^{f^{-1}}(z))$ their formal earnings if they had lower realization of fixed cost of informal employment κ and (potentially) worked informally.[22] Note that $\rho(z)$ can be zero or positive. Now we can define the *tax burden of staying formal* as $\Delta T(z) = T(z) - T(\rho(z))$. The tax revenue impact of the extensive margin responses of these workers is then

$$\Delta T(z) \cdot \mathrm{d}h^f(z) = -\Delta T(z) \cdot \pi(z) \cdot h^f(z) \cdot \mathrm{d}\Delta T(z), \tag{18}$$

where $\pi(z)$ is the semi-elasticity of the density of formal workers with respect to the tax burden of staying formal, defined as

$$\pi(z) = \frac{g_\theta\big(\tilde{\kappa}(\theta)\big)}{1 - G_\theta\big(\tilde{\kappa}(\theta)\big)} \frac{1}{h^f(z)} \quad \text{for } \theta \text{ such that } \overline{y}^f(\theta) = z. \tag{19}$$

Thus, the extensive margin responses are more costly in terms of government revenue when the tax burden of staying formal is higher and when the density of agents at the threshold $\tilde{\kappa}(\theta)$ is higher, which translates into larger semi-elasticity of earnings density $\pi(z)$. Aggregating extensive responses across all income levels, we have

$$\mathrm{BE}_{\mathrm{ext}} = -\int_0^\infty \Delta T(z) \cdot \pi(z) \cdot h^f(z) \cdot \mathrm{d}\Delta T(z)\, dz. \tag{20}$$

Optimal tax formula. At the optimal tax schedule, no small tax reform can result in a gain in the welfare-adjusted tax revenue. Hence, the sum of all the effects of the tax reform needs to be zero:

$$\mathrm{ME} + \mathrm{WE} + \mathrm{BE}_{\mathrm{int}} + \mathrm{BE}_{\mathrm{ext}} = 0 \tag{21}$$

for an arbitrary direction of the tax reform. The following theorem expresses this condition as a Diamond–Saez formula for economies with an informal sector.

Theorem 1. *Suppose that the optimal tax schedule is twice continuously differentiable. The optimal tax rate at earnings y satisfies*

$$\begin{aligned} &\frac{T'(y)}{1 - T'(y)} \cdot \tilde{\varepsilon}^f(y) \cdot y \cdot h^f(y) + \frac{T'\big(\overline{s}(y)\big)}{1 - T'\big(\overline{s}(y)\big)} \cdot \tilde{\varepsilon}^s\big(\overline{s}(y)\big) \cdot \overline{s}(y) \cdot h^s\big(\overline{s}(y)\big) \\ &\quad = \int_y^\infty \big[1 - \overline{\lambda}(z)\big]\, dH(z) - \int_y^\infty \Delta T(z) \cdot \pi(z) \cdot \mathbb{1}\big(\rho(z) \le y\big)\, dH^f(z), \end{aligned} \tag{22}$$

where $\overline{s}(y)$*, equal to* $\min\{z \in \mathrm{Im}(\underline{y}^f) : z \ge y\}$ *if min exists and* 0 *otherwise, indicates the formal earnings of moonlighting workers distorted by a raise of* $T'(y)$.

[22]By assumptions made, $\overline{y}^f$ is strictly increasing and, thus, its inverse function $\overline{y}^{f^{-1}}$ exists.

Formula (22) equates costs and benefits of increasing marginal tax rate $T'(y)$. The left-hand side consists of the deadweight loss from distorting the formal workers and the moonlighting workers. Note that we combine the deadweight loss from smooth and jumping responses of moonlighting workers into a single term by using mapping $y \mapsto \bar{s}(y)$, which points to formal earnings where the distorted moonlighting workers are located. The deadweight loss terms increase in (i) the marginal tax rate, as the reduction in formal income implies a higher tax loss if it is taxed at the higher rate, (ii) the density of formal income, and (iii) the formal income reduction per worker in response to a higher tax rate, i.e., the product of formal income and the income elasticity.

There are two important differences between the deadweight loss terms of formal and moonlighting workers. The first difference relates to the location of responses. The raise of $T'(y)$ triggers intensive margin responses of formal workers with earnings y. In contrast, the intensive margin responses of moonlighting workers happen at formal earnings level $\bar{s}(y)$ that can be strictly greater than y. In particular, if $\bar{s}(y) = y$, then the moonlighting workers respond smoothly, while in the case of $\bar{s}(y) > y$, they respond by jumping to discretely lower formal earnings. The second difference relates to the size of responses. Conditional on the local progressivity of the tax schedule, the moonlighting workers are more elastic than the formal workers. Thus, ceteris paribus, increasing tax rates at incomes that are earned predominantly by moonlighting workers is more distortionary.

The last term on the right-hand side represents the tax loss from increased participation in the shadow economy. Importantly, moonlighting modifies the incidence of these extensive margin responses. Absent moonlighting, an increase of $T'(y)$ increases incentives for informality for all formal workers earning more than y. With moonlighting, the incentives for informality increase only at some formal earnings levels $z > y$—those for which $\rho(z) \leq y$ holds. Intuitively, if formal workers earning $z > y$ retained formal income of $\rho(z) > y$ even if they started moonlighting, then the raise of $T'(y)$ would not affect their incentives to take up an informal job. The remaining terms on the right-hand side of the formula capture the mechanical and welfare impacts of the reform, which are standard.

The novelty of the tax formula is due to the moonlighting responses, namely complementing formal earnings with income from an informal job. There are other settings that give rise to phenomena similar to moonlighting where such a formula could be applied, e.g., the model of home production or the problem of a local tax authority with residents who can work partly outside its jurisdiction as seasonal workers; see Mirrlees (1982) for an early investigation of a similar problem.

3.1 *How does a shadow economy affect optimal tax rates?*

We examine the impact of a shadow economy on the optimal tax rates in two ways, similarly to the approach of Scheuer and Werning (2017). First, we fix the formal income distribution and other sufficient statistics, and compare the prescription of the optimal tax formula with the tax schedules that would be chosen by the tax authority that believed certain informality responses do not happen. Tax schedules chosen under such beliefs can be described with well known formulas from the literature. This analysis is most informative for choosing tax policy based on a given, observed formal income distribution. Second, we compare the optimal top tax rate with and without a shadow economy for given model primitives while allowing the formal income distribution to adjust. This comparison is useful for the counterfactual analysis. It informs us how the optimal top tax rate would change if we could costlessly shut down the informal sector.

3.1.1 *Comparison for a fixed formal income distribution* Taking the income distribution and other sufficient statistics as given, we will compare the prescriptions of the optimal tax formula and two other formulas, corresponding to different beliefs regarding the informality, which we describe below.

In the first case, the tax authority acknowledges the mobility between the formal and the informal sectors, but ignores the possibility of moonlighting. The tax formula corresponding to such beliefs is

$$\begin{aligned}\frac{T_{\mathrm{I}}'(y)}{1-T_{\mathrm{I}}'(y)}\cdot\overline{\varepsilon}(y)\cdot y\cdot h(y) &= \int_y^\infty\big[1-\overline{\lambda}(z)\big]\,dH(z)\\ &\quad-\int_y^\infty \Delta T(z)\cdot\pi(z)\cdot\mathbb{1}\big(\rho(z)=0\big)\,dH^f(z),\end{aligned}\tag{23}$$

where $\overline{\varepsilon}(y)$ is the average formal earnings elasticity at formal income y and $\pi(z)\cdot\mathbb{1}(\rho(z)=0)$ is the semi-elasticity of participation in the formal labor market at formal income z with respect to tax burden at z. The indicator function makes sure that only responses that reduce formal earnings to zero are accounted for. Such a formula was derived by Saez (2002) and Jacquet, Lehmann, and Van der Linden (2013) in the model with intensive margin responses and endogenous participation in the labor market.

In the second case, the tax authority ignores all informality responses. Then it would set the income tax according to the formula

$$\frac{T_{\mathrm{II}}'(y)}{1-T_{\mathrm{II}}'(y)}\cdot\overline{\varepsilon}(y)\cdot y\cdot h(y) = \int_y^\infty\big[1-\overline{\lambda}(z)\big]\,dH(z).\tag{24}$$

Here, the planner effectively believes in an extreme version of the segmented market hypothesis, where the allocation of workers to the formal and the informal sectors is given and policy invariant. In this view, the tax schedule affects only the labor supply of formal workers on the intensive margin. Hence, this tax formula coincides with the formula of Diamond (1998) and Saez (2001), derived in the model with intensive margin of labor supply alone.

The following proposition compares tax rates implied by the optimal formula, denoted by T_{opt}', with the rates prescribed by the two other formulas.

Proposition 2. *Fix the distribution of formal income, the schedule of Pareto weights $\overline{\lambda}$, and the values of all other sufficient statistics required to compute the optimal tax schedule according to Theorem 1. Suppose that the status quo tax schedule has nonnegative marginal tax rates and that all tax formulas prescribe a tax schedule that is twice differentiable. Then $T_{\mathrm{opt}}'(y)\le T_{\mathrm{I}}'(y)\le T_{\mathrm{II}}'(y)$ for all y.*

We obtain a clear ordering of marginal tax rates at each income level. The optimal tax formula prescribes the lowest rates, followed by the rates set when only moonlighting is ignored, and the highest rates are chosen when all informality responses are ignored. The intuition is simple: the optimal tax formula correctly incorporates the entire fiscal cost of raising tax rates, while the other formulas miss some cost terms: the deadweight loss from jump responses of moonlighting workers and—in the case of formula I only partially—the extensive margin responses. Note that this result holds for any schedule of Pareto weights. Thus, the same ranking of tax rates holds also for the upper bound of the Pareto efficient (or, alternatively, revenue-maximizing) marginal tax rates: that case corresponds to setting $\overline{\lambda}(z) = 0$ at each positive income level. Also note that in Section 4, we conduct a similar analysis quantitatively while allowing for the endogenous adjustment of the income distribution.

3.1.2 *Comparison for fixed primitives: Top tax rate* Let τ be the tax rate in top bracket $[z^*, \infty)$. As a first step, let us determine how the top tax rate influences the upper tail of formal earnings. It is useful to denote the ratio of shadow and formal productivity at the top by $\phi = \lim_{\theta \to \overline{\theta}} w^s(\theta)/w^f(\theta)$.

Lemma 3. *Suppose that in the top bracket, (i) formal productivity $w^f(\theta)$ is Pareto distributed with coefficient α, (ii) fixed cost of shadow employment κ is Pareto distributed with coefficient γ, and (iii) the Frisch elasticity of labor supply is ε. Suppose further that marginal tax rates are nondecreasing. Then the tail parameter of the formal income distribution $\alpha_y = \lim_{y \to \infty} yh(y)/(1 - H(y))$ satisfies*

$$\alpha_y = \begin{cases} \dfrac{\alpha}{1+\varepsilon} & \text{if } \tau < 1 - \phi \\ \dfrac{\alpha}{1+\varepsilon} + \gamma & \text{if } \tau > 1 - \phi. \end{cases} \tag{25}$$

The inverse of parameter α_y describes the *thickness* of the upper tail of the formal income distribution. When the top tax rate is sufficiently low, none of the most productive types works informally and the thickness of the formal income tail is exactly the same as in the standard Mirrlees model. As soon as the top tax rate crosses a tipping point $1 - \phi = 1 - \lim_{\theta \to \overline{\theta}} w^s(\theta)/w^f(\theta)$, a positive fraction of top earners joins the shadow economy. As a result, the thickness of the upper tail falls to a discretely lower value.[23] The size of this discrete fall is increasing with γ, the tail parameter of the fixed cost distribution. Intuitively, if γ is high, there are many workers with a low fixed cost of shadow employment who reduce their formal income and join the shadow economy. If instead γ is low, there are few workers with a low fixed cost of shadow employment and the formal income distribution is less affected.[24]

[23]Note that this statement does not contradict the differentiability of the aggregate tax revenue with respect to tax reforms, since we are discussing the impact on the shape of the income distribution in the limit as $y \to \infty$. For any finite y, the ratio $yh(y)/(1 - H(y))$ is differentiable with respect to the top tax rate.

