Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium*

Dominik Sachs
European University Institute

Aleh Tsyvinski
Yale University

Nicolas Werquin
Toulouse School of Economics

July 22, 2016

Abstract

We study the incidence and the optimal design of arbitrary nonlinear income taxes in a Mirrleesian economy with a continuum of endogenous wages. Using a variational approach, we characterize the general equilibrium effects of reforming any baseline tax system on individual labor supplies, wages, and utilities, as well as on government revenue and social welfare. We show that for a general production function, analyzing the economic incidence of taxation reduces mathematically to solving an integral equation, for which we derive a general analytical solution and provide a clear economic interpretation. For specific production functions, such as CES and Translog, this solution becomes particularly tractable and insightful. In particular, we show theoretically that starting from a tax schedule that closely approximates the U.S. tax code, general equilibrium forces raise the benefits on government revenue of increasing the progressivity of the tax schedule. We then specialize our analysis of tax incidence to the study of optimal taxes. The variational approach in this case allows us to gain sharp economic insights about the design of optimal taxes in general equilibrium, in addition to those obtained from the more traditional mechanism design approach. When the production function is CES we then provide a strikingly simple closed-form solution for the optimal top tax rate and show that the U-shape of optimal marginal tax rates in partial equilibrium gets reinforced. One insight from our joint analysis of tax incidence and optimal taxation is that the results about the optimum tax schedule may actually be reversed when considering reforms of a suboptimal (e.g., the U.S.) tax code. We finally provide extensive numerical simulations for both tax incidence and optimal taxes. The welfare gains from taking into account general equilibrium effects are quantitatively significant.

*We thank Costas Arkolakis, Michael Peters, and Gianluca Violante for helpful comments and suggestions, and especially Philip Ushchev for detailed and insightful discussions.
Introduction

In this paper we study the incidence and the optimal design of nonlinear income taxes in a general equilibrium Mirrlees (1971) economy. We consider an aggregate production function with a continuum of labor inputs and imperfect substitutability between skills, so that the wage of each skill type, equal to its marginal product, is endogenous.

We connect two classical strands of the public finance literature that have so far been somewhat disconnected: the tax incidence literature (Harberger, 1962; Kotlikoff and Summers, 1987; Fullerton and Metcalf, 2002), and the literature of optimal nonlinear income taxation Mirrlees (1971); Stiglitz (1982); Diamond (1998); Saez (2001). The object of the tax incidence analysis (in our context with one sector and several labor inputs) is to characterize the effects of tax policy on the distribution of individual wages, labor supplies, and utilities, as well as on government revenue and social welfare. That is, we are interested in understanding who bears the economic burden of taxes. What makes this question conceptually and technically difficult (and interesting) is that the economic burden is typically very different, in a general equilibrium environment, from the statutory burden of taxes: the agent who pays the check to the tax authority adjusts her labor supply behavior, which in turn affects the wages that all other individuals face in a non-trivial way, making them bear part of the economic burden. Our goal is thus to analyze the fundamental economic forces underlying these distributional effects of taxation to guide economic policy.

We extend the existing tax incidence literature by considering arbitrary nonlinear taxes and a continuum of endogenous wages. In partial equilibrium (when wages are fixed), the effects of a tax change on the labor supply of a given agent can be straightforwardly derived as a function of the elasticity of labor supply of that agent (see Saez (2001)). The key difficulty in general equilibrium is that this direct partial equilibrium effect affects the wage, and thus the labor supply, of every other individual. This in turn feeds back into the wage distribution, which further impacts labor supply decisions, and so on. Moreover, even if we are only interested in the aggregate (e.g., social welfare) effects of the change in taxes, we must derive independently each individual’s response to the perturbation, since both the baseline tax schedule and the tax reform itself are allowed to be arbitrarily nonlinear. Solving for the fixed point in each type’s labor supply adjustment is the key step in the analysis of tax incidence and the primary technical challenge of our paper.

We first show that this a priori complex problem of deriving the effects of tax reforms on individual labor supply reduces mathematically to solving an integral equation, i.e., an equation where the unknown is a function that appears under an integral sign. The mathematical theory of integral equations allows us to derive an abstract analytical solution to this problem for a general production function, which furthermore has a clear economic meaning. Specifically, this solution can be represented as an infinite sum, where the first term is the direct (partial equilibrium) impact of the reform, and each of the subsequent terms captures a successive round of general equilibrium cross-wage feedback effects. Having derived this general expression for the change in labor supply, the formulas giving the incidence of any nonlinear tax reform on individual wages, utilities, government revenue, and social welfare can be easily derived.
Our second set of results is that for specific functional forms of the technology, in particular CES and Translog, the integral equation admits simpler closed form solutions, which become particularly tractable and insightful. Specifically, we can obtain such simple solutions when the cross-wage elasticities (and hence the kernel of the integral equation) can be expressed as the sum of multiplicatively separable functions.

In particular, we show that this occurs for a CES production function, since in this case a change in the labor supply of a given type has the same effect (in percentage terms) on the wage of every other type. We start by deriving a transparent closed form solution for the tax incidence of any local tax reform in this context. Next, we consider a particularly important baseline tax system, namely, the “constant rate of progressivity” (CRP) tax schedule,\(^1\) that closely approximates the current U.S. tax code. We show two striking results for this case. First, the general equilibrium effects of any (linear or nonlinear) tax reform on aggregate revenue are equal to zero if the baseline tax schedule is linear (although they are non-zero at the individual level). Second, and more generally, if the baseline tax system is nonlinear, we show that the revenue gains of raising the marginal tax rate at a given income level are lower than in partial equilibrium for low incomes, and higher for high incomes. In other words, starting from the progressive U.S. tax code, general equilibrium forces raise the benefits of increasing the progressivity of the tax schedule.

At first sight this result may seem to be at odds with the familiar insight of Stiglitz (1982) in the two-income model, which we generalize to a continuum of incomes in our section on optimal taxation. Indeed, those results say that the optimal tax rates should be lower at the top, and higher at the bottom, of the income distribution, relative to the partial equilibrium benchmark. That is, the optimal tax schedule should be more regressive when the general equilibrium forces are taken into account. The reason for the difference with our tax incidence results is that we consider here reforms of the current (suboptimal) U.S. tax code, with low marginal tax rates at the bottom. Instead, the results about the optimum tax schedule use as a benchmark the optimal partial equilibrium tax schedule, which features high marginal tax rates at the bottom (see Diamond (1998); Saez (2001)). The key insight is therefore that results about the optimum tax schedule may actually be reversed when considering reforms of the current U.S. tax code. This underlines the importance of our tax reform approach, and leads us to conclude that one should be cautious, in practice, when applying the results of the general equilibrium optimal tax theory. This insight is reminiscent of Guesnerie (1977), who shows in the setting of Diamond and Mirrlees (1971) that directions of local tax reforms may be desirable despite being suboptimal from an optimal taxation viewpoint.

We then consider the case of the Translog production function, for which the elasticities of substitution are non-constant. This technology is a particularly important example as it can be used as a second-order approximation to any production function (see details in Christensen et al. (1973)). We first propose a particularly tractable specification of the Translog technology with a continuum of inputs, for which the cross-wage elasticities are decreasing in the distance between types – the smaller the distance between skills, the stronger substitutes they are. Moreover, this

\(^1\)See, e.g., Heathcote et al. (2014) for a recent study in the context of the general equilibrium.
specification is such that the distance between types reduces their substitutability linearly, which again allows us to derive a simple closed-form solution to the integral equation. We obtain that the general equilibrium effects of tax reforms on individual labor supply in this case can be expressed as a simple affine function of the own-wage elasticity and the cross-wage elasticity with an “average” (endogenous) type. We show moreover in the Appendix how this technique can be fruitfully extended to other production functions.