[24]Regarding the knife-edge case $\tau = 1 - \phi$, we can construct productivity schedules such that α_y takes any value between $\alpha/(1+\varepsilon)$ and $\alpha/(1+\varepsilon) + \gamma$.

Now we will compare the optimal top tax rate from our model, denoted by τ^*, with the the optimal top rate in the model where the informal sector does not exist, denoted by τ_M. Since the model without the informal sector is just the standard Mirrlees model, we call τ_M a Mirrleesian top tax rate. In this comparison, we take as given model primitives (the distribution of productivity and cost types, the productivity schedules, and the schedule of Pareto weights) and we allow the income distribution and all other sufficient statistics to endogenously adjust to the top tax rate. To obtain analytical results, we consider the limiting top bracket as $z^* \to \infty$. The following proposition shows that the shadow economy leads to a (weakly) lower optimal top tax rate.

Proposition 3. *Suppose that the assumptions of Lemma 3 hold and that the average Pareto weight, weighted by earnings in the top bracket, is $\lambda \in [0, 1)$. Consider a sequence of top brackets $[z^*, \infty)$ such that $z^* \to \infty$ and suppose that when $\tau > 1 - \phi$, then there exists an upper bound on formal earnings of moonlighting workers that is independent of z^*. Then $\tau_M = \frac{1-\lambda}{1-\lambda+\frac{\alpha}{1+\varepsilon}\cdot\varepsilon}$ for all z^* and*

$$\lim_{z^*\to\infty} \tau^* = \begin{cases} \tau_M & \text{if } \tau_M \leq 1-\phi \\ 1-\phi & \text{if } \tau_M > 1-\phi. \end{cases} \tag{26}$$

From Lemma 3, we know that setting the top tax rate above $1 - \phi$ makes the upper tail of formal earnings thinner, as many top workers join the shadow economy. In the proof of Proposition 3, we show that the resulting tax revenue loss dominates any possible redistributive gain when z^* is sufficiently high. Thus, the shadow economy effectively imposes an upper bound $1 - \phi$ on the optimal top tax rate.[25] In contrast, a top rate below $1 - \phi$ does not give incentives for informality at the top. Thus, if $\tau_M < 1 - \phi$, the optimal rate is equal to τ_M.

Piketty, Saez, and Stantcheva (2014) study the optimal top tax rate with income shifting on the intensive margin only, when the shifted income is taxed with exogenous tax rate t. Setting $t = 0$, their income shifting is equivalent to informality. They show that in this case, the shifting responses affect the optimal top tax rate only by increasing the elasticity of reported (formal) income. Our result, derived in a somewhat different environment with intensive and extensive margin informality responses, is similar in flavor.[26] Namely, when $\tau < 1 - \phi$, then there are no informality responses and the elasticity of formal income is as in the standard Mirrlees model. On the other hand, when

[25]The assumption that formal earnings of moonlighting workers are bounded from above when $\tau > 1 - \phi$ rules out a contrived case where, as $z^* \to \infty$, the tax schedule below z^* adjusts such that $\underline{y}^f(\overline{\theta}) \to \infty$. Such a scenario is not policy relevant: even if the optimal nonlinear tax schedule had a top tax rate above $1 - \phi$, its marginal tax rate would cross $1 - \phi$ at some finite earnings level, providing an upper bound for $\underline{y}^f$.

[26]In Piketty, Saez, and Stantcheva (2014), the monetary cost of shifting income x is $d(x)$, which is an increasing and convex function satisfying $d'(0) = 0$. As a result, whenever $\tau > 0$, all agents misreport a positive share of income. In our model, the monetary cost of misreporting income by x can be expressed as $\kappa + (1 - w^s(\theta)/w^f(\theta))x$. Thus, agents misreport income only when the tax rate is sufficiently high (i.e., when $\tau > 1 - w^s(\theta)/w^f(\theta)$) and if they misreport, they do it to such an extent that they leave the top tax bracket entirely.

τ increases above $1 - \phi$, top workers start joining the informal sector, which drastically increases the elasticity of formal income and renders such a tax increase suboptimal.

4. Quantitative analysis

In this section, we explore the quantitative importance of our theoretical results. We estimate the model with data from Colombia and analyze the impact of informality responses on the the optimal tax schedule.

4.1 *Estimation*

We estimate the model using the household survey from Colombia, which allows us to identify the sector and hourly wage at the main job. We restrict attention to individuals aged 24–50 years without children (34,000 observations), since they face a tax and transfer schedule that does not depend on choices absent from our modeling framework, such as assets, the number of children, or college attainment. Below we explain how we identify informality in the data and introduce our estimation strategy. Further details are provided in Appendix D.

Identifying informality. We identify the main job of a given worker as informal if the worker reports not contributing to the mandatory social insurance programs. Since social insurance contributions are paid jointly with payroll taxes and the withheld part of the personal income tax, a worker who pays contributions is automatically subject to income taxation. Thus, this approach is particularly well suited for our exercise.[27] We find that 58% of all workers in Colombia in 2013 were employed informally at a main job, a result consistent with other indicators of informality in Colombia.[28] The average wage in the informal sector is about half of the average wage in the formal sector and the distributions of wages in the two sectors overlap significantly. Of the workers with a formal main job, about 6% have a secondary job. Some of them could be moonlighting in the shadow economy. However, the available data do not allow us to identify the sector of work in the second job.[29] Hence, we treat the informality status of the second job as a latent variable.

Estimation strategy. Using only the information on wages and sector of work does not allow us to identify the model unless we impose additional restrictions on the distribution of types (θ, κ) and productivity schedules. The reason is that any observed wage

[27]Detecting informality via social security contributions is broadly consistent with the methodology of the International Labour Organization (ILO (2013)) and is used by the Ministry of Labor of Colombia (ILO (2014)), as well as by Goldberg and Pavcnik (2003), Guataquí, García, and Rodríguez (2010), and Mora and Muro (2017) in the studies of Colombia.

[28]The official statistical agency of Colombia (DANE) follows an alternative measure of informality based on the size of the establishment, status in employment, and educational level of workers. They find that 57.3% and 56.7% of workers were informal in the first two quarters of 2013 (ILO (2014)), which is very close to 58% we find for the entire 2013.

[29]The only information we have about the second job is the number of hours worked in the week prior to the survey, the income from that job in the month prior to the survey, and whether the worker is an employee or a self-account worker. There is no information on the social security payments tied to this job.

distribution and sector allocation can be an equilibrium outcome of the model in which formal and shadow productivities are equal for all workers and all the sorting is driven by the fixed cost of informal employment κ.[30]

Our identification approach is based on the assumption that productivities are log normal, conditional on observable characteristics of workers.[31] Specifically, we map sectoral log productivities of individual i to her observable characteristics X_i as

$$\log(w_i^f) = \log(w_0^f) + \gamma^f \cdot (X_i\beta + \epsilon_i) \tag{27}$$

$$\log(w_i^s) = \log(w_0^s) + \gamma^s \cdot (X_i\beta + \epsilon_i), \tag{28}$$

where β is a vector of parameters and ϵ_i is a normally distributed error term.[32] In those equations, $X_i\beta + \epsilon_i$ is the empirical counterpart of the productivity type θ in the model. Parameters γ^f and γ^s allow for different slopes of the sectoral productivity schedules, while w_0^f and w_0^s control the levels.[33] This log-linear specification coincides with the commonly used Mincerian regressions (Heckman, Lochner, and Todd (2006)).

We cannot estimate the above equations directly as we would face a clear selection problem. Furthermore, we also need to estimate the distribution of the fixed cost κ. Thus, we carry on a structural estimation of the model. First, we assume that the fixed cost of informal employment κ follows a generalized Pareto distribution with the productivity-dependent scale parameter $\sigma_\kappa(w^f(\theta) - w_\kappa)^{\alpha_\kappa}$, where σ_κ, w_κ, and α_κ are parameters to be estimated. When $\alpha_\kappa = 0$, the distribution of κ is independent of the productivity type θ. Second, we specify that disutility from labor is given by $v(n) = \Gamma n^{1+1/\varepsilon}/(1+1/\varepsilon)$ with intensive margin elasticity ε equal to 0.33 following Chetty (2012).

The observed log hourly wage $\mathcal{W}$ and the indicator of having a formal main job I^f for an individual with characteristics X are then drawn according to

$$(\mathcal{W}, I^f) = \begin{cases} (\log(w_0^f) + \gamma^f \cdot \theta + u, 1) & \text{with prob. } 1 - G_\theta(\tilde{\kappa}(\theta)) \\ (\log(w_0^s) + \gamma^s \cdot \theta + u, 0) & \text{with prob. } G_\theta(\tilde{\kappa}(\theta)), \end{cases} \tag{29}$$

where the productivity type is given by $\theta = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma_\epsilon^2)$, $u \sim N(0, \sigma_u^2)$ is a measurement error, $\tilde{\kappa}(\theta)$ is the formality threshold above which the worker is formal, and $G_\theta(\kappa)$ is the cdf of fixed cost κ. We complement this specification with a Pareto tail for the top 1% of wages. We estimate the parameters by maximum likelihood.

[30] Suppose that, in the data, hourly wages (from the two sectors jointly) have cumulative density F_w and share $x(w)$ of workers with hourly wage w is informal. Then set $w^f(\theta) = w^s(\theta) = \theta$, $F(\theta) = F_w(\theta)$, and $G_\theta(\tilde{\kappa}(\theta)) = x(\theta)$ to match the observed distributions exactly.

[31] This assumption was made by Magnac (1991) in the study of the Colombian informal sector, although he does not restrict the productivity type to be one dimensional.

[32] Vector X contains typical regressors from Mincerian wage equations such as age, gender, education level, and experience. Following Pratap and Quintin (2006), who emphasize the importance of the establishment size to explain the differences of average wages across the formal and the informal sectors, we also include job and firm characteristics such as the task performed by the worker and the size of the firm. For the full list of regressors, see Appendix D.

[33] We impose that β and ϵ_i are identical in the two equations to remain consistent with the assumption that the underlying productivity type θ is one dimensional.

Our estimation approach is closely related to the more general estimation of Roy models (Heckman and Honore (1990), French and Taber (2011)). In these models, the distribution of sectoral productivities is two dimensional. In contrast, in our model the two dimensions of heterogeneity are given by productivity type θ and fixed cost of informal employment κ, where, importantly, the two can be correlated. As opposed to the identification approach proposed for Roy models, where observables are used to obtain exclusion restrictions, here we use the observable characteristics of workers to pin down the distribution of the productivity type θ. Consequently, the unconditional distribution of θ depends on the distribution of characteristics in the population.

As we discussed above, moonlighting cannot be recovered from the survey directly. We do not impose, however, that workers with a formal main job are exclusively formal. Instead, we treat the moonlighting margin as an unobservable in the estimation of the model. The estimated model will then imply moonlighting behavior which is consistent with the observed data on hourly wages and sector of the main job.

Estimation results. The left panel of Figure 4 presents the estimated productivity profiles and the density of productivity types. The bottom 25% of workers are more productive in the shadow sector, while the median worker is 6% more productive formally. We find that the comparative advantage in shadow labor decreases with the productivity type.[34] Thus, as assumed in the theoretical analysis, the single-crossing condition holds. The right panel of Figure 4 compares the fraction of workers with a formal main job by quintiles of hourly wages in the data and in the estimated model. The empirical share of workers with a formal main job increases sharply with hourly wage. The model tracks the data well, showing that our parametric specification is compatible with the observed sorting of workers across sectors. The model also predicts that no workers are moonlighting. That is intuitively consistent with our theoretical findings linking tax progressivity and moonlighting (recall Proposition 1), since the empirical tax and transfer

(a) Productivity profiles and type distribution (b) Formality and hourly wages

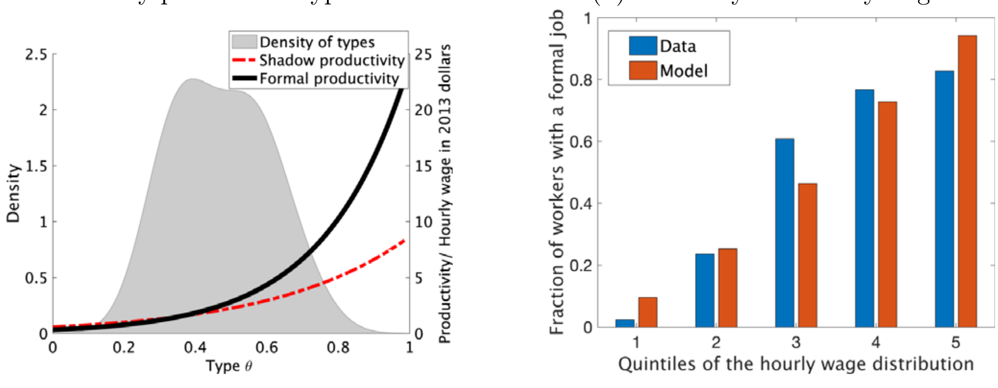

FIGURE 4. Estimation results. The productivity type distribution in panel (a) is obtained as a kernel density estimate of the distribution of $X\beta$ in the sample.