Our third set of results concerns the derivation of optimal taxes in this general equilibrium setting. We start with a general production function. Our first result gives the characterization of the optimum using both the mechanism design approach as in Mirrlees (1971) (or Stiglitz (1982) in general equilibrium) and the variational (or “tax reform”) approach in the spirit of Saez (2001). While these two approaches deliver, of course, the same result for the optimal taxes, the mechanism design characterization is quite complicated both theoretically and economically, as it depends both on the pattern and the size of the multipliers on the binding incentive constraints for all types. We thus turn to the variational approach (as a special case of our tax incidence analysis discussed above), which allows us to get a sharper intuition and decompose the key economic forces behind the design of optimal taxes in general equilibrium. Note that in principle, our tax incidence analysis immediately delivers a characterization of the optimum tax schedule, by equating the welfare effects of any tax reform to zero. However, we propose a specific tax reform that leads to a substantially simpler optimal formula. Specifically, we combine the Saez (2001) partial equilibrium nonlinear tax perturbation at a given income level, with an additional reform that keeps all other agents’ labor supplies fixed. This counteracting reform allows us to significantly simplify the general integral equation obtained in our tax incidence analysis, by effectively canceling all of the feedback general equilibrium effects. It has, however, additional revenue and welfare implications which form the core of our economic decomposition of the optimum. This technique allows us to isolate the novel effects that arise purely from general equilibrium forces. Intuitively, the Saez (2001) perturbation at a given income level affects the wage of all other agents. Thus the counteracting tax reform must perturb, roughly speaking, the tax rate of every other agent by the same amount by which their wage has changed. The key insight here is that there is a crucial difference between the changes in the wage and in the marginal tax rate induced by the combination of these perturbations. Indeed, a change in the wage affects only the revenue levied at a given income level. In contrast, a change in the marginal tax rate at a given income level increases uniformly the total tax liability paid by everyone above this income by a lump-sum amount. Summing these two effects yields our characterization of the optimum. In passing, this analysis also allows us to shed a new light on the difference between Saez (2004) and Scheuer and Werning (2016), on the one hand, for whom the optimum tax schedule obtained in partial equilibrium applies regardless of the technology, and Stiglitz (1982), Rothschild and Scheuer (2014), Ales et al. (2015), and our paper, on the other hand.

We then show that the optimal tax formula can be represented as an integral equation, with the net-of-tax rate as the unknown function. This equation becomes particularly simple if the production function is CES,\(^2\) in which case we derive a simple analytical expression for the optimal

\(^2\)The CES production function is the most parsimonious extension of the partial equilibrium model, which is
The solution highlights two main differences between the partial equilibrium and the general equilibrium optimal taxes. These two differences come from the two types of wage effects that occur in general equilibrium: the own-wage effects (of a given type’s labor supply on their own wage) and the cross wage-effects (of a given type’s labor supply on other types’ wages). The first of these general equilibrium forces is reflected in a general equilibrium-corrected inverse elasticity in the standard optimal tax formula. The second effect is represented as a simple combination of the Pareto weights and the constant elasticity of substitution. Perhaps surprisingly, the economic consequence of this latter effect is that if the government values the welfare of a given individual less (more) than average, her marginal tax rate should be higher (lower) than in partial equilibrium – this makes her work less and earn a higher wage, at the expense of reducing everyone else’s wage.

We then show how the two classical results from the Mirrleesian literature, namely the top tax rate of Saez (2001), and the U-shape result for optimal marginal tax rates of Diamond (1998), extend to the framework with endogenous wages when the production function is CES. We first obtain a strikingly simple closed form for the optimal top tax rate in terms of one additional parameter (the elasticity of substitution, that has been estimated in the literature), which immediately yields direct policy recommendations for the optimum tax on high incomes. We then turn to the bulk of the optimum tax schedule, and show that the general equilibrium forces make the familiar U-shape of marginal tax rates even more pronounced. In particular, in the increasing part of the U, marginal tax rates increase faster with income than in the partial equilibrium case, leading to a more progressive pattern of marginal tax rates.

Our final set of results is a quantitative analysis of both the tax incidence and the optimal taxes in the general equilibrium setting. The starting point of our numerical analysis is the calibration of the U.S. tax code based Heathcote et al. (2014). To meaningfully compare the general equilibrium optimal taxes to the partial equilibrium optimum, we further construct a particularly policy-relevant benchmark. Specifically, we compare our policy recommendations to those that one would obtain by applying the analysis of Diamond (1998), calibrating the model to the same income distribution, and making the same assumptions about the utility function. We theoretically decompose the source of the differences between the optimal policy and this partial equilibrium benchmark by providing a general equilibrium wedge accounting analysis. This analytical exercise highlights in particular the roles of the general equilibrium adjustments coming from the own-wage effects, the cross-wage effects, and the endogeneity of the hazard rates of the wage distribution. The simulations of optimal policies confirm our theoretical insights and their quantitative significance. For the Rawlsian case, we find that optimal marginal tax rates are also U-shaped and that this U-shape gets more pronounced than the partial equilibrium benchmark; the differences are magnified by the elasticity of substitution and the structural labor supply elasticity. Welfare gains from setting taxes optimally, relative to the partial equilibrium optimum benchmark, are significant and up to 3% gains in consumption equivalent.

---

3 As described in the context of tax incidence, we can also derive an explicit solution in the Translog case.
 Related Literature. This paper is related to the literature on tax incidence, in particular Harberger (1962), Shoven and Whalley (1984), and Kotlikoff and Summers (1987) and Fullerton and Metcalf (2002) for comprehensive surveys. This literature has typically focused on models with two inputs in production, capital and labor, each taxed at a constant rate. We briefly set up and analyze this incidence problem in a model with one sector and two inputs in Appendix A.1, in order to highlight the connection and unify the notations with our general environment. This paper extends the tax incidence framework in two directions. First, in our model production has a continuum of (labor) inputs; this extension introduces several technical difficulties which we show how to address. Second, and most importantly, we analyze the incidence of arbitrary nonlinear tax schedules. These are the two most important features that are necessary to study labor income taxation in the workhorse Mirrlees (1971) framework in general equilibrium.

The second main strand in the literature that our paper relates to is that of optimum nonlinear income taxation (Mirrlees, 1971; Diamond, 1998). As described in the previous paragraph, our goal is to extend this canonical taxation model to the general equilibrium environment, therefore contributing to the unification of the two major strands in the taxation literature (tax incidence and optimal tax design). Specifically, our benchmark is the Diamond (1998) version of the (Mirrlees, 1971) model, where the utility function has no income effects. We extend this setup by letting the wage of each type be endogenous and depend on the labor supply of every other type through an aggregate production function.

The optimal taxation problem in general equilibrium with arbitrary nonlinear tax instruments (and a mechanism design approach) has originally been studied by Stiglitz (1982). That paper characterizes optimal taxes in a model with two types and endogenous wages. The key result of Stiglitz (1982) is that at the optimum tax system, general equilibrium forces lead to a more regressive tax schedule than in partial equilibrium, as lowering the top tax rate and raising the bottom tax rate compresses the wage distribution, which relaxes the incentive constraint of the high type.

In the recent optimal taxation literature, there are two strands that relate to our work. First, a series of important contributions by Scheuer (2014); Rothschild and Scheuer (2013, 2014); Scheuer and Werning (2015), Chen and Rothschild (2015), and Ales et al. (2015) explicitly focus on optimal taxation in general equilibrium models and form the modern analysis of optimal nonlinear taxes in general equilibrium. Specifically, Rothschild and Scheuer (2013, 2014) generalize Stiglitz (1982) to a setting with \( N \) sectors and a continuum of (infinitely substitutable) skills in each sector, leading to a multidimensional screening problem. Ales et al. (2015) endogenize the production function by incorporating an assignment model into the Mirrlees framework and study the implications of technological change for optimal taxation. Our model is simpler than those of Rothschild and Scheuer (2013, 2014) and Ales et al. (2015). In particular, different types earn different wages (there is no “overlap” in the wage distributions of different types, as opposed to the framework of Rothschild and Scheuer (2013, 2014)), and the production function is exogenous (as opposed to being microfounded in an assignment model as in Ales et al. (2015)). Our simpler setting allows us to get a sharper characterization for the effects of taxes on individual and aggregate welfare, using
the tools of the theory of integral equations,\footnote{Integral equations arise naturally in settings where there is a continuum of types and the tax rate at one income level affects the behavior of an individual earning a different income. This is also the case in the context of optimization frictions in Farhi and Gabaix (2015).} and to directly compare our results to the (Mirrlees, 1971) benchmark. We detail the differences between our results and theirs below.

Second, Heathcote et al. (2014) study optimal tax progressivity in a model with restricted tax schedules of the CRP form in general equilibrium.\footnote{See also Heathcote and Tsuijyama (2016) for a detailed analysis of the performance of various restricted functional forms to approximate the fully optimal taxes.} Their production function is CES with continuum of skills, as in several parts of our paper. On the one hand, our model is simpler than their dynamic framework. On the other hand, we do not restrict ourselves to a particular functional form for taxes. Our papers share, however, one important goal – to derive simple closed form expressions for the effects of tax reforms in general equilibrium. Moreover, several key results of our paper (see, e.g., Corollary 3 and the quantitative analysis in Section 4) use their calibration of the CPR schedule as an approximation of the current U.S. tax system to study the effects of the tax reforms.

Aside from the methodological contribution of using integral equations, we broadly view our contribution relative to these two strands of the literature to be threefold.

• First, our study encompasses, but is not restricted to, the analysis of the optimum tax schedule with nonlinear taxes. As we already described, our focus is on the incidence problem of reforming any, potentially suboptimal, tax system (this leads in particular to a characterization of the optimum, by equating the welfare effects of these reforms to zero), and where the tax reform is not a priori restricted to any specific functional form of taxes. This more general analysis is important, because as we discussed earlier, the insights that are obtained for the optimum tax schedule may be reversed when considering reforms of suboptimal tax codes.

• Second, we derive our tax formulas from a very different angle than the mechanism design papers. In particular, our tax incidence formulas cannot be obtained with the traditional mechanism design tools used by those papers, which by construction can only characterize the full optimum. Instead, we unify the tax incidence and the optimal tax design problems into a common framework by using a variational, or “tax reform” approach, originally pioneered by Piketty (1997) and Saez (2001), and extended to several other contexts by, e.g., Kleven et al. (2009) (for the case of multidimensional screening), and Golosov et al. (2014) (for dynamic stochastic models). In this paper we extend this approach to the general equilibrium framework with endogenous wages. We show that our resulting optimal tax formula coincides with those obtained using the traditional mechanism design approach (e.g., our Proposition 4 using the mechanism design approach is similar to that derived by Rothschild and Scheuer (2013, 2014) and Ales et al. (2015)), but the tax incidence analysis uncovers different fundamental economic forces and provides an alternative, and arguably clearer, economic intuition for the characterization of the optimum (see, e.g., Proposition 5 and Corollary 4 ).
Third, our characterization of the optimum tax schedule is derived in the canonical Mirrleesian taxation framework with a continuum of incomes as in Diamond (1998) and Saez (2001), and therefore allows us to generalize the stark insights of Stiglitz (1982) in the two-type model to a policy-relevant environment, both theoretically and quantitatively. In particular, we revisit the key insights obtained in the partial equilibrium model, namely the closed-form formula for the optimal top tax rate and the U-shape of optimal marginal tax rates (see Diamond (1998) and Saez (2001)). In contrast, the typical two-type analyses can only give limited qualitative intuitions. More broadly, our quantitative analysis is in the spirit of those of Feldstein (1973) and Allen (1982), who also specified the production function to simple functional forms, but focused on the case of linear taxes.

The final set of papers that ours relates to are Saez (2004) and Scheuer and Werning (2016). These two papers show that under certain assumptions, an irrelevance result applies, namely, that the optimal tax formula derived by (Mirrlees, 1971) in partial equilibrium applies directly in general equilibrium irrespective of the technology (which is in stark constrast with the results of Stiglitz (1982)). Our model with a continuum of types fits into these authors’ framework by interpreting each type as a separate sector. In particular, Scheuer and Werning (2016) show that the (Mirrlees, 1971) optimal tax formula can be derived as a special case of Diamond and Mirrlees (1971) (written for a continuum of goods); because the analysis of the latter model holds irrespective of the technology, it follows that the partial equilibrium optimal tax characterization of (Mirrlees, 1971) also holds for any, not necessarily linear, production function. The key assumption needed for this irrelevance result to hold is that there exists a sufficiently rich number of tax instruments available to the government; namely, each sector must be taxed at a specific rate, as opposed to having a single (possibly nonlinear) tax schedule independent of the agent’s sector (or type). To analyze labor income taxation in general equilibrium, we assume instead in this paper that all individuals in the economy are subject to the same nonlinear tax schedule – in which case general equilibrium forces are present.

This paper is organized as follows. Section 1 describes our framework and defines the key structural elasticity variables. In Section 2 we generalize and analyze the tax incidence problem to a continuum of wages and nonlinear income taxes. In Section 3 we derive optimal taxes in general equilibrium. Finally, we present quantitative results for the U.S. in Section 4. In Appendix A we provide a short primer on tax incidence, in a version of our model with two types. The proofs are gathered in Appendix B.

---

6See also Green and Phillips (2015) which is a quantitative exploration of the Stiglitz two-type model.
1 Environment

1.1 Equilibrium

Individual and firm behavior

Individuals have preferences over consumption \( c \) and labor supply \( l \) given by \( U(c, l) = u(c - v(l)) \), where the functions \( u \) and \( v \) are twice continuously differentiable and strictly increasing, \( u \) is concave, and \( v \) is strictly convex. In particular, note that \( U \) implies no income effects on labor supply. There is a continuum of exogenous types (productivities) \( \theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+ \), distributed according to the p.d.f. \( f_\theta(\cdot) \) with c.d.f. \( F_\theta(\cdot) \). Each type is composed of a continuum of agents.

An individual of type \( \theta \) earns a wage \( w(\theta) \), which she takes as given. She chooses her labor supply \( l(\theta) \) and earns taxable income \( y(\theta) = w(\theta) l(\theta) \). The government imposes a tax schedule \( T(y) \) on income, where the function \( T: \mathbb{R}_+ \to \mathbb{R} \) is twice continuously differentiable. Individual \( \theta \) solves:

\[
l(\theta) = \arg \max_{l \in \mathbb{R}_+} u[w(\theta) l - T(w(\theta) l) - v(l)].
\]

The optimal labor supply \( l(\theta) \) chosen by individual \( \theta \) is the solution to the first-order condition:

\[
[1 - T'(w(\theta) l(\theta))] w(\theta) = v'(l(\theta)). \tag{1}
\]

We sometimes denote explicitly by \( l(\theta, T) \) the labor supply of individual \( \theta \) given a baseline tax schedule \( T \). We denote by \( U(\theta) \) her indirect utility function. Finally, denote by \( L(\theta) \equiv f_\theta(\theta) \) the total amount of labor supplied by individuals of type \( \theta \).

There is a continuum of mass 1 of identical firms that produce output \( Y \) using the labor of all types \( \theta \). The first technical difficulty is to rigorously define a production function with a continuum of inputs. Following Hart (1979) and Fradera (1986) (see also Scheuer and Werning (2016)), we represent the continuum of labor inputs as a Borel measure on a compact metric space. Specifically we define \( \mathcal{L} = \{L(\theta)\}_{\theta \in \Theta} \) to be a measure on \( (\Theta, \mathcal{B}(\Theta)) \), so that for any Borelian set \( B \in \mathcal{B}(\Theta) \) (e.g., an interval in \( \Theta \)), \( \mathcal{L}(B) \) is the total amount of labor supplied by individuals with productivity \( \theta \in B \). Denote by \( \mathcal{M} \) the space of all finite, non-negative Borel measures \( \mathcal{L} \) on \( \Theta \). We then define the production function \( \mathcal{F} \) as:

\[\mathcal{Y} = \mathcal{F}(\mathcal{L}) = \mathcal{F} \left( \{L(\theta)\}_{\theta \in \Theta} \right).\]

We assume that the production function has constant returns to scale. The representative firm chooses the demand of inputs (labor of each type), taking as given the wage \( w(\theta) \), to maximize its profit

\[
\max_{\mathcal{L}} \left[ \mathcal{F}(\mathcal{L}) - \int_{\Theta} w(\theta) L(\theta) d\theta \right].
\]

As a result, in equilibrium firms earn no profits and the wage \( w(\theta) \) is equal to the marginal productivity of type-\( \theta \) labor. Formally, \( w(\theta) \) is equal to the Gateaux derivative of the production function
when the labor effort schedule $L$ is perturbed in the direction $\delta \theta$, where $\delta \theta$ is the Dirac measure at $\theta$:

$$w(\theta) = \lim_{\mu \to 0} \frac{1}{\mu} \left\{ F(L + \mu \delta \theta) - F(L) \right\}. \quad (2)$$

That is, the wage of an individual of type $\theta$ is equal to the increase in output when the labor effort of agents of type $\theta$, $L(\theta)$, is marginally increased, while the labor effort of all other agents remains constant (since $\delta \theta(\theta') = 0$ for all $\theta' \neq \theta$). For intuitive purposes, one can interpret equation (2) as $w(\theta) = \partial \partial_{L(\theta)} F(L(\theta'))_{\theta' \in \Theta}$ for all $\theta \in \Theta$.