[34]Given our assumptions, $w^s(\theta)/w^f(\theta)$ is strictly decreasing when $\gamma^s - \rho^f < 0$. The point estimate of $\gamma^s - \gamma^f$ is -1.74 with a standard error of 0.08.

schedule in Colombia is only weakly progressive.[35] In the quantitative analysis, we show that moonlighting emerges and becomes important for the optimal policy once the tax progressivity is more pronounced.

In our sample of focus and in our model, 44% of agents work informally. To understand how the model matches the empirical extent of informality, consider two counterfactual exercises. First, let us set the fixed cost of informal employment κ to zero for all agents. In this case, the share of informal workers increases to 65%. The fixed cost effectively prevents many middle class workers from joining the informal sector. Second, we keep the fixed cost distribution as estimated but remove all tax distortions by replacing the empirical tax schedule with a lump-sum tax. Our model then implies that the share of informal workers falls to 25%. Thus, tax distortions explain a bit less than half of the size of the Colombian shadow economy. The remaining part is explained by the fact that workers from the bottom productivity quartile are more productive informally and face low fixed cost of informal employment.[36]

4.2 *Optimal tax schedule and the role of the informal sector*

In this subsection, we derive the optimal tax schedules for Colombia. We then compare them to benchmark tax schedules obtained when various informality responses are ignored due to misspecified beliefs of the planner. We consider two cases of misspecified beliefs, as in Proposition 2. In the first case, the planner ignores moonlighting but acknowledges the mobility of workers between sectors. In the second case, the planner ignores all informality responses: both the moonlighting and the mobility between the two sectors. The latter case can be interpreted as a belief in an extreme version of the segmented market hypothesis, where the allocation of workers between the sectors of work is immutable. As we explained before Proposition 2, the tax schedules chosen under these two cases of misspecified beliefs coincide with well known formulas from the optimal tax literature.[37] Importantly, we allow the income distribution to endogenously adjust to the chosen tax schedule.

The tax schedules we present for the two benchmark cases follow the notion of self-confirming policy equilibrium (SCPE), developed by Rothschild and Scheuer (2016). Since the distribution of income is endogenous to tax policy, we find the tax schedules implied by each formula iteratively: a tax schedule implies an income distribution that, together with a tax formula, results in a new tax schedule. A SCPE is a fixed point of this mapping. In such an equilibrium, the income distribution and the tax schedule are consistent with the beliefs of the planner. The planner has no incentives to adjust the policy and does not discover its misperceptions, which in our case correspond to being

[35]The Colombian tax and transfer schedule is regressive at low income levels (due to phasing out of benefits) and the progressivity at higher levels is limited, with the marginal tax rates below 38% for all but the highest earners.

[36]That is broadly consistent with evidence from Pratap and Quintin (2006), who show that low productivity Argentinian workers earn a higher hourly wage in the informal sector.

[37]Ignoring only moonlighting responses means that the planner follows the tax formula from Saez (2002) and Jacquet, Lehmann, and Van der Linden (2013), whereas ignoring all informality responses implies the formula from Diamond (1998) and Saez (2001).

unaware of various informality responses. In principle, each tax formula can admit multiple SCPE. We report the equilibrium that yields the highest welfare. Each tax schedule is required to generate the same revenue as the actual Colombian income tax.

We assume that Pareto weights follow $\lambda(\theta) = r(1 - F(\theta))^{r-1}$ as in Rothschild and Scheuer (2013). The parameter $r \geq 1$ captures the strength of redistributive preferences and is equal to the Pareto weight placed on the least productive agents. The average weight is always equal to 1 and the weight of the most productive agents converges to 0 when $r > 1$. We consider two cases of social preferences, $r = 1.1$ and $r = 1.7$, which we call weakly and strongly redistributive, respectively. The Pareto weight placed on the 90th percentile of θ is approximately 0.9 for weakly redistributive and 0.3 for strongly redistributive social preferences.

Figure 5 depicts the optimal tax schedules and the tax schedules chosen when either moonlighting responses or all informality responses are ignored, and Table 1 shows the implied distributions of workers between the sectors. Remarkably, the order of tax rates predicted by Proposition 2 continues to hold, even though the assumption of identical income distributions is clearly not satisfied. We find that ignoring all informality responses leads to higher tax rates at each income level than ignoring only moonlighting responses, while the optimal tax rates are the lowest.

The optimal tax schedules are close to fully progressive: the marginal tax rates almost always increase with income. At low income levels, tax rates are low and roughly constant; they start to rise close to the median income (approx. \$10,000) and stabilize at the top. A stronger taste for redistribution shifts the schedule upward while roughly preserving this shape. Thus, the optimal fraction of workers with exclusively formal employment decreases from 71% to 57% with the strength of redistributive preferences. The bulk of the remaining workers are employed exclusively informally. The share of moonlighting individuals is small and increases from 1.4% to 3.4% as redistributive preferences become stronger.

When all informality responses are ignored, the tax schedules feature very high marginal tax rates at low income levels, approaching 100% at the bottom. The tax rates are decreasing through most of the income distribution and increase again as they approach the top income tail, generating a familiar U shape (Diamond (1998), Saez (2001)).

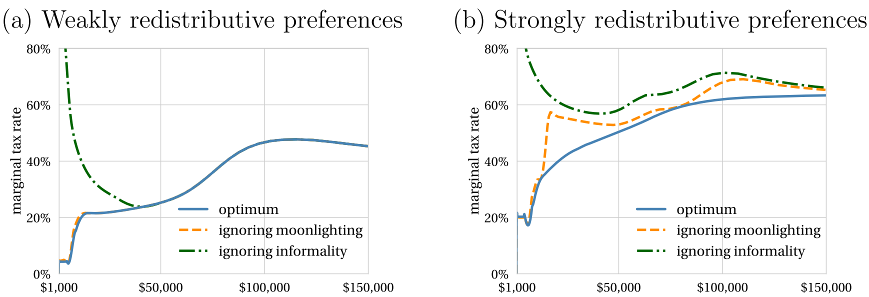

Figure 5. Equilibrium tax schedules. In the optimum with weakly (strongly) redistributive social preferences the 50th, 95th, and 99th percentiles of formal income are approximately \$10,500 (\$9400), \$45,000 (\$40,000), and \$87,000 (\$78,000), respectively.

Table 1. Allocation of workers between sectors and welfare loss.

	Share of Workers by Sector of Work			Welfare Loss
	Only Formal	Moonlighting	Only Informal	(Rel. to Optimum, % of Cons.)
Weakly redistributive preferences				
Optimum	70.8%	1.4%	27.8%	
Ignoring moonlighting	70.1%	2.1%	27.8%	0%
Ignoring informality	32.6%	0%	67.4%	13.5%
Strongly redistributive preferences				
Optimum	57.4%	3.4%	39.2%	
Ignoring moonlighting	52.6%	7.6%	39.9%	2.4%
Ignoring informality	23.6%	0%	76.4%	24.8%

High tax rates push most of the low and medium productivity workers to the shadow sector. Nevertheless, from the planner's perspective, the tax schedule seems optimal. That is because the implied density of formal income at low and medium income levels—and, hence, the perceived deadweight loss from taxation—is, in fact, low. We find that ignoring all informality responses when setting the tax policy effectively doubles the share of shadow workers relative to the optimum. Although taxes rates are on average higher than in the optimum, tax progressivity is actually lower, since tax rates increase the most at low earnings. Consistent with our theoretical findings linking progressivity and moonlighting, the share of moonlighting workers falls relative to the optimum: virtually all shadow workers are exclusively informal. The welfare loss from ignoring all informality responses is catastrophic and ranges from 13.5% to 24.8% of consumption depending on the social welfare function, as reported in Table 1. In other words, accounting for informality brings a huge welfare gain.

When we compare the optimal tax schedule with the tax schedule when only the moonlighting responses are ignored, it is clear that the impact of moonlighting depends crucially on redistributive preferences. When preferences for redistribution are strong, the moonlighting responses reduce the marginal tax rates above the median formal income by up to 20 percentage points (p.p.) The moonlighting responses—complementing formal income with additional informal earnings—are important higher in the income distribution compared to the responses of switching from entirely formal to entirely informal employment. Intuitively, a secondary informal job is tempting for workers with well paid formal jobs who face high marginal tax rates and for whom transitioning to entirely informal employment is too costly. On the other hand, when preferences for redistribution are weak, the moonlighting responses have little effect on the optimal tax schedule. This is because the tax rates for highly productive workers are not substantial enough to create incentives for informality.

When the preferences for redistribution are strong, ignoring moonlighting results in a share of moonlighting workers that more than doubles the optimal value, with a large welfare loss equivalent to a 2.4% drop in consumption. Since the tax schedule is excessively progressive, with the tax rates too high above median formal income but approximately optimal below, we should expect moonlighting to become more prevalent.

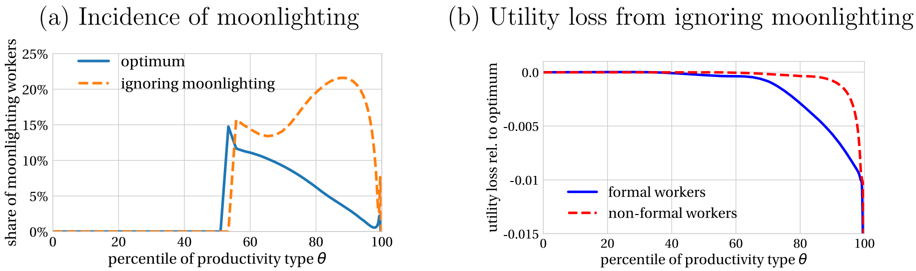

Figure 6. Consequences of ignoring moonlighting.

However, why is increased moonlighting so damaging for social welfare? We find that the sorting of workers across sectors is substantially different in comparison to the optimal allocation (see Figure 6, panel (a)). Relative to the optimum, moonlighting is induced among workers with higher productivity, mostly from the top quartile of the productivity distribution. The median percentile of productivity type of moonlighting workers increases from 66% at the optimum to 78%. As the most productive workers who face high marginal tax rates replace a fraction of their formal earnings with shadow earnings, the tax revenue is eroded heavily. In fact, although the overall level of taxes is substantially higher at high income levels when moonlighting is ignored relative to the optimum (e.g., the average tax rate at the 95th percentile of the formal earnings is higher by 8 p.p.), the overall tax revenue is actually slightly lower. This means that the least productive workers receive a lower transfer. Since the tax schedule chosen when the moonlighting responses are ignored generates a lower tax revenue while imposing higher distortions, it is Pareto inefficient. Indeed, all agents in the economy loose relative to the optimum, although losses are concentrated among the most productive workers (see Figure 6, panel (b)).

5. Conclusions

This paper studies the optimal income taxation when agents can earn income in a shadow economy that is unobserved by the government. The key technical contribution is allowing for workers to supply labor simultaneously to the formal and the informal sectors, which we call moonlighting. We show theoretically and quantitatively that the optimal tax schedule that accounts for informality responses features lower tax rates throughout the income distribution. In particular, the possibility of workers migrating to entirely informal employment restricts tax rates at low and medium income levels, while the possibility of moonlighting is relevant at higher levels of income.

Theoretical tools we developed could be used in other settings. Our tax formula applies when agents can simultaneously work in two, broadly understood, sectors and the tax schedule can be optimized over the income from only one of the sectors. Examples of such settings include the model of home production or the problem of a local tax authority with residents who can work partly outside its jurisdiction.