It follows from definition (2) that the wage $w(\theta)$ of type $\theta$ can be represented as a functional $\omega: \Theta \times \mathbb{R}^+ \times \mathcal{M} \to \mathbb{R}^+$ that has three arguments: the individual’s type $\theta \in \Theta$, the labor supply $L(\theta) \in \mathbb{R}^+$ that we perturb, and the measure $\mathcal{L} \in \mathcal{M}$ which describes each agent’s labor effort:

$$w(\theta) = \omega(\theta, L(\theta), \mathcal{L}). \quad (3)$$

We make the following assumption throughout the paper:

**Assumption 1.** The wage and earnings functions $\theta \mapsto w(\theta)$ and $\theta \mapsto y(\theta)$ are strictly increasing and differentiable.

That is, individuals with higher productivities earn a higher wage and a higher income. This monotone relationship allows us to define the densities of wages and incomes as

$$f_\theta(\theta) = f_w(w(\theta))w'(\theta) = f_y(y(\theta))y'(\theta).$$

with c.d.f. $F_w(\cdot)$ and $F_y(\cdot)$ respectively.

### 1.2 Social welfare

The government chooses the income tax schedule $T(\cdot)$ and evaluates social welfare according to a weighted utilitarian social objective. Denote by $\tilde{f}_\theta(\cdot)$ the probability density function (i.e., the schedule of Pareto weights, normalized such that $\int_{\Theta} \tilde{f}_\theta(x) \, dx = 1$) that the government uses to weigh individual indirect utilities in the social objective. The case $\tilde{f}_\theta(\theta) = f_\theta(\theta)$ for all $\theta \in \Theta$ corresponds to a utilitarian planner. Social welfare is then defined by:

$$G \equiv \int_{\Theta} u[\ell(\theta) - T(w(\theta)l(\theta)) - v(l(\theta))] \tilde{f}_\theta(\theta) \, d\theta, \quad (4)$$

Denote by $\lambda$ the marginal value of public funds. We define the social marginal welfare weight

---

7We verify numerically ex post that this assumption is satisfied.

8We do not take a stand on how the marginal value of public funds is determined, which ensures the generality of the government’s problem we analyze. In particular, $\lambda$ can be defined exogenously as follows. Suppose that there is a public good which brings social utility $V(\cdot)$; then $\lambda = V'(\cdot)$ is the social value of marginally increasing the funding of the public good. Alternatively, $\lambda$ can be defined endogenously by imposing that all the perturbations of the tax system that we consider are revenue neutral, i.e., satisfy $\int_{\Theta} T[w(\theta)l(\theta)] f_\theta(\theta) \, d\theta \geq 0$. For instance, if lump-sum transfers are used by the government to redistribute the excess tax revenue (or loss) from local tax reforms, then $\lambda$ is equal to the change in the value of the social objective (4) in response to a $1$ lump-sum transfer uniformly.
(see, e.g., Saez and Stantcheva (2016)) associated with individuals of type \( \theta \) as

\[
g(\theta) = \frac{u'[c(\theta) - v(l(\theta))]f_\theta(\theta)\bar{\lambda}f_\theta(\theta)}{\lambda f_\theta(\theta)}
\]

which summarizes the government’s redistributive preferences.

### 1.3 Elasticity concepts

#### Partial equilibrium labor supply elasticities

We define several labor supply elasticity concepts. Consider an individual with type \( \theta \) and income \( y = y(\theta) \). Denote her marginal tax rate by \( \tau(\theta) = T'(y(\theta)) \).

We first define the labor supply elasticity with respect to the marginal tax rate along the linearized budget constraint as:

\[
\varepsilon_{l,1-\tau}(\theta) = \frac{\partial \ln l(\theta)}{\partial \ln (1 - \tau(\theta))} \bigg|_{w(\theta)} = \frac{u'(l(\theta))}{l(\theta)v''(l(\theta))}, \quad (5)
\]

where the second equality is proved in the Appendix. Note that this is a partial equilibrium elasticity, since the change in labor supply is computed keeping the individual wage constant. Assumption 1 allows us to write interchangeably \( \varepsilon_{l,1-\tau}(\theta) \) or \( \varepsilon_{l,1-\tau}(y(\theta)) \) depending on the context, and similarly for all the elasticities that we encounter.

Second, we define the elasticity along the nonlinear budget constraint, as the total labor supply response to a perturbation of the marginal tax rates, keeping wages constant. This total response includes an indirect effect generated by the nonlinearity of the tax schedule: as an individual changes her labor supply by \( dl \), the marginal tax rate she faces changes by \( dT'(wl) = T''(wl) wdl \), which induces a further labor supply response determined by the elasticity \( \varepsilon_{l,1-\tau} \). The following elasticity solves for this fixed point and writes:

\[
\tilde{\varepsilon}_{l,1-\tau}(\theta) = \frac{d\ln l(\theta)}{d\ln (1 - \tau(\theta))} \bigg|_{w(\theta)} = \frac{1 - T'(y(\theta))}{1 - T'(y(\theta)) + \varepsilon_{l,1-\tau}(\theta) y(\theta) T''(y(\theta))} \varepsilon_{l,1-\tau}(\theta), \quad (6)
\]

where the second equality is proved in the Appendix.

Finally, we define the labor supply elasticities with respect to a change in the wage, along the nonlinear budget constraint, as the total labor supply response to an exogenous perturbation of the wage rate, but still ignoring the general equilibrium wage effects induced by this initial response. That is,

\[
\tilde{\varepsilon}_{l,w}(\theta) = \frac{d\ln l(\theta)}{d\ln w(\theta)} \bigg|_{w(\theta)} = \frac{1 - T'(y(\theta)) - y(\theta) T''(y(\theta))}{1 - T'(y(\theta)) + \varepsilon_{l,1-\tau}(\theta) y(\theta) T''(y(\theta))} \varepsilon_{l,1-\tau}(\theta), \quad (7)
\]

distributed among all individuals in the economy (which does not entail any labor supply responses since the utility function has no income effects).

---

9See, e.g., Saez (2001), Golosov et al. (2014).

where the second equality is proved in the Appendix.

**Wage elasticities**

We now define the elasticities that are specific to the general equilibrium framework. For any pair \((\theta, \theta') \in \Theta^2\), define the wage elasticity \(\gamma(\theta, \theta')\) as the effect of a marginal increase in the labor supply of type \(\theta'\), \(L(\theta')\), on the wage of type \(\theta\), \(w(\theta)\). Start by defining these elasticities heuristically: intuitively, \(\gamma(\theta, \theta')\) is the derivative of \(\ln w(\theta)\) with respect to \(\ln L(\theta')\), that is,

\[
\gamma(\theta, \theta') = \frac{\partial \ln w(\theta)}{\partial \ln L(\theta')} = \frac{L(\theta') \mathcal{F}'_{\theta,\theta'}(\mathcal{L})}{\mathcal{F}'_{\theta}(\mathcal{L})}.
\]

Note that there are two cases to consider, namely the *cross-wage elasticity*, if \(\theta \neq \theta'\), and the *own-wage elasticity*, if \(\theta = \theta'\). Specifically, denote by \(\bar{\gamma}(\theta, \theta') = \gamma(\theta, \theta')\), for all \(\theta \neq \theta'\), the impact of the labor supply of type \(\theta'\) on the wage of another type \(\theta\). We then complete the definition of \(\bar{\gamma}(\theta, \theta')\) on \(\Theta\) by letting \(\bar{\gamma}(\theta', \theta') = \lim_{\theta \to \theta'} \bar{\gamma}(\theta, \theta')\), and we assume that the function \(\theta \mapsto \bar{\gamma}(\theta, \theta')\) is continuously differentiable. Now, in general we have \(\gamma(\theta', \theta') \neq \bar{\gamma}(\theta', \theta')\): the function \(\theta \mapsto \gamma(\theta, \theta')\) “jumps” when \(\theta = \theta'\), i.e., when we consider the impact of the labor supply of type-\(\theta'\) on their own wage. We thus define \(\bar{\bar{\gamma}}(\theta', \theta')\) as the “size” of this jump, i.e., the difference between \(\gamma(\theta', \theta')\) and \(\bar{\gamma}(\theta', \theta')\).

Formally, \(\gamma(\theta, \theta')\) is defined as the Gateaux derivative of the functional \(\omega(\theta, L(\theta), \mathcal{L})\) (introduced in (3)), when \(\mathcal{L}\) is perturbed in the Dirac direction \(\delta_{\theta'}\). If \(\theta' \neq \theta\), the labor supply \(L(\theta')\) appears only through \(\mathcal{L}\) in the arguments of \(\omega(\theta, L(\theta), \mathcal{L})\), while it also appears directly as the second argument of the functional if \(\theta' = \theta\). Thus we have the following definition:

**Definition 1.** Define the cross-wage elasticity as

\[
\bar{\gamma}(\theta, \theta') = \lim_{\mu \to 0} \frac{1}{\mu} \{\ln \omega(\theta, L(\theta), \mathcal{L} + \mu \delta_{\theta'}) - \ln \omega(\theta, L(\theta), \mathcal{L})\}, \tag{8}
\]

and the own-wage elasticity as

\[
\bar{\bar{\gamma}}(\theta, \theta) = \frac{\partial \ln \omega(\theta, L(\theta), \mathcal{L})}{\partial \ln L(\theta)}. \tag{9}
\]

The wage elasticity \(\gamma(\theta, \theta')\), for any \((\theta, \theta')\), is then defined as

\[
\gamma(\theta, \theta') = \bar{\gamma}(\theta, \theta') + \bar{\bar{\gamma}}(\theta', \theta') \delta_{\theta'}(\theta). \tag{10}
\]

Note that (10) is an equality of generalized functions, since the right hand side (and hence the left hand side as well) has a Dirac term for \(\theta = \theta'\).

**General equilibrium labor supply elasticities**

If wages are endogenous, the partial equilibrium labor supply elasticities (5), (6), and (7) do not directly inform us about the effects of a tax change on labor supply. As soon as one individual (say,
θ) adjusts her labor supply, the wages of all other types θ' ≠ θ are affected, which leads them in turn to shift their labor supply. This then feeds back into the wage of every individual θ'' ∈ Θ, thus triggering an infinite adjustment process. We study in detail the total labor supply effects of this chain of responses in Section 2.

At this point, we introduce for notational simplicity a simple version of general equilibrium elasticities, which capture the response of type-θ labor supply to changes in the net of tax rate $1 - \tau(\theta)$ and the wage $w(\theta)$, taking into account the endogeneity of the wage $w(\theta)$ to its own labor supply $L(\theta)$, but keeping everyone else’s labor supply fixed. Specifically, we define:

$$\hat{E}_{l,1-\tau}(\theta) = \frac{\hat{\varepsilon}_{l,1-\tau}(\theta)}{1 - \hat{\gamma}(\theta,\theta)\hat{\varepsilon}_{l,w}(\theta)}, \quad \text{and} \quad \hat{E}_{l,w}(\theta) = \frac{\hat{\varepsilon}_{l,w}(\theta)}{1 - \hat{\gamma}(\theta,\theta)\hat{\varepsilon}_{l,w}(\theta)}.$$ (11)

Intuitively, an increase in the labor supply of type-θ individuals, triggered by an increase in (say) their net of tax rate (and measured by the labor supply elasticity $\hat{\varepsilon}_{l,1-\tau}(\theta)$), induces a decrease in their own wage (measured by the own-wage elasticity $\hat{\gamma}(\theta,\theta)$), and hence a further decrease in their labor supply (measured by the elasticity $\hat{\varepsilon}_{l,w}(\theta)$). Solving for this fixed point, we can write

$$d \ln l(\theta) = \hat{\varepsilon}_{l,1-\tau}(\theta) + \hat{\varepsilon}_{l,w}(\theta) \hat{\gamma}(\theta,\theta) d \ln l(\theta),$$

which yields the expression for $\hat{E}_{l,1-\tau}(\theta)$ in (11) (the expression for $\hat{E}_{l,w}(\theta)$ is derived identically). Thus $\hat{E}_{l,1-\tau}(\theta)$ and $\hat{E}_{l,w}(\theta)$ measure the effects of perturbing tax rates and wages on the labor supply of type-θ agents, taking into account the dampening general equilibrium effect that their behavior induces on their own wage.

1.4 Examples

In order to illustrate the concepts introduced so far, we now derive expressions of the wages and elasticities for two of the most common production functions used in the literature, namely the CES and the Translog technologies. In the Appendix we define other useful production functions and derive their corresponding properties.

Example 1. (CES technology.) The production function has constant elasticity of substitution (CES) if

$$\mathcal{F} \left( \{ L(\theta) \}_{\theta \in \Theta} \right) = \left[ \int_{\Theta} a(\theta) (L(\theta))^\rho \, d\theta \right]^{1/\rho},$$

for some constant $\rho \in (-\infty, 1]$. The wage schedule is given by

$$w(\theta) = a(\theta) (L(\theta))^{\rho-1} \left[ \int_{\Theta} a(x) (L(x))^\rho \, dx \right]^{\frac{1}{\rho}-1},$$

which highlights the dependence of the wage functional on the three variables $\theta \in \Theta$, $L(\theta) \in \mathbb{R}_+$,
and the measure $\mathcal{L} = \{ L(\theta') \}_{\theta' \in \Theta} \in \mathcal{M}$. The cross wage elasticities are given by

$$\bar{\gamma}(\theta, \theta') = (1 - \rho) \frac{\alpha(\theta') L(\theta')}{\int_{\Theta} \alpha(x) L(x)^{\rho} dx}, \forall \theta, \theta' \in \Theta,$$

and the own-wage elasticities are given by

$$\bar{\gamma}(\theta, \theta) = \rho - 1, \forall \theta \in \Theta.$$

Note in particular that $\bar{\gamma}(\theta, \theta) < 0$ is constant, and that $\bar{\gamma}(\theta, \theta') > 0$ does not depend on $\theta$, implying that a change in the labor supply of type $\theta'$ has the same effect (in percentage terms) on the wage of every other type $\theta \neq \theta'$. Finally the elasticity of substitution between any two labor inputs $(L(\theta), L(\theta'))$, defined as the effect on the relative wage of types $\theta$ and $\theta'$ of an increase in their relative supply of labor, is constant and equal to

$$\sigma(\theta, \theta') = -\frac{\partial \ln (w(\theta)/w(\theta'))}{\partial \ln (L(\theta)/L(\theta'))} = -\frac{1}{\rho - 1}.$$

We denote by $\sigma$ this constant elasticity of substitution whenever we specify a CES production function. The cases $\sigma = 1$ and $\sigma = 0$ correspond respectively to the Cobb-Douglas and Leontieff production functions.

The second main example considered in the literature is the Translog production function, which can be used as a second-order approximation to any production function (see details in Christensen et al. (1973)). We now formally define this production function for our continuous input setting.