Appendix A: Proofs from Section 2 and Section 3

Proof of Lemma 1. The strict *Spence–Mirrlees* single-crossing condition holds if, keeping the formal income level fixed, the marginal rate of substitution $\frac{v'(n)}{w^f(\theta)}$ is strictly decreasing with θ, where n is the total labor supply. For formal workers, it follows from the strict convexity of v. For workers who supply labor to the informal sector, we have $v'(n) = w^s(\theta)$ and the single-crossing follows from $\frac{w^s(\theta)}{w^f(\theta)}$ being strictly decreasing. Given the single-crossing condition, the formal earnings schedule is increasing by Theorem 7.2 in Fudenberg and Tirole (1991). □

Proof of Proposition 1. The second-order condition of the moonlighting θ worker is $-T''(\underline{y}^f(\theta)) \le 0$. It cannot be satisfied if $T''(\underline{y}^f(\theta)) < 0$. If the second-order condition holds as equality, then the marginal tax rate is locally constant and the agent is indifferent between $\underline{y}^f(\theta)$ and $\underline{y}^f(\theta) + e$ for sufficiently small $e > 0$. By Assumption 2, the indifferent agent chooses higher formal earnings. Thus, the moonlighting workers never choose formal earnings where the second-order condition holds as equality. □

Lemma 4. *The following statements hold. (i) The second-order condition with respect to formal earnings holds as a strict inequality for all workers. (ii) All formal workers have a unique utility-maximizing formal earnings level conditional on remaining formal. (iii) All workers with type $\theta > \underline{\theta}$, where $\underline{y}^f(\theta)$ is continuous, have a unique utility-maximizing formal earnings level conditional on earning some informal income.*

Proof. (i) The proof of Proposition 1 shows that for moonlighting workers, it is straightforward to extend the argument to formal workers. (ii) Given Assumption 3, we need to check it only for $\theta > \underline{\theta}$. Suppose, on the contrary, that a formal θ worker is indifferent between $\overline{y}^f(\theta)$ and strictly lower earnings y'. By continuity, there is a formal worker with productivity type $\theta' < \theta$ such that $\overline{y}^f(\theta) > \overline{y}^f(\theta') > y'$. Thus, there exists a possible equilibrium in which the θ worker chooses earnings y' and the earnings choices are not increasing in θ, which contradicts Lemma 1. Note that such equilibrium is ruled out by Assumption 2, but Lemma 1 does not depend on this assumption. (iii) The proof is analogous to that of statement (2) for the formal workers. □

Proof of Lemma 2. We will prove the simple case first and then generalize. Suppose that the original tax schedule is strictly progressive ($T''(y) > 0$) apart from the regressivity interval $[y_1, y_2]$. Consequently, the formal earnings schedule of moonlighting workers $\underline{y}^f(\cdot)$ is discontinuous at some θ_d, with $\underline{y}^f(\theta_d^-) < y_1$ and $\underline{y}^f(\theta_d) > y_2$. Furthermore, assume that the moonlighting $\underline{\theta}$ worker has a unique utility-maximizing choice of formal earnings. Fix the direction of the reform $\mathrm{d}T \in \mathcal{C}^2$ and denote by μ the size of the reform.

First we will show that $\overline{y}^f(\theta)$, $\underline{y}^f(\theta)$, and θ_d are continuously differentiable (denoted c.d.) with μ at $\mu = 0$ for almost all θ. Second, we show that almost everywhere $\tilde{\kappa}(\theta)$ is either c.d. with μ or it does not affects tax revenue to the first order. Then we will show that it implies the existence of $\mathrm{d}\mathcal{TR}$.

The first-order condition for $\overline{y}^f(\theta)$ at the reformed tax schedule $T + \mu \cdot \mathrm{d}T$ is

$$H_1(\overline{y}^f(\theta), \mu) \equiv \bigl(1 - T'(\overline{y}^f(\theta)) - \mu \cdot \mathrm{d}T'(\overline{y}^f(\theta))\bigr) \cdot w^f(\theta) - v'\left(\frac{\overline{y}^f(\theta)}{w^f(\theta)}\right) = 0. \tag{30}$$

By Lemma 4, the utility-maximizing earnings level is unique, which means that this equation is sufficient to pin down the equilibrium earnings level following a reform for μ small enough. Furthermore, the second-order condition holds as a strict inequality: $\frac{\partial H_1(\overline{y}^f(\theta), \mu)}{\partial \overline{y}^f(\theta)} < 0$. We can apply the implicit function theorem (IFT). which implies that $\frac{\partial y^f(\theta)}{\partial \mu}|_{\mu=0} = \mathrm{d}y^f(\theta)[T, \mathrm{d}T]$ exists and is continuous.[38] In an analogous way we can show that $\underline{y}^f(\theta)$ is c.d. with μ in the intervals $(\underline{\theta}, \theta_d)$ and $(\theta_d, \overline{\theta})$.

Now we will show that θ_d is c.d. with μ. Denote auxiliary formal income intervals $I_1 = [0, \underline{y}^f(\theta_d^-) + \epsilon]$ and $I_2 = [\underline{y}^f(\theta_d) - \epsilon, \infty)$ for small $\epsilon > 0$. Define auxiliary earnings schedules

$$\underline{y}_j^f(\theta, \mu) \equiv \arg\max_{y^f \in I_j} \left\{ \max_{y^s \geq 0} R(y^f) - \mu \cdot \mathrm{d}T(y^f) + y^s - v\left(\frac{y^f}{w^f(\theta)} + \frac{y^s}{w^s(\theta)}\right)\right\} \tag{31}$$

for all θ and $j \in \{1, 2\}$. In words, $\underline{y}_i^f(\theta, \mu)$ is the utility-maximizing formal earnings choice of a moonlighting θ worker following a tax reform of size μ when the choice is confined to I_j. Note that $\underline{y}_i^f(\theta, \mu)$ is c.d. with θ and μ. We can now write the indifference condition (8) as

$$\begin{aligned} H_2(\theta_d, \mu) \equiv\ & T\bigl(\underline{y}_2^f(\theta_d, \mu)\bigr) - T\bigl(\underline{y}_1^f(\theta_d, \mu)\bigr) + \mu \cdot \bigl(\mathrm{d}T\bigl(\underline{y}_2^f(\theta_d, \mu)\bigr) - \mathrm{d}T\bigl(\underline{y}_1^f(\theta_d, \mu)\bigr)\bigr) \\ & - \bigl(T'\bigl(\underline{y}_1^f(\theta_d, \mu)\bigr) + \mu \cdot \mathrm{d}T'\bigl(\underline{y}_1^f(\theta_d, \mu)\bigr)\bigr) \cdot \bigl(\underline{y}_2^f(\theta_d, \mu) - \underline{y}_1^f(\theta_d, \mu)\bigr) \\ =\ & 0. \end{aligned} \tag{32}$$

Notice that $\frac{\partial H_2}{\partial \theta_d}|_{\mu=0} = -T''(\underline{y}_1^f(\theta_d, 0)) \cdot \frac{d\underline{y}_1^f(\theta_d, 0)}{d\theta_d} \cdot (\underline{y}_2^f(\theta_d, 0) - \underline{y}_1^f(\theta_d, 0))$ is negative: $T'' > 0$ holds for moonlighting workers and the continuity of T' implies that $\frac{\partial \underline{y}_1^f(\theta_d, 0)}{\partial \theta_d} > 0$. Then, by IFT, θ_d is c.d. with μ.

Regarding threshold $\tilde{\kappa}(\theta)$, it is straightforward to show that it is c.d. with μ when $\tilde{\kappa}(\theta) > 0$ by applying IFT to (2). However, $\tilde{\kappa}(\theta)$ can be non-differentiable with μ when $\tilde{\kappa}(\theta) = 0$, as it cannot decrease below 0. It would be problematic if, for a positive measure of productivity types, we had $\overline{y}^f(\theta) = \underline{y}^f(\theta) > \underline{y}^f(\theta^-)$ and $\tilde{\kappa}(\theta) = 0$ (in words, the $(\theta, 0)$ worker is indifferent between two formal income levels, the lower of which involves informal activity), since then a movement of these workers to informality would have a first-order impact on tax revenue.[39] Below we show that in such cases, the expected utility $V(\theta, 0)$ is non-differentiable with respect to the productivity type. Since

[38]See, for instance, de Oliveira (2014, Theorem 5).

[39]One can construct an economy in which a positive measure of productivity types has $\tilde{\kappa}(\theta) = 0$ and is indifferent between staying formal or joining the shadow economy, but conditional on joining the shadow economy they would choose formal incomes in the neighborhood of their original earnings, implying that the overall impact on tax revenue is zero to the first order. For example, suppose that $1 - T'(\overline{y}^f(\theta)) = \frac{w^s(\theta)}{w^f(\theta)}$

$V(\cdot,0)$ has a measure zero of non-differentiability points by Milgrom and Segal (2002) (Corollary 1 and footnote 10), there is a measure zero of such cases and they do not affect the behavior of $\mathcal{TR}$. To show that $V(\cdot,0)$ is not differentiable at θ, denote $\overline{n}^f=\frac{\overline{y}^f(\theta)}{w^f(\theta)}$, $\underline{n}^f=\frac{\underline{y}^f(\theta^-)}{w^f(\theta)}$, and $n^m=v'^{-1}(w^s(\theta))$. The directional derivatives of $V(\theta,0)$ are $V_\theta(\theta^-,0)=(\frac{\dot{w}^f(\theta)}{w^f\theta}\underline{n}^f+\frac{\dot{w}^s(\theta)}{w^s\theta}(n^m-\underline{n}^f))v'(n^m)$ and $V_\theta(\theta^+,0)=\frac{\dot{w}^f(\theta)}{w^f\theta}\overline{n}^f v'(\overline{n}^f)$, where $\dot{w}^i\equiv\frac{dw^i(\theta)}{d\theta}$. We have $V_\theta(\theta^+,0)>V_\theta(\theta^-,0)$ when

$$\frac{\dot{w}^f(\theta)}{w^f(\theta)}\big(\overline{n}^f v'(\overline{n}^f)-\underline{n}^f v'(n^m)\big)>\frac{\dot{w}^s(\theta)}{w^s(\theta)}\big(n^m v'(n^m)-\underline{n}^f v'(n^m)\big). \tag{33}$$

By Assumption 1, $\frac{\dot{w}^f(\theta)}{w^f(\theta)}>\frac{\dot{w}^s(\theta)}{w^s(\theta)}$. We also have $\overline{n}^f\ge n^m$, since otherwise $w^s(\theta)>(1-T'(\overline{y}^f(\theta))w^f(\theta)$ and type $(\theta,0)$ would be strictly better off moonlighting than staying formal, which contradicts $\tilde{\kappa}(\theta)=0$. Thus, the above inequality holds.

Now we will apply the Leibniz integral rule.[40] Write the tax revenue as

$$\begin{aligned}\mathcal{TR}=&\int_{\underline{\theta}}^{\overline{\theta}}T\big(\overline{y}^f(\theta)\big)\cdot\big(1-G_\theta\big(\tilde{\kappa}(\theta)\big)\big)\,dF(\theta)\\&+\int_{\underline{\theta}}^{\theta_d}T\big(\underline{y}_1^f(\theta)\big)\cdot G_\theta\big(\tilde{\kappa}(\theta)\big)\,dF(\theta)+\int_{\theta_d}^{\overline{\theta}}T\big(\underline{y}_2^f(\theta)\big)\cdot G_\theta\big(\tilde{\kappa}(\theta)\big)\,dF(\theta).\end{aligned} \tag{34}$$

For each integral on the right-hand side, since the integrand as well as the edges of the integrating region are either constant or c.d. with μ at $\mu=0$, the derivative of each integral with μ at $\mu=0$ exists. Thus, the derivative of the left-hand side exists and we denote it by $\mathrm{d}\mathcal{TR}[T,\mathrm{d}T]$.

Below we describe how to generalize this result to more complex cases.

Case 1. Multiple discontinuities. Since $\underline{y}^f$ is increasing, it has countably many discontinuity points $\{\theta_{d1},\theta_{d2},\dots\}$. For each discontinuity point, we need to keep track of two auxiliary intervals: those containing the highest and the second-highest utility-maximizing formal earnings level. Define countably many income intervals $I_1,I_2,\dots,I_i,\dots$ in the manner $I_1=[0,\underline{y}^f(\theta_{d1})+\epsilon]$, $I_2=[\underline{y}^f(\theta_{d1})-\epsilon,\underline{y}^f(\theta_{d2})+\epsilon]$, $I_i=[\underline{y}^f(\theta_{di-1})-\epsilon,\underline{y}^f(\theta_{di})+\epsilon]$, and so on for small $\epsilon>0$. For each I_i, define the auxiliary earnings schedule $\underline{y}_i^f(\theta,\mu)$ as in (31). Following the steps from the simple case, it is easy to show that each θ_{di} is c.d. with μ at $\mu=0$. Thus, we can write $\mathcal{TR}$ as a sum of countably many integrals, each with the integrand and the edges of the integration region that are c.d. with μ at $\mu=0$. By the Leibniz integral rule, $\mathrm{d}\mathcal{TR}$ exists.

Case 2. Non-unique utility-maximizing choice for moonlighting $\underline{\theta}$ workers. Treat $\underline{\theta}$ as a discontinuity point and apply the case of multiple discontinuities above. □

for all θ, such that for all types we have $\tilde{\kappa}(\theta)=0$ and $\overline{y}^f(\theta)=\underline{y}^f(\theta)=\underline{y}^f(\theta^-)$. Increasing marginal tax rates uniformly increases informality, but this increase of informality has no first-order impact on tax revenue.

[40]See Theorem 2.4.1 in Casella and Berger (2002).

Proof of Theorem 1. Integrating by parts, the welfare effect from (12) becomes

$$\mathrm{WE} = -\left[\mathrm{d}T(y)\int_0^y \overline{\lambda}(z)h(z)\,dz\right]_{y=0}^{y\to\infty} + \int_0^\infty \mathrm{d}T'(y)\int_0^y \overline{\lambda}(z)h(z)\,dz\,dy. \tag{35}$$

Recall that the average Pareto weight $\int_0^\infty \overline{\lambda}(z)h(z)\,dz$ is 1. Then the first term is $-\lim_{y\to\infty}\mathrm{d}T(y) = -\int_0^\infty \mathrm{d}T'(y)\,dy$. Rearranging, we arrive at

$$\mathrm{WE} = -\int_0^\infty \left[\int_y^\infty \overline{\lambda}(z)h(z)\,dz\right]\mathrm{d}T'(y)\,dy. \tag{36}$$

Analogously, we can express mechanical effect ME as $\int_0^\infty [(1-H(y)]\,\mathrm{d}T'(y)\,dy$.

Using $\overline{s}(y)$, defined explicitly in the statement of the theorem, the behavioral effect due to intensive responses $\mathrm{BE}_{\mathrm{int}}$ can be written as

$$-\int_0^\infty \left[\frac{T'(y)}{1-T'(y)}\tilde{\varepsilon}^f(y)yh^f(y) + \frac{T'(\overline{s}(y))}{1+T'(\overline{s}(y))}\tilde{\varepsilon}^s(\overline{s}(y))\overline{s}(y)h^s(\overline{s}(y))\right]\mathrm{d}T'(y)\,dy,$$

where $\overline{s}(y)$ stands for the formal income level at which the moonlighting workers respond on the intensive margin to the reform: $\overline{s}(y) = y$ if the earning schedule of moonlighting workers is locally continuous and $\overline{s}(y) > y$ otherwise. If there are no moonlighting workers with earnings $\geq y$, then there are no intensive margin responses of moonlighting workers, $\overline{s}(y) = 0$, and the second term on the right-hand side disappears.

Write the behavioral effect due to extensive responses (20) as

$$\mathrm{BE}_{\mathrm{ext}} = -\int_0^\infty \Delta T(z)\pi(z)h^f(z)\int_0^\infty \mathbb{1}(\rho(z)\leq y\leq z)\,\mathrm{d}T'(y)\,dy\,dz \tag{37}$$

$$= -\int_0^\infty \left[\int_y^\infty \Delta T(z)\pi(z)h^f(z)\mathbb{1}(\rho(z)\leq y)\,dz\right]\mathrm{d}T'(y)\,dy, \tag{38}$$

where, in the second row, we exchanged the order of integration and expressed the second inequality within the identity function as a limit of the integration interval.

Plugging the terms into optimality condition (21) leads to

$$\begin{aligned}\int_0^\infty \Bigg[&-\frac{T'(y)}{1-T'(y)}\tilde{\varepsilon}^f(y)yh^f(y) - \frac{T'(\overline{s}(y))}{1+T'(\overline{s}(y))}\tilde{\varepsilon}^s(\overline{s}(y))\overline{s}(y)h^s(\overline{s}(y))\\ &+\int_y^\infty [1-\overline{\lambda}(z)]\,dH(z) + \int_y^\infty \Delta T(z)\pi(z)\mathbb{1}(\rho(z)\leq y)\,dH^f(z)\Bigg]\mathrm{d}T'(y)\,dy = 0,\end{aligned} \tag{39}$$

which holds for an arbitrary reform if and only if formula (22) holds for all y. □

Proof of Proposition 2. Note that we treat the entire term $\Delta T(z)\cdot\pi(z)$, which stands for the elasticity of $h^f(z)$ with respect to $\Delta T(z)$, as fixed. Consider raising the tax rate at y. Formula I (equation 23) does not account for the intensive margin responses of moonlighting workers when $\overline{s}(y) > y$ as well as for extensive margin responses at formal earnings z where $0 < \rho(z) \leq y$. Formula II (equation 24) in addition does not account for extensive responses when $\rho(z) = 0$. Since the status quo tax schedule has nonnegative marginal tax rates, both intensive and extensive margin responses have a nonpositive impact on the tax revenue. Thus, formulas I and II underestimate the cost of raising the tax rate relative to the actual cost, and formula II underestimates it further relative to formula I. This implies the ranking of tax rates. □

LEMMA 5. *Suppose that marginal tax rates are nondecreasing, the tax rate in the top bracket* $[z^*, \infty)$ *is* $\tau \geq 1 - \phi$, $\underline{y}^f$ *is bounded from above by* $\tilde{z}$, *and Frisch elasticity in the top bracket is* ε. *Consider* θ^* *such that* $\overline{y}^f(\theta^*) \geq z^*$. *Then* $\psi(\theta^*) \geq \frac{\tilde{\kappa}(\theta)}{w^f(\theta)^{1+\varepsilon}} \geq \psi_{lb}(\theta^*)$ *for all* $\theta \geq \theta^*$, *where* $\psi(\theta^*) = \frac{1}{1+\varepsilon}((\frac{w^s(\theta^*)}{w^f(\theta^*)})^{1+\varepsilon} - (1-\tau)^{1+\varepsilon})$ *and*

$$\psi_{lb}(\theta^*) = \psi(\overline{\theta}) - \left(\frac{w^s(\theta^*)}{w^f(\theta^*)} - (1-\tau)\right) \cdot \frac{\tilde{z}}{w^f(\theta^*)^{1+\varepsilon}}$$
$$+ (1-\tau)^{1+\varepsilon} - \frac{R(\overline{y}^f(\theta^*)) - R(\underline{y}^f(\theta^*))}{\overline{y}^f(\theta^*) - \underline{y}^f(\theta^*)} \cdot (1-\tau)^{\varepsilon}. \tag{40}$$

PROOF OF LEMMA 5. Without loss of generality, $v(n) = (1 + \frac{1}{\varepsilon})^{-1} \cdot n^{1+\frac{1}{\varepsilon}}$. One can show that threshold $\tilde{\kappa}(\theta) = V(\theta, 0) - \lim_{\kappa\to\infty} V(\theta, \kappa)$ is

$$\tilde{\kappa}(\theta) = w^f(\theta)^{1+\varepsilon} \cdot \psi(\theta) - \left(\frac{w^s(\theta)}{w^f(\theta)} - \frac{R(\overline{y}^f(\theta)) - R(\underline{y}^f(\theta))}{\overline{y}^f(\theta) - \underline{y}^f(\theta)}\right) \cdot \underline{y}^f(\theta) \tag{41}$$
$$+ \big((1-\tau) \cdot w^f(\theta)\big)^{1+\varepsilon} - \frac{R(\overline{y}^f(\theta)) - R(\underline{y}^f(\theta))}{\overline{y}^f(\theta) - \underline{y}^f(\theta)} \cdot \overline{y}^f(\theta), \tag{42}$$

where $\psi(\theta)$ is defined in the lemma. The following statements are true: (i) $\psi(\theta)$ is strictly decreasing in θ, (ii) $\frac{R(\overline{y}^f(\theta))-R(\underline{y}^f(\theta))}{\overline{y}^f(\theta)-\underline{y}^f(\theta)}$ is decreasing in θ because of nondecreasing marginal tax rates and it converges to $1-\tau$ as $\theta \to \overline{\theta}$, and (iii) $\frac{w^s(\theta)}{w^f(\theta)} - \frac{R(\overline{y}^f(\theta))-R(\underline{y}^f(\theta))}{\overline{y}^f(\theta)-\underline{y}^f(\theta)} > 0$, since otherwise a moonlighting θ worker would be better off remaining fully formal. Note that formal θ workers supply less labor than the non-formal ones. It follows that the bounds from the lemma hold. □

PROOF OF LEMMA 3. Define the local Pareto parameter at earnings y as $a(y) = \frac{h(y)\cdot y}{1-H(y)}$. When $1 - \tau > \phi$, workers with sufficiently high θ are formal and we have[41]

$$\lim_{y\to\infty} a(y) = \underbrace{\lim_{\theta\to\overline{\theta}} \frac{f(\theta)\cdot w^f(\theta)}{1-F(\theta)} \cdot \left(\frac{dw^f(\theta)}{d\theta}\right)^{-1}}_{=\alpha} \cdot \underbrace{\left(\frac{d\overline{y}^f(\theta)}{dw^f(\theta)} \cdot \frac{w^f(\theta)}{\overline{y}^f(\theta)}\right)^{-1}}_{=(1+\varepsilon)^{-1} \text{ by } (4)} = \frac{\alpha}{1+\varepsilon}. \tag{43}$$

[41]Formal productivity being Pareto distributed implies that $1 - F(\theta) = k \cdot w^f(\theta)^{-\alpha}$ and $f(\theta) \cdot (\frac{\mathrm{d}w^f(\theta)}{\mathrm{d}\theta})^{-1} = \alpha \cdot k \cdot w^f(\theta)^{-(\alpha+1)}$.

When $1-\tau<\phi$, we can use the bounds on $\frac{\tilde{\kappa}(\theta)}{w^f(\theta)^{1+\varepsilon}}$ from Lemma 5 to bound the local Pareto parameter:

$$\frac{\psi_{lb}(\theta)^{-\gamma}}{\psi(\theta)^{-\gamma}}\frac{w^f(\theta)^{-\gamma(1+\varepsilon)}f(\theta)}{\displaystyle\int_\theta^{\overline{\theta}} w^f(t)^{-\gamma(1+\varepsilon)}\,dF(t)}\frac{\overline{y}^f(\theta)}{\frac{d\overline{y}^f(\theta)}{d\theta}}$$

$$\geq a\big(\overline{y}^f(\theta)\big)\geq\frac{\psi(\theta)^{-\gamma}}{\psi_{lb}(\theta)^{-\gamma}}\frac{w^f(\theta)^{-\gamma(1+\varepsilon)}f(\theta)}{\displaystyle\int_\theta^{\overline{\theta}} w^f(t)^{-\gamma(1+\varepsilon)}\,dF(t)}\frac{\overline{y}^f(\theta)}{\frac{d\overline{y}^f(\theta)}{d\theta}}.$$

Since $\psi_{lb}(\theta)\to\psi(\overline{\theta})$ as $\theta\to\overline{\theta}$ and formal wages are Pareto distributed, both sides converge to $\frac{\alpha}{1+\varepsilon}+\gamma$ as $\theta\to\overline{\theta}$. □

Proof of Proposition 3. For derivations of τ_M, see, e.g., Piketty and Saez (2013). Regarding τ^*, if $\tau_M<1-\phi$, then all top productivity types have a strict preference for formality as $z^*\to\infty$. Thus, the Mirrleesian tax formula still applies in the model with a shadow economy for sufficiently high z^* and $\lim_{z^*\to\infty}\tau^*=\tau_M$.