**Example 2. (Translog technology.)** The production function is transcendental logarithmic (Translog) if

$$\ln \mathcal{F}(\{ L(\theta) \}_{\theta \in \Theta}) = \alpha_0 + \int_{\Theta} \alpha_\theta \ln L(\theta) d\theta + \frac{1}{2} \int_{\Theta} \bar{\beta}_{\theta, \theta} \ln L(\theta) \ln L(\theta) \, d\theta + \frac{1}{2} \int_{\Theta \times \Theta} \tilde{\beta}_{\theta, \theta'} (\ln L(\theta)) (\ln L(\theta')) \, d\theta d\theta', \tag{13}$$

where $\bar{\beta}_{\theta, \theta'} = \tilde{\beta}_{\theta', \theta}$ and

$$\int_{\Theta} \alpha_\theta d\theta' = 1, \text{ and } \bar{\beta}_{\theta, \theta} = \int_{\Theta} \tilde{\beta}_{\theta, \theta'} d\theta', \forall \theta, \theta' \in \Theta. \tag{14}$$

These restrictions ensure that the technology has constant returns to scale. When $\bar{\beta}_{\theta, \theta'} = 0$ for all $\theta, \theta'$, the production function is Cobb-Douglas.

The wage schedule is given by

$$w(\theta) = \frac{\mathcal{F}(\mathcal{L})}{L(\theta)} \left\{ \alpha_\theta + \bar{\beta}_{\theta, \theta} \ln L(\theta) + \int_{\Theta} \tilde{\beta}_{\theta, \theta'} \ln L(\theta') \, d\theta' \right\}.$$
The cross-wage elasticities are given by
\[
\bar{\gamma}(\theta, \theta') = \chi_{\theta'} + \frac{\bar{\beta}_{\theta, \theta'}}{\chi_\theta}, \quad \forall \theta, \theta' \in \Theta,
\] (15)
where for all \( \theta, \chi_\theta = \frac{w(\theta)L(\theta)}{F(L)} \) denotes the type-\( \theta \) labor share of output. The own-wage elasticities are given by
\[
\bar{\gamma}(\theta, \theta) = -1 + \frac{\bar{\beta}_{\theta, \theta}}{\chi_\theta}, \quad \forall \theta \in \Theta.
\]
Finally, the elasticities of substitution are given by
\[
\sigma(\theta, \theta') = 1 + \left[1 + \frac{1}{\chi_\theta} + \frac{1}{\chi_{\theta'}}\right] \bar{\beta}_{\theta, \theta'},
\]
(All the derivations are in the Appendix.)

We can obtain further results about the (non-constant) substitution elasticities derived from the Translog production function by specifying the exogenous coefficients \( \bar{\beta}_{\theta, \theta'} \) of formula (13). We propose to set the following structure. Suppose that the coefficients \( \bar{\beta}_{\theta, \theta'} \) form a bivariate lognormal distribution with correlation coefficient \( \rho \), i.e.,
\[
\bar{\beta}_{\theta, \theta'} = \frac{1}{\theta\theta'2\pi s^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(\ln \theta - m)^2}{s^2} + \frac{(\ln \theta' - m)^2}{s^2} - 2\rho (\ln \theta - m)(\ln \theta' - m)\right]\right)
\equiv \hat{\phi}_{m,s,\rho}(\theta, \theta'),
\] (16)
which implies
\[
\bar{\beta}_{\theta, \theta} = \frac{1}{\theta\sqrt{2\pi s^2}} \exp\left(-\frac{1}{2} \left[\frac{(\ln \theta - m)^2}{s^2}\right]\right) \equiv \hat{\phi}_{m,s}(\theta),
\]
where \( \hat{\phi}_{m,s} \) is the density of a log-normal distribution with parameters \((m, s^2)\), and \( \hat{\phi}_{m,s,\rho} \) is the corresponding bivariate lognormal density. Evaluating the wage elasticities \( \gamma(\theta, \theta') \) to a first-order Taylor approximation as \( \rho \to 0 \) yields:

**Lemma 1.** Suppose that the production function is Translog with coefficients \( \bar{\beta}_{\theta, \theta'} \) given by (16). A first-order Taylor expansion as \( \rho \to 0 \) yields the following distance-dependent wage elasticities:
\[
\bar{\gamma}(\theta, \theta') = \rho \to 0 \chi_{\theta'} + \chi_{\theta}^{-1} \left[1 - \frac{\rho}{2s^2} (\ln \theta - \ln \theta')^2\right] \hat{\phi}_{m,s,0}(\theta, \theta'),
\] (17)
and
\[
\bar{\gamma}(\theta, \theta) = \rho \to 0 -1 + \chi_{\theta}^{-1} \hat{\phi}_{m,s}(\theta),
\]
where \( \hat{\phi}_{m,s,0}(\theta, \theta') = \hat{\phi}_{m,s}(\theta) \hat{\phi}_{m,s}(\theta') \) denotes the product of two independent lognormal distributions with the same mean and variance.

**Proof.** See Appendix. □
Expression (17) shows that the cross-wage elasticities $\bar{\gamma}(\theta, \theta')$ are decreasing in the distance between the types $\theta$ and $\theta'$, measured by $(\ln \theta - \ln \theta')^2$. This implies that the smaller the distance between $\theta$ and $\theta'$, the stronger substitutes these types are. This is in contrast with the CES case, where the labor supply of a given type $\theta'$ affects all types $\theta$ identically. Moreover, a given increase in the distance between types $\theta$ and $\theta'$ reduces their substitutability by a larger amount when the correlation $\rho$ is larger and the variance $\sigma$ is lower. Note in passing that the proposed specification is a (minor) contribution of the paper. The functional form we use allows to tractably capture the distance-dependence of the elasticity of substitution in the continuous input setting and thus generalizes the CES case.\footnote{This is reminiscent of Teulings (1995)’s assignment model.}

2 General tax incidence analysis

In this section we derive the first-order effects of arbitrary local tax reforms on: (i) individual labor supplies, wages, and utilities, and (ii) government revenue and social welfare. The crucial step, and technical insight, consists in showing that the problem of analyzing the economic incidence of taxes reduces mathematically to solving an integral equation. Having established this, we can use the tools of the well-developed mathematical theory of integral equations to study tax incidence. We first derive the formulation of this integral equation and provide the analytical solution for a general production function in Sections 2.1 and 2.2. We show that the general solution to this equation, expressed in terms of the resolvent kernel, has a straightforward economic interpretation, capturing the successive rounds of general equilibrium feedback effects. We then show that for specific functional forms of the technology, in particular CES (Section 2.3) and Translog (Section 2.4), the solution admits particularly tractable and insightful closed forms. In the case of a CES production function, we further specify the baseline tax schedule to a two-parameter functional form that closely approximates the U.S. tax system (following Heathcote et al. (2014)) and derive an operational theoretical expression for the general equilibrium incidence of any local tax reform of the current tax code.

2.1 Effects of tax reforms on labor supply

The aim of this section and the next is to derive the first-order effects on individual and aggregate behavior of arbitrary local perturbations (“tax reforms”) of a given baseline tax schedule. As in partial equilibrium (Saez, 2001), the crucial part of the analysis of this general incidence problem consists of solving for each individual’s labor supply change in terms of behavioral elasticities. This problem is, however, much more involved in the general equilibrium setting. Once this step is solved, the effects on individual wages, utilities, government revenue, and social welfare can be easily derived. We address the former question in this section, and the latter in Section 2.2.

In partial equilibrium, the effects of a tax reform on the labor supply of a given agent, say $\theta$, can be straightforwardly derived as a function of the elasticity of labor supply (5) (see Saez (2001)).
The simplicity of the solution to this fixed point problem comes from the fact that a change in the tax rate of a given individual involves only a change in labor effort from that type, without any effects on the other agents. Instead, the key difficulty in general equilibrium is that this direct (“partial equilibrium”) effect of the perturbation on the labor supply of individual $\theta$ affects the wage, and thus the labor supply, of every other agent $\theta' \neq \theta$. This in turn feeds back into the wage of $\theta$, which further impacts her labor supply, and so on. Representing the total effect of this infinite chain reaction triggered by the reform, coupled with the fact that both the baseline tax schedule and the tax reform itself are allowed to be arbitrarily nonlinear, is thus a complicated task.