Suppose that $\tau\geq 1-\phi$ and consider z^* high enough such that $\tau>1-\frac{w^s(\theta^*)}{w^f(\theta^*)}$, where θ^* satisfies $\overline{y}^f(\theta^*)=z^*$. Consider increasing the top tax rate marginally. The combined fiscal and welfare gains of doing so are given by

$$\Phi\big(\tau,z^*\big)=1-\lambda-\frac{\tau}{1-\tau}\cdot a\cdot\varepsilon-\mathbb{E}\big[\pi(z)\cdot\Delta T(z)|z\geq z^*\big],\tag{44}$$

where $a=\frac{\overline{z}}{\overline{z}-z^*}$ for $\overline{z}$ equal to the average earnings in the top bracket (derivations are standard and omitted). Focussing on the the last term, we have

$$\pi\big(\overline{y}^f(\theta)\big)\cdot\Delta T\big(\overline{y}^f(\theta)\big)=\frac{g_\theta\big(\tilde{\kappa}(\theta)\big)\cdot\tilde{\kappa}(\theta)}{1-G_\theta\big(\tilde{\kappa}(\theta)\big)}\cdot\frac{w^f(\theta)^{1+\varepsilon}}{\tilde{\kappa}(\theta)}\cdot\frac{\Delta T\big(\overline{y}^f(\theta)\big)}{w^f(\theta)^{1+\varepsilon}}.\tag{45}$$

Let us bound this term from both sides for any $\theta\geq\theta^*$. The first factor on the right-hand side is γ when $\tau>1-\phi$; the second factor can be bounded by Lemma 5. Using the assumption of nondecreasing marginal tax rates, we can bound the third factor from above as

$$\frac{\Delta T\big(\overline{y}^f(\theta)\big)}{w^f(\theta)^{1+\varepsilon}}\leq\frac{\tau\cdot\overline{y}^f(\theta)}{w^f(\theta)^{1+\varepsilon}}=\tau\cdot(1-\tau)^\varepsilon\tag{46}$$

and from below as

$$\frac{\Delta T\big(\overline{y}^f(\theta)\big)}{w^f(\theta)^{1+\varepsilon}}=\frac{T\big(\overline{y}^f(\theta)\big)-T\big(\underline{y}^f(\theta)\big)}{\overline{y}^f(\theta)-\underline{y}^f(\theta)}\left((1-\tau)^\varepsilon-\frac{\underline{y}^f(\theta)}{w^f(\theta)^{1+\varepsilon}}\right)\tag{47}$$

$$\geq\frac{T\big(\overline{y}^f(\theta^*)\big)-T\big(\underline{y}^f(\theta^*)\big)}{\overline{y}^f(\theta^*)-\underline{y}^f(\theta^*)}\left((1-\tau)^\varepsilon-\frac{\tilde{z}}{w^f\big(\theta^*\big)^{1+\varepsilon}}\right).\tag{48}$$

The implied bounds on $\pi(\overline{y}^f(\theta)) \cdot \Delta T(\overline{y}^f(\theta))$ have a common limit as $\theta^* \to \overline{\theta}$, which is $\gamma \cdot \frac{(1+\varepsilon)\cdot\tau\cdot(1-\tau)^\varepsilon}{\phi^{1+\varepsilon}-(1-\tau)^{1+\varepsilon}} \equiv \xi(\tau)$. Thus, $\xi(\tau)$ is also the limit of $\mathbb{E}[\pi(z)\cdot\Delta T(z)|z \geq z^*]$ as $z^* \to \infty$. Note that the integral $\int_{1-\phi}^{\tau} \xi(t)\,dt$ is infinite for any $\tau \in (1-\phi, 1)$,

$$\begin{aligned}\int_{1-\phi}^{\tau} \xi(t)\,dt &\geq k_1 \int_{1-\phi}^{\tau} \frac{1}{\phi^{1+\varepsilon}-(1-t)^{1+\varepsilon}}\,dt \\ &= k_1 \int_0^{x(\tau)} \frac{1}{x}\left(\frac{dx}{dt}\right)^{-1} dx \geq k_2 \int_0^{x(\tau)} \frac{1}{x}\,dx = +\infty, \end{aligned} \tag{49}$$

where $x(t) = \phi^{1+\varepsilon} - (1-t)^{1+\varepsilon}$, $k_1 = \gamma \cdot (1+\varepsilon)\cdot \min_{t\in[1-\phi,\tau]}\{t\cdot(1-t)^\varepsilon\}$, and $k_2 = \frac{k_1}{(1+\varepsilon)\cdot\phi^\varepsilon}$. In the limit as $z^* \to \infty$, the tax revenue loss from crossing the tipping point $1-\phi$ always dominates any redistributive gains, which are finite. Thus, when $\tau_M \geq 1-\phi$, τ^* is equal to $1-\phi$ for sufficiently high z^* or converges to it from below. □

Appendix B: Detailed derivations of the impact of tax reforms

Consider an arbitrary tax reform in direction $\mathrm{d}T$ of size $\mu > 0$. Let us evaluate the first-order condition for formal workers (4) at the reformed tax schedule. Suppressing the productivity type θ for brevity, we obtain

$$(1 - \left((T'(\overline{y}^f) + \mu\cdot \mathrm{d}T'(\overline{y}^f) + T''(\overline{y}^f)\cdot\mu\cdot \mathrm{d}\overline{y}^f\right))w^f = v'\left(\frac{\overline{y}^f + \mu\cdot \mathrm{d}\overline{y}^f}{w^f}\right). \tag{50}$$

Subtract (4), divide by μ, and evaluate in the limit as $\mu \to 0$ to get

$$-\mathrm{d}T'(\overline{y}^f) - T''(\overline{y}^f)\cdot \mathrm{d}\overline{y}^f = v''\left(\frac{\overline{y}^f}{w^f}\right)\cdot\frac{\mathrm{d}\overline{y}^f}{(w^f)^2}. \tag{51}$$

Substituting $\mathrm{d}[1-T'(\overline{y}^f)]$ for $-\mathrm{d}T'(\overline{y}^f)$ and rearranging yields the earnings elasticity of formal workers

$$\tilde{\varepsilon}^f = \frac{\mathrm{d}\overline{y}^f}{\mathrm{d}[1-T'(\overline{y}^f)]}\frac{1-T'(\overline{y}^f)}{\overline{y}^f} = \left(\frac{1}{\varepsilon} + \frac{T''(\overline{y}^f)\cdot\overline{y}^f}{1-T'(\overline{y}^f)}\right)^{-1}, \tag{52}$$

where $\varepsilon = \frac{v'}{v''}\frac{w^f}{\overline{y}^f}$ is the Frisch elasticity or the elasticity of earnings along the linear tax schedule.

Following analogous steps with respect to the first-order condition of the moonlighting workers (6), we obtain $-\mathrm{d}T'(\underline{y}^f) - T''(\underline{y}^f)\cdot \mathrm{d}\overline{y}^f = 0$. The term involving v'' is not present, since the total labor supply does not change. Rearranging yields

$$\tilde{\varepsilon}^s = \frac{\mathrm{d}\underline{y}^f}{\mathrm{d}[1-T'(\underline{y}^f)]}\frac{1-T'(\underline{y}^f)}{\underline{y}^f} = \frac{1-T'(\underline{y}^f)}{T''(\underline{y}^f)\cdot\underline{y}^f}. \tag{53}$$

Suppose that $\underline{y}^f(\cdot)$ is discontinuous at some productivity type θ_d such that $\underline{y}^f(\theta_d) > \underline{y}^f(\theta_d^-)$. Denote these two earnings levels by $\overline{s}$ and $\underline{s}$. Express the indifference condition

(8) as $T(\overline{s}) - T(\underline{s}) = T'(\overline{s}) \cdot (\overline{s} - \underline{s})$. Differentiating yields

$$\begin{aligned} &\mathrm{d}T(\overline{s}) - \mathrm{d}T(\underline{s}) + T'(\overline{s}) \cdot \mathrm{d}\overline{s} - T'(\underline{s}) \cdot \mathrm{d}\underline{s} \\ &\quad = \mathrm{d}T'(\overline{s}) \cdot (\overline{s} - \underline{s}) + T''(\overline{s}) \cdot \mathrm{d}\overline{s} \cdot (\overline{s} - \underline{s}) + T'(\overline{s}) \cdot (\mathrm{d}\overline{s} - \mathrm{d}\underline{s}). \end{aligned} \tag{54}$$

Note that $T'(\overline{s}) = T'(\underline{s})$, which follows from interior first-order condition (6) holding for type θ_d at both $\overline{s}$ and $\underline{s}$. Consequently, the right terms on the two sides cancel out and we get

$$\mathrm{d}\overline{s} = \frac{\mathrm{d}\big[T(\overline{s}) - T(\underline{s})\big]}{T''(\overline{s}) \cdot (\overline{s} - \underline{s})} - \frac{\mathrm{d}T'(\overline{s})}{T''(\overline{s})}. \tag{55}$$

The first term on the right-hand side captures the change of $\overline{s}$ due to moonlighting workers jumping to (or from) a discretely lower earnings $\underline{s}$ in response to change of the relative tax burden $\mathrm{d}[T(\overline{s}) - T(\underline{s})]$. The second term corresponds to the marginal response to change in the tax rate $\mathrm{d}T'(\overline{s})$. So as to isolate the jumping responses, suppose that $\mathrm{d}T'(\overline{s}) = 0$. Then we have that

$$\frac{\mathrm{d}\overline{s}}{\mathrm{d}\big[R(\overline{s}) - R(\underline{s})\big]} \frac{R(\overline{s}) - R(\underline{s})}{\overline{s}} = -\frac{1 - T'(\overline{s})}{\overline{s} \cdot T''(\overline{s})} = -\tilde{\varepsilon}^s(\overline{s}), \tag{56}$$

where $R(y) = y - T(y)$ is the formal after-tax income and we use $(R(\overline{s}) - R(\underline{s})) = (1 - T'(\overline{s})) \cdot (\overline{s} - \underline{s})$, which follows from (8). Thus, the jumping responses following the change in the relative after-tax income (or average net-of-tax rate) $\mathrm{d}[R(\overline{s}) - R(\underline{s})]$ are described by exactly the same elasticity as the marginal responses to a change of the (marginal) net-of-tax rate $\mathrm{d}[1 - T'(\overline{s})]$. The minus sign is due to the fact that $\overline{s}$ going up means that agents are jumping down—the income responses are of the opposite sign to $\mathrm{d}\overline{s}$.

Regarding the extensive margin responses, the earnings density of formal workers is $h^f(\overline{y}^f(\theta)) = (1 - G_\theta(\tilde{\kappa}(\theta)) \cdot f(\theta) \cdot (\frac{d\overline{y}(\theta)}{d\theta})^{-1}$. Differentiation yields

$$\begin{aligned} &\mathrm{d}h^f\big(\overline{y}^f(\theta)\big) + h^{f\prime}\big(\overline{y}^f(\theta)\big) \cdot \mathrm{d}\overline{y}^f(\theta) \\ &\quad = h^f\big(\overline{y}^f(\theta)\big) \cdot \Bigg(-\frac{g_\theta\big(\tilde{\kappa}(\theta)\big) \cdot \mathrm{d}\tilde{\kappa}(\theta)}{1 - G_\theta\big(\tilde{\kappa}(\theta)\big)} + \bigg(\frac{d\overline{y}(\theta)}{d\theta}\bigg) \cdot \mathrm{d}\Bigg[\bigg(\frac{d\overline{y}(\theta)}{d\theta}\bigg)^{-1}\Bigg]\Bigg). \end{aligned} \tag{57}$$

We are interested in the impact of the tax reform on the earnings density through the threshold $\tilde{\kappa}(\theta)$; the other terms are related to the intensive margin responses that we already accounted for. Ignoring the intensive margin terms, we obtain (the absolute value of) the semi-elasticity of the density of formal workers with respect to the tax burden of staying formal,

$$\pi(\theta) = -\frac{\mathrm{d}h^f\big(\overline{y}^f(\theta)\big)}{\mathrm{d}\big[T(\overline{y}^f(\theta) - T\big(\underline{y}^f(\theta)\big)\big]} \cdot \frac{1}{h^f\big(\overline{y}^f(\theta)\big)} = \frac{g_\theta\big(\tilde{\kappa}(\theta)\big)}{1 - G_\theta(\tilde{\kappa}(\theta)} \cdot \frac{1}{h^f\big(\overline{y}^f(\theta)\big)}, \tag{58}$$

where we used $\mathrm{d}\tilde{\kappa}(\theta) = \mathrm{d}[T(\overline{y}^f(\theta) - T(\underline{y}^f(\theta))]$, implied by (2).