Note that one way to side-step this problem would be to define, for each specific tax reform, a “policy elasticity” (as in Hendren (2015), Piketty and Saez (2013)), as each individual’s total labor supply response to the corresponding tax reform. It would then be straightforward to derive the effects of the same tax reform on, say, social welfare, as a function of this policy elasticity variable. However, the key difficulty of the incidence problem consists precisely of expressing the labor supply response in terms of the structural elasticity parameters introduced in Section 1.3. Doing so allows us to uncover the fundamental economic forces underlying the incidence of taxation. It also allows us to represent the effects of any tax reform in terms of a common set of structural elasticity parameters, without requiring the case-by-case evaluation of the labor supply effects of each specific reform that one might consider implementing in practice.

Formally, consider an arbitrary perturbation $h (\cdot) \in C^1 [\mathbb{R}_+]$ of a baseline, potentially suboptimal, tax schedule $T (\cdot)$. Our aim is to compute the Gateaux derivative of individual $\theta$’s labor supply functional $l_\theta : T \mapsto l (\theta; T)$ in the direction $h$, i.e.,

$$
\frac{d\hat{l} (\theta, h)}{d\theta} = \lim_{\mu \to 0} \frac{1}{\mu} \left \{ \ln l_\theta (T + \mu h) - \ln l_\theta (T) \right \}.
$$

(18)

This expression captures the total labor supply change in response to the tax reform $h$, taking into account all of the feedback effects due to the endogeneity of wages. Our aim is to derive the individual labor supply effects of the tax reform $h$ in terms of the structural labor supply and wage elasticities. We now show how this a priori complex problem reduces mathematically to solving an integral equation, i.e., an equation where the unknown is a function that appears under an integral sign.

Lemma 2. The effect of a perturbation $h$ of the baseline tax schedule $T$ on individual labor supply, $d\hat{l} (\cdot, h)$, is the solution to the following integral equation:

$$
d\hat{l} (\theta, h) = -\tilde{E}_{l,1,-\tau} (\theta) \frac{h' (y (\theta))}{1 - T' (y (\theta))} + \tilde{E}_{l,w} (\theta) \int_{\Theta} \tilde{\gamma} (\theta, \theta') d\hat{l} (\theta', h) d\theta',
$$

(19)

for all $\theta \in \Theta$.

Proof. See Appendix.

---

12The notation $d\hat{l} (\theta, h)$ ignores for simplicity the dependence of the Gateaux derivative on the baseline tax schedule $T$. 

Formula (19) is a Fredholm integral equation of the second kind, i.e., an equation of the form
\[ \varphi(\theta) = \psi(\theta) + \int_{\Theta} K(\theta, \theta') \varphi(\theta') \, d\theta', \]
where \( \varphi(\theta) = d\hat{l}(\theta, h) \) is the unknown function, \( \psi(\theta) \propto h'(y(\theta)) \) is a known function, and \( K_1(\theta, \theta') = \tilde{E}_{l,w}(\theta) \tilde{\gamma}(\theta, \theta') \) is the kernel of the integral equation (see, e.g., Zemyan (2012)). The key technical implication of this lemma is that the general analysis of the incidence of taxation, with a continuum of labor inputs and arbitrary nonlinear taxes, can be mathematically formulated as the problem of characterizing the solution to a linear integral equation (19).

We now sketch the proof of equation (19). The net-of-tax rate faced by individual \( \theta \) changes, in percentage terms, by \(-\frac{h'(y(\theta))}{1 - T'(y(\theta))}\) due to the tax reform. By definition of the elasticity of labor supply along the nonlinear tax schedule (6), this tax change induces a direct percentage change in labor supply \( l(\theta) \) equal to \(-\tilde{\varepsilon}_{l,1-\tau}(\theta) \frac{h'}{1 - T'} \). This is the expression we would obtain in the partial equilibrium setting (see p. 217 in Saez (2001)). In particular, in the partial equilibrium environment, the effects of tax reforms on labor supply are given by a degenerate case of the general integral equation (19), in which the kernel of the integral is uniformly zero. Now, in general equilibrium, type-\( \theta \) labor supply is also impacted indirectly by the change in all other individuals’ labor supplies. Specifically, for all \( \theta' \in \Theta \), the change in labor supply of type \( \theta' \), \( d\ln l_{\theta'} \), induces a change in the wage of type \( \theta \) equal to \( \gamma(\theta, \theta') \times d\ln l_{\theta'} \), and hence in turn a change in the labor supply of type \( \theta \) equal to \( \tilde{\varepsilon}_{l,w}(\theta) \gamma(\theta, \theta') \, d\ln l_{\theta'} \). Summing the direct effect of the tax change, and all of these indirect wage effects, we obtain that type \( \theta \) changes her labor supply by
\[
d\hat{l}(\theta, h) = -\tilde{\varepsilon}_{l,1-\tau}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta))} + \tilde{E}_{l,w}(\theta) \int_{\Theta} \gamma(\theta, \theta') \, d\hat{l}(\theta, h) \, d\theta'.
\]

Finally, disentangling the own- and cross-wage effects in the previous integral (using equation (10)), we can easily rewrite this integral equation in the form (19), where the labor supply elasticities \( \tilde{\varepsilon}_{l,1-\tau}(\theta) \) and \( \tilde{\varepsilon}_{l,w}(\theta) \) are now weighted by the own-wage GE effects \( \tilde{\gamma}(\theta, \theta) \tilde{E}_{l,1-\tau}(\theta) \) to yield \( \tilde{E}_{l,w}(\theta) \) and \( \tilde{E}_{l,w}(\theta) \) (defined in (11)). The kernel of the integral equation, which features only the cross-wage effects \( \tilde{\gamma}(\theta, \theta') \), is now smooth.

Since Lemma 2 reduces the tax incidence problem to characterizing the solution to an integral equation, the next step of our analysis consists of solving for \( d\hat{l}(\theta, h) \). The mathematical theory of integral equations is well developed (see, e.g., Tricomi (1985), Kress (2014), and, for a concise introduction, Zemyan (2012)). In particular, the resolvent formalism technique allows us to derive a general analytic representation of the solution to (19); we do so in Proposition 1 below. Moreover, we can derive simpler closed-form solutions in many cases (see Polyanin and Manzhirov (2008)), which becomes particularly handy when we specify functional forms for the production function (see Sections 2.3 and 2.4 below). Finally, numerical techniques are widely available and can be easily implemented (see, e.g., Press (2007)), leading to straightforward quantitative simulations of the incidence of arbitrary tax reforms.

**Proposition 1.** Assume that \( \int_{\Theta} |K_1(\theta, \theta')|^2 \, d\theta d\theta' < 1 \), where \( K_1(\theta, \theta') = \tilde{E}_{l,w}(\theta) \tilde{\gamma}(\theta, \theta') \).\(^{13}\) The

---

\(^{13}\)This technical condition ensures that the infinite series (21) converges and can be verified numerically. When
The solution to the integral equation (19) is given by

\[ d\hat{l}(\theta, h) = -\tilde{E}_{l,1-\tau}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta))} - \int_{\Theta} \mathcal{R}(\theta, \theta') \hat{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} d\theta', \]  

(20)

where for all \( \theta, \theta' \in \Theta \), the resolvent kernel \( \mathcal{R}(\theta, \theta') \) is given by the Neumann series

\[ \mathcal{R}(\theta, \theta') = \sum_{n=1}^{\infty} K_n(\theta, \theta'), \]  

(21)

in which the \((n+1)\)-th iterated kernel given, for all \( n \geq 1 \), by

\[ K_{n+1}(\theta, \theta') = \int_{\Theta} K_n(\theta, \theta'') K_1(\theta'', \theta') d\theta''. \]

Moreover, the solution is unique.