While all the above elasticities are expressed as a function of the productivity type θ, we can express them equivalently as functions of formal earnings, since $\theta \mapsto \overline{y}^f(\theta)$ is strictly increasing over the entire domain and $\theta \mapsto \underline{y}^f(\theta)$ is strictly increasing when $\underline{y}^f(\theta) > 0$.

Appendix C: Two-dimensional heterogeneity in productivities

We can extend our optimal tax formula to an economy in which the distribution of formal and shadow productivities is two dimensional. Suppose that the type is given by $(\alpha, \theta, \kappa) \in \mathcal{A} \times [\underline{\theta}, \overline{\theta}] \times [0, \infty) \subseteq \mathbb{R}^3_+$, $\mathcal{A}$ finite, such that a worker of type (α, θ, κ) has productivities $w^f_\alpha(\theta)$ and $w^s_\alpha(\theta)$. Assume that the productivity schedules satisfy Assumption 1 for each $\alpha \in \mathcal{A}$. Denote the probability density function (pdf) of α by $A(\cdot)$, the cdf of θ conditional on α by $F_\alpha(\cdot)$, and the cdf of κ conditional on (α, θ) by $G_{\alpha,\theta}(\cdot)$. All cumulative density functions are assumed to be twice continuously differentiable. We maintain Assumption 2 and suppose that Assumption 3 holds for each α to rule out formal workers responding with jumps. In what follows, x_α denotes variable x evaluated at type α and x without a subscript indicates the average across α types. For instance, $h_\alpha(z)$ is the formal earnings density of workers with type α at earnings z, while $h(z) = \sum_{\alpha \in \mathcal{A}} h_\alpha(z) \cdot A(\alpha)$ is the overall formal earnings density at z.

Consider initial tax $T \in \mathcal{C}^2$ and direction of tax reform $\mathrm{d}T \in \mathcal{C}^2$. The optimal tax schedule needs to satisfy

$$\sum_{\alpha \in \mathcal{A}} [\mathrm{ME}_\alpha + \mathrm{WE}_\alpha + \mathrm{BE}_{\mathrm{int},\alpha} + \mathrm{BE}_{\mathrm{ext},\alpha}] \cdot A(\alpha) = 0. \tag{59}$$

The following theorem expresses this condition as a tax formula.

Theorem 2. *Suppose that the optimal tax schedule is twice continuously differentiable. The optimal tax rate at earnings y satisfies*

$$\begin{aligned} &\frac{T'(y)}{1 - T'(y)} \cdot \tilde{\varepsilon}^f(y) \cdot y \cdot h^f(y) + \sum_{s \in \overline{S}(y)} \frac{T'(s)}{1 - T'(s)} \cdot \tilde{\varepsilon}^s(s) \cdot s \cdot h^s(s) \\ &\quad = \int_y^\infty \left[1 - \overline{\lambda}(z)\right] dH(z) - \int_y^\infty \overline{\Delta T}(z) \cdot \overline{\pi}(z, y)\, dH^f(z), \end{aligned} \tag{60}$$

where (i) $\overline{S}(y) = \{\overline{s}_\alpha(y) : \alpha \in \mathcal{A}\}$, for $\overline{s}_\alpha(y)$ defined as $\min\{z \in \mathrm{Im}(\underline{y}^f_\alpha) : z \geq y\}$ if min exists and 0 otherwise, is a set of formal earnings at which moonlighting workers are responding on the intensive margin to the raise of $T'(y)$, (ii) $\overline{\Delta T}(z) = \sum_\alpha \Delta T_\alpha(z) \cdot h^f_\alpha(z) \cdot A(\alpha)$ is the average tax burden of staying formal at z, and (iii) $\overline{\pi}(z, y) = \sum_{\alpha \in \mathcal{A}} \frac{\Delta T_\alpha(z)}{\overline{\Delta T}(z)} \cdot \pi_\alpha(z) \cdot \mathbb{1}(\rho_\alpha(z) \leq y) \cdot \frac{h^f_\alpha(z) \cdot A(\alpha)}{h^f(z)}$ is the average semi-elasticity of density of formal workers at earnings z with respect to $T'(y)$ weighted by the tax burden of staying formal.

Proof. Following the steps in the proof of Theorem 1, it is easy to show that optimality condition (21) holds if and only if at every $y > 0$ we have

$$\sum_\alpha A(\alpha)\bigg[-\frac{T'(y)}{1-T'(y)}\tilde{\varepsilon}^f(y)yh^f_\alpha(y) - \frac{T'(\overline{s}_\alpha(y))}{1+T'(\overline{s}_\alpha(y))}\tilde{\varepsilon}^s_\alpha(\overline{s}_\alpha(y))\overline{s}_\alpha(y)h^s_\alpha(\overline{s}(y))$$
$$+\int_y^\infty [1-\overline{\lambda}_\alpha(z)]\,dH_\alpha(z) + \int_y^\infty \Delta T_\alpha(z)\pi_\alpha(z)\mathbb{1}(\rho_\alpha(z)\le y)\,dH^f_\alpha(z)\bigg] = 0. \tag{61}$$

By averaging each term within square brackets over α types separately, it is easy to express the average mechanical and welfare effects terms in the equation above as in (60). The same is true for the term capturing the intensive responses of formal workers, since all formal workers with earnings y have identical elasticity $\tilde{\varepsilon}^f(y)$.

The intensive responses of moonlighting workers can now happen at multiple earnings levels simultaneously, captured by the set $\overline{S}(y) = \{\overline{s}_\alpha(y) : \alpha \in \mathcal{A}\}$. Furthermore, for each $s \in \overline{S}(y)$, we have $h^s(s) = \sum_\alpha h^s_\alpha(s)\mathbb{1}(\overline{s}_\alpha(y) = s)A(\alpha)$, i.e., at each earnings level $s \in \overline{S}(y)$, there are only the moonlighting workers who respond on the intensive margin to a change of $T'(y)$. To shows this, suppose that there is some type (α_1, θ_1) and earnings level $y > 0$ for which $\underline{y}^f_{\alpha_1}(\theta_1^-) < y < \underline{y}^f_{\alpha_1}(\theta_1) = \overline{s}_{\alpha_1}(y)$. Suppose there is some other type (α_2, θ_2) for which $\underline{y}^f_{\alpha_2}(\theta_2) = \underline{y}^f_{\alpha_1}(\theta_1)$. Then (6) implies that $\frac{w^s_{\alpha_1}(\theta_1)}{w^f_{\alpha_1}(\theta_1)} = \frac{w^s_{\alpha_2}(\theta_2)}{w^f_{\alpha_2}(\theta_2)}$, and then (8) implies that a moonlighting worker of type (α_2, θ_2) is indifferent between $\underline{y}^f_{\alpha_1}(\theta_1^-)$ and $\underline{y}^f_{\alpha_1}(\theta_1)$. Thus, $\overline{s}_{\alpha_2}(y) = \overline{s}_{\alpha_1}(y)$.

Finally, the average extensive margin term can be expressed as in formula (60) by moving the summation sign under the integral, and then multiplying and dividing the sum by $\overline{\Delta T}(z)$, defined in the statement of the theorem. □

The formula is remarkably similar to that found in the case of one-dimensional heterogeneity in productivities. There are two differences. First, raising the marginal tax rate at a single income level may lead to intensive margin responses of moonlighting workers at multiple formal income levels. Suppose the marginal tax rate at y is increased. There can be moonlighting workers of type α_1 at formal earnings y responding marginally, as well as moonlighting workers of type α_2 at formal earnings $y' > y$ responding by jumping down. If the tax schedule features multiple regions of regressivity, a change of marginal tax rate can trigger jump responses of moonlighting workers at multiple earnings levels. In contrast, in the single-dimensional case considered in the main text, the moonlighting workers always responded at a single earnings level. The second difference concerns the extensive margin responses. Formal workers with identical earnings z but different types α can have different propensity to join the shadow economy (controlled by the semi-elasticity $\pi_\alpha(z)$) as well as different earnings conditional on joining the shadow economy (implying different $\Delta T_\alpha(z)$). Thus, we need to consider the average semi-elasticity of density of formal workers as well as the average tax revenue loss from joining the informal sector.

Appendix D: Estimation details

We use the 2013 wave of the household survey by the official statistical agency of Colombia (DANE). We restrict attention to individuals aged 24–50 years without children (34,000 observations). The indicator variable I^f is set equal to 1 if the worker reports being affiliated to all three components of social security: pension system, health insurance, and labor accidents insurance. The variable I^f indicates formality of the main job and does not imply that the worker is exclusively formal.

Table 2. Description of variables and parameter estimates.

Variable	Description	Estimate	Std. Error
Productivity schedules			
$\log(w_0^f)$	Formal log productivity at the bottom	0.003	(1e−4)
γ^f	Slope of formal log productivity	4.64	(0.06)
$\log(w_0^s)$	Informal log productivity at the bottom	0.006	(1e−4)
γ^s	Slope of informal log productivity	2.90	(0.06)
α_w	Tail parameter of w^f distribution	2.25	(0.03)
Type distribution			
σ_ϵ	Std. deviation of θ given observables	0.09	(2e−3)
σ_κ	Parameter of the fixed cost distribution	1.38	(0.03)
α_κ	Parameter of the fixed cost distribution	0.88	(0.01)
w_κ	Parameter of the fixed cost distribution	0.018	(2e−4)
σ_u	Std. dev. of the wage measurement error	0.53	(3e−3)
Preferences			
Γ	Weight on labor supply in the utility function	0.032	(8e−4)
Coefficients of variables included in X			
Gender	Dummy variable equal to 1 for women	−0.08	(2e−3)
Age	Age of the worker	0.04	(1e−4)
Age^2	Age squared	−5e−4	(6e−6)
Educ	Number of education years	0.02	(5e−4)
Degree	Highest degree achieved (no degree to doctorate)	0.05	(1e−3)
Work	Number of months worked in the last year	0.02	(7e−4)
Exper	Number of months worked in the last job	6e−5	(6e−5)
1stJob	Dummy for the first job (1 if it is the first job)	2e−4	(2e−5)
S-Man	Dummy for the manufacturing sector	−0.04	(1e−3)
S-Fin	Dummy for financial intermediation	0.14	(3e−3)
S-Ret	Dummy for the sales and retailers sector	−0.012	(3e−4)
B-city	Dummy for a firm in one of the two largest cities	0.10	(3e−3)
Size	Categories for the number of workers	0.11	(2e−3)
Lib	Dummy for a liberal occupation	0.25	(6e−3)
Admin	Dummy for an administrative task	−5e−3	(1e−4)
Seller	Dummy for sellers and related	4e−3	(1e−4)
Services	Dummy for a service task	−0.02	(6e−4)
Union	Dummy for labor union affiliation (1 if yes)	0.17	(4e−3)
Agency	Dummy for agency hiring (1 if yes)	−0.015	(3e−4)
Senior	Number of months of the worker in the firm	7e−4	(1e−5)

Note: Standard errors are obtained by case resampling bootstrap using 150 draws.

We use two questions of the survey to construct our measure of the hourly wage $\mathcal{W}$: first the worker is asked what was her income at the main job last month; second, what is the number of hours she "normally" works at that job. We use the ratio of the reported income and hours in those questions to compute our measure of the hourly wage. Since the normal number of hours need not correspond to last month's number of hours, we explicitly introduce a measurement error.[42] If the worker is identified to be formal at the main job, we include the statutory payroll taxes that are paid by the employer in the computation of the pre-tax income at the main job. Finally, we construct the empirical tax and transfer schedule by taking into account the personal income tax, payroll taxes that are not linked to benefits, and cash transfers to low income individuals.

The parameter estimates are reported in Table 2. The estimated density of types and the model fit are shown in Figure 4 in the main text.

References

Ábrahám, Aarpád and Nicola Pavoni (2005), "The efficient allocation of consumption under moral hazard and hidden access to the credit market." *Journal of the European Economic Association*, 3, 370–381. [756]

Allingham, Michael G. and Agnar Sandmo (1972), "Income tax evasion: A theoretical analysis." *Journal of Public Economics*, 1, 323–338. [752]

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998), "Tax compliance." *Journal of Economic Literature*, 36, 818–860. [752]

Atkinson, Anthony Barnes (1990), "Public economics and the economic public." *European Economic Review*, 34, 225–248. [754]

Bachas, Pierre Jean, Lucie Gadenne, and Anders Jensen (2020), "Informality, consumption taxes and redistribution." [753]

Beaudry, Paul, Charles Blackorby, and Dezsö Szalay (2009), "Taxes and employment subsidies in optimal redistribution programs." *American Economic Review*, 99, 216–242. [752]

Bergstrom, Katy and William Dodds (2021), "Optimal taxation with multiple dimensions of heterogeneity." *Journal of Public Economics*, 200, 104442. [751, 758, 759]

Boadway, Robin, Maurice Marchand, and Pierre Pestieau (1994), "Towards a theory of the direct-indirect tax mix." *Journal of Public Economics*, 55, 71–88. [753]

Boadway, Robin and Motohiro Sato (2009), "Optimal tax design and enforcement with an informal sector." *American Economic Journal: Economic Policy*, 1, 1–27. [752]

Casella, George and Roger L. Berger (2002), *Statistical Inference*, second edition. Duxbury, Pacific Grove, CA. [777]

[42]We further assume that survey respondents correctly reveal their gross income from the main job, regardless of whether the main job is formal or informal. Other papers making this assumption include Meghir, Narita, and Robin (2015) for Brazil and López García (2015) for Chile.

Chetty, Raj (2012), “Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply.” *Econometrica*, 80, 969–1018. [769]

Cremer, Helmuth and Firouz Gahvari (1996), “Tax evasion and the optimum general income tax.” *Journal of Public Economics*, 60, 235–249. [752]

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2013), “The demand for, and consequences of, formalization among informal firms in Sri Lanka.” *American Economic Journal: Applied Economics*, 5, 122–150. [754]

de Oliveira, Oswaldo (2014), “The implicit and the inverse function theorems: Easy proofs.” *Real Analysis Exchange*, 39, 207–218. [776]

De Paula, Aureo and Jose A. Scheinkman (2010), “Value-added taxes, chain effects, and informalit.” *American Economic Journal: Macroeconomics*, 2, 195–221. [753]

Di Nola, Alessandro, Georgi Kocharkov, Almuth Scholl, and Anna-Mariia Tkhir (2020), “The aggregate consequences of tax evasion.” [752]

Diamond, Peter A. (1998), “Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates.” *American Economic Review*, 83–95. [754, 765, 771, 772]

Doligalski, Paweł and Luis Rojas (2016), “Optimal redistribution with a shadow economy.” EUI Working Papers. [752, 754]

Ebert, Udo (1992), “A reexamination of the optimal nonlinear income tax.” *Journal of Public Economics*, 49, 47–73. [754]

Emran, M. Shahe and Joseph E. Stiglitz (2005), “On selective indirect tax reform in developing countries.” *Journal of Public Economics*, 89, 599–623. [752]

French, Eric and Christopher Taber (2011), “Identification of models of the labor market.” In *Handbook of Labor Economics*, volume 4 (Orley Ashenfelter and David Card, eds.), 537–617, Elsevier. [770]

Fudenberg, Drew and Jean Tirole (1991), *Game Theory*. MIT Press, Cambridge, MA. [775]

Goldberg, Pinelopi Koujianou and Nina Pavcnik (2003), “The response of the informal sector to trade liberalization.” *Journal of Development Economics*, 72, 463–496. [768]

Golosov, Mikhail and Aleh Tsyvinski (2007), “Optimal taxation with endogenous insurance markets.” *The Quarterly Journal of Economics*, 122, 487–534. [756]

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin (2014), “A variational approach to the analysis of tax systems.” Technical report, National Bureau of Economic Research. [759]

Gomes, Renato, Jean-Marie Lozachmeur, and Alessandro Pavan (2017), “Differential taxation and occupational choice.” *The Review of Economic Studies*, rdx022. [752]

Gorodnichenko, Yuriy, Jorge Martinez-Vazquez, and Klara Sabirianova Peter (2009), "Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia." *Journal of Political Economy*, 117, 504–554. https://ideas.repec.org/a/ucp/jpolec/v117y2009i3p504-554.html. [750]

Guataquí, Juan Carlos, Andrés Felipe García, and Mauricio Rodríguez (2010), "El Perfil de la Informalidad Laboral en Colombia." Perfil de Coyuntura Económica. Available at https://ideas.repec.org/a/col/000165/008964.html. [768]

Heckman, James J. and Bo E. Honore (1990), "The empirical content of the Roy model." *Econometrica*, 1121–1149. [770]

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2006), "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond." *Handbook of the Economics of Education*, 1, 307–458. [769]

Hendren, Nathaniel (2019), "Efficient welfare weights." NBER Working Paper No. 20351. [759]

Henley, Andrew, G. Reza Arabsheibani, and Francisco G. Carneiro (2009), "On defining and measuring the informal sector: Evidence from Brazil." *World Development*, 37, 992–1003. [750]

Huang, Jason and Juan Rios (2016), "Optimal tax mix with income tax non-compliance." *Journal of Public Economics*, 144, 52–63. [753]

Hussmanns, Ralf and Brigitte du Jeu (2002), "ILO compendium of official statistics on employment in the informal sector." ILO Working Papers, International Labour Organization. Available at https://EconPapers.repec.org/RePEc:ilo:ilowps:993540423402676. [750]

ILO (2013), "Measuring informality: A statistical manual on the informal sector and informal employment." Technical report, International Labour Organization. Available at http://www.ilo.org/global/publications/ilo-bookstore/order-online/books/WCMS_222979/lang--en/index.htm. [768]

ILO (2014), "Trends in informal employment in Colombia: 2009–2013." Technical report, International Labour Organization. Available at http://www.ilo.org/americas/sala-de-prensa/WCMS_245885/lang--en/index.htm. [768]

ILO (2018), "Women and men in the informal economy: A statistical picture (third edition)." Technical report, International Labour Organization. Available at https://www.ilo.org/global/publications/books/WCMS_626831/lang--en/index.htm. [749, 751]

Jacobs, Bas (2015), "Optimal inefficient production." Report, Erasmus University Rotterdam. [752]

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden (2013), "Optimal redistributive taxation with both extensive and intensive responses." *Journal of Economic Theory*, 148, 1770–1805. [765, 771]

Jacquet, Laurence and Etienne Lehmann (2021), "Optimal income taxation with composition effects." *Journal of the European Economic Association*, 19, 1299–1341. [759]

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez (2011), "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica*, 79, 651–692. [750]

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez (2009), "The optimal income taxation of couples." *Econometrica*, 77, 537–560. [752]

Kleven, Henrik Jacobsen, Wolfram F. Richter, and Peter Birch Sørensen (2000), "Optimal taxation with household production." *Oxford Economic Papers*, 52, 584–594. [753]

Kopczuk, Wojciech (2001), "Redistribution when avoidance behavior is heterogeneous." *Journal of Public Economics*, 81, 51–71. [752]

Leal Ordóñez, Julio (2014), "Tax collection, the informal sector, and productivity." *Review of Economic Dynamics*, 17, 262–286. [752]

López García, Italo (2015), "Human capital and labor informality in Chile: A life-cycle approach." Technical report, RAND working paper series. [786]

Magnac, Th. (1991), "Segmented or competitive labor markets." *Econometrica*, 165–187. [753, 754, 769]

Meghir, Costas, Renata Narita, and Jean-Marc Robin (2015), "Wages and informality in developing countries." *American Economic Review*, 105, 1509–1546. [786]

Milgrom, Paul and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 583–601. [757, 777]

Mirrlees, James A. (1971), "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, 175–208. [750]

Mirrlees, James A. (1982), "Migration and optimal income taxes." *Journal of Public Economics*, 18, 319–341. [764]

Monteiro, Joana CM and Juliano J. Assunção (2012), "Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises." *Journal of Development Economics*, 99, 105–115. [750]

Mora, John and Juan Muro (2017), "Dynamic effects of the minimum wage on informality in Colombia." LABOUR. [768]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [754]

Olovsson, Conny (2015), "Optimal taxation with home production." *Journal of Monetary Economics*, 70, 39–50. [753]

Piketty, Thomas and Emmanuel Saez (2013), "Optimal labor income taxation." In *Handbook of Public Economics*, volume 5 (Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds.), 391. Newnes. [752, 758, 780]

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva (2014), "Optimal taxation of top labor incomes: A tale of three elasticities." *American Economic Journal: Economic Policy*, 6, 230–271. [752, 767]

Pratap, Sangeeta and Erwan Quintin (2006), "Are labor markets segmented in developing countries? A semiparametric approach." *European Economic Review*, 1817–1841. [754, 769, 771]

Rocha, Rudi, Gabriel Ulyssea, and Laísa Rachter (2018), "Do lower taxes reduce informality? Evidence from Brazil." *Journal of Development Economics*, 134, 28–49. [750, 754]

Rothschild, Casey and Florian Scheuer (2013), "Redistributive taxation in the Roy model." *The Quarterly Journal of Economics*, 128, 623–668. [772]

Rothschild, Casey and Florian Scheuer (2014), "A theory of income taxation under multidimensional skill heterogeneity." Technical report, National Bureau of Economic Research. [752]

Rothschild, Casey and Florian Scheuer (2016), "Optimal taxation with rent-seeking." *The Review of Economic Studies*, 83, 1225–1262. [771]

Saez, Emmanuel (2001), "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, 68, 205–229. [754, 759, 765, 771, 772]

Saez, Emmanuel (2002), "Optimal income transfer programs: Intensive versus extensive labor supply responses." *The Quarterly Journal of Economics*, 117, 1039–1073. [765, 771]

Scheuer, Florian (2014), "Entrepreneurial taxation with endogenous entry." *American Economic Journal: Economic Policy*, 6, 126–163. [752]

Scheuer, Florian and Iván Werning (2017), "The taxation of superstars." *The Quarterly Journal of Economics*, 132, 211–270. [764]

Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro (2011), "Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007." In *Handbook on the Shadow Economy* (Friedrich Schneider, ed.), 9–77, Edward Elgar Cheltenham. [749]

Schneider, Friedrich and Dominik H. Enste (2000), "Shadow economies: Size, causes, and consequences." *Journal of Economic Literature*, 38, 77. [749]

Selin, Håkan and Laurent Simula (2020), "Income shifting as income creation?" *Journal of Public Economics*, 182, 104081. [752]

Slemrod, Joel and Wojciech Kopczuk (2002), "The optimal elasticity of taxable income." *Journal of Public Economics*, 84, 91–112. [752]

Slemrod, Joel and Shlomo Yitzhaki (2002), "Tax avoidance, evasion, and administration." *Handbook of Public Economics*, 3, 1423–1470. [752]

Statistics Poland (2019), "Unregistered employment in Poland in 2017." Technical report, Statistics Poland. Available at https://stat.gov.pl/en/

topics/labour-market/working-unemployed-economically-inactive-by-lfs/unregistered-employment-in-poland-in-2017,13,5.html. [750]

Tazhitdinova, Alisa (2017), "Increasing hours worked: Moonlighting responses to a large tax reform." Available at SSRN 3047332. [750]

Waseem, Mazhar (2018), "Taxes, informality and income shifting: Evidence from a recent pakistani tax reform." *Journal of Public Economics*, 157, 41–77. [750]

Co-editor Florian Scheuer handled this manuscript.

Manuscript received 14 October, 2020; final version accepted 6 May, 2022; available online 31 May, 2022.