**Proof.** See Appendix.

Equation (20) is the general solution to the integral equation (19). It is given as a function of the resolvent kernel \( \mathcal{R}(\theta, \theta') \) of the integral equation, which is itself given in the form of an infinite series of the iterated kernels \( K_n(\theta, \theta') \). While the general solution may seem complex, in practice there is a large literature on the computation of such solutions.\(^\text{14}\)

We show that this abstract mathematical representation of the solution to (19) has a very clear economic meaning. The first term in the right hand side of (20) is the direct (“partial equilibrium”) effect of the tax reform on labor supply \( l(\theta) \), already described above for equation (19). The second term captures the infinite sequence of general equilibrium effects between the labor supplies of types \( \theta \) and \( \theta' \): each term in the series represents one round of feedback effects occurring through wages. Consider for example the first iterated kernel, \( n = 1 \), which writes

\[ \int_{\Theta} K_1(\theta, \theta') \hat{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} d\theta'. \]

This integral expresses the fact that for any \( \theta' \), the percentage change in the labor supply of \( \theta' \) due to the tax reform (including the own-wage effects), \( \hat{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} \), induces a change in the wage of type \( \theta \) given by \( \hat{\gamma}(\theta, \theta') \), and hence a change in her labor supply given by \( K_1(\theta, \theta') = \hat{E}_{l,w}(\theta) \hat{\gamma}(\theta, \theta') \). The integrand thus accounts for the effect of a given type \( \theta' \) on the labor supply of \( \theta \) through direct cross-wage effects.

Next, the second iterated kernel, \( n = 2 \), accounts for the effects of \( \theta' \) on the labor supply of \( \theta \), indirectly through the behavior of third parties \( \theta'' \); that is, \( \theta' \) affects \( \theta'' \), which in turn affects \( \theta \).

---

\(^{14}\)For example, numerical methods are readily available (see e.g. Press (2007) and Section 2.6 in Zemyan (2012) and the references therein).
This term writes

\[
\int_{\Theta} \left\{ \int_{\Theta} K_1(\theta, \theta') K_1(\theta'', \theta') d\theta'' \right\} \tilde{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} d\theta'.
\]

This integral expresses the fact that for any \( \theta' \), the percentage change in the labor supply of \( \theta' \) due to the tax reform, \( \tilde{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} \), induces a change in the wage of all other types \( \theta'' \) given by \( \gamma(\theta'', \theta') \), and hence a change in the labor supply of \( \theta'' \) given by \( K_1(\theta'', \theta') = \tilde{E}_{l,w}(\theta'') \gamma(\theta'', \theta') \). This in turn affects the labor supply of type \( \theta \) (through the cross-wage effects \( \gamma(\theta, \theta'') \)) by the amount \( K_1(\theta, \theta'') = \tilde{E}_{l,w}(\theta) \gamma(\theta, \theta'') \).

An inductive reasoning shows similarly that the terms \( n \geq 3 \) account for the effects of \( \theta' \) on the labor supply of \( \theta \) through \( n \) stages of cross-wage effects, e.g. (for \( n = 3 \)), \( \theta' \rightarrow \theta'' \rightarrow \theta''' \rightarrow \theta \).

### 2.2 Effects of tax reforms on wages and welfare

We now turn to the incidence analysis of a tax reform \( h \) on individual wages and utilities (Corollary 1), and on government revenue and social welfare (Corollary 2). The corresponding Gateaux derivatives \( d\tilde{w}(\theta, h), d\tilde{U}(\theta, h), d\tilde{R}(T, h), d\tilde{W}(T, h) \) are defined as in (18). The results below show that, having characterized the solution to the integral equation (19), all the other incidence effects of a tax reform can be easily derived as a function of the variable \( d\tilde{l}(\theta, h) \) derived in (20).

**Corollary 1.** The first-order effects of a perturbation \( h \) of the baseline tax schedule \( T \) on wages are given by

\[
d\tilde{w}(\theta, h) = \tilde{e}_{l,1-\tau}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta))} + d\tilde{l}(\theta, h). \tag{22}
\]

The first-order effects of a perturbation \( h \) of the baseline tax schedule \( T \) on individual utilities are given by

\[
d\tilde{U}(\theta, h) = - u'(\theta) \left[ h(y(\theta)) + (1 - T'(y(\theta))) y(\theta) d\tilde{w}(\theta, h) \right]. \tag{23}
\]

**Proof.** See Appendix. \( \square \)

Equation (22) shows that the effect of the reform \( h \) on the wage \( w(\theta) \) is given by a simple transformation of the effect on labor supply, \( d\tilde{l}(\theta, h) \). To understand this expression, note that we can rewrite (22) as

\[
d\tilde{l}(\theta, h) = -\tilde{e}_{l,1-\tau}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta))} + \tilde{e}_{l,w}(\theta) d\tilde{w}(\theta, h).
\]

This expression is very intuitive. The left hand side is the percentage change in \( \theta \)'s labor supply due to the tax reform \( h \). The right hand side expresses this change as the direct effect on \( l(\theta) \) of the percentage perturbation of marginal tax rates \( \frac{h'}{1 - T'} \) (through the elasticity \( \tilde{e}_{l,1-\tau}(\theta) \)), plus the indirect (general equilibrium) effect through the change in the wage \( d\tilde{w}(\theta, h) \) (via the elasticity \( \tilde{e}_{l,w}(\theta) \)).
Equation (23) characterizes the impact of the tax reform on indirect utilities. The first term, \(-u'(\theta)h(y(\theta))\), is due to the fact that the tax increase \(h(y(\theta))\) makes the individual poorer, which reduces her welfare by the marginal utility \(u'(\theta)\). The second term in the r.h.s. of (23) accounts for the change in consumption, and hence welfare, due to the wage change.\(^{15}\) By the envelope theorem, the change in labor supply due to the tax reform has a zero first-order impact on the individual’s utility.

We now derive the incidence of the tax reform \(h\) on government revenue and social welfare.

**Corollary 2.** The first-order effect of a perturbation \(h\) of the baseline tax schedule \(T\) on tax revenue is given by

\[
dR(T,h) = \int_R h(y) f_y(y) \, dy - \int_R \frac{T'(y)}{1 - T'(y)} \left[ d\tilde{l}(\theta, h) + d\hat{w}(\theta, h) \right] y f_y(y) \, dy
\]

and the first-order effect on social welfare (measured in terms of public funds) is given by

\[
W(T,h) = R(T,h) + \int_R g(y) \left[ -h(y) + yd\hat{w}(\theta, h) \right] f_y(y) \, dy.
\]

**Proof.** See Appendix. \(\square\)

The interpretation of equation (24) is as follows. The tax reform \(h\) induces two effects on government revenue. The first term in the right hand side is the mechanical effect of the perturbation, i.e., the statutory increase in tax revenue for a given individual behavior. The second term in the right hand side of (24) describes the impact of labor supply and wage changes on government revenue. In partial equilibrium, the change in labor supply of type \(\theta\) would be equal to the change in marginal tax rates \(h'(y(\theta))\) times the labor supply elasticity \(\tilde{\epsilon}_{l,1-\tau}(\theta)\), i.e.,

\[
d\tilde{l}_{PE}(\theta, h) = -\tilde{\epsilon}_{l,1-\tau}(\theta) h'(y(\theta)).
\]

This labor supply change would in turn impact government revenue by an amount proportional to the marginal tax rate at income \(y, T'(y(\theta))\).

In general equilibrium, there are two novel effects. The first is that the behavioral change in labor supply of type \(\theta\), \(d\tilde{l}(\theta, h)\), is generally different from its partial equilibrium counterpart \(d\tilde{l}_{PE}(\theta, h)\), since we saw that it includes an infinite sequence of the feedback effects through wages (see equation (20)). Suppose in particular, as in Saez (2001), that the tax reform \(h\) increases the marginal tax rate at income level \(y^*\) only, and thus increases the total tax liability by a uniform lump sum amount for incomes \(y > y^*\).\(^{16}\) In this case the elasticities \(d\tilde{l}(\theta, \mathbb{I}_{\{y \geq y^*\}})\) are non-zero for all \(\theta \in \Theta\), while in partial equilibrium the second integral in (20) would reduce to the value of the integrand at \(y^*\), so that \(d\tilde{l}_{PE}(\theta, \mathbb{I}_{\{y \geq y^*\}})\) would be equal to zero for all \(\theta \neq \theta^*\). Finally, the second difference is

\(^{15}\)Note that using (22), we can express the incidence of the tax reform on utilities solely in terms of the labor supply policy elasticity \(d\tilde{l}(\theta, h)\).

\(^{16}\)Formally, \(h'\) is the Dirac delta function at \(y^*\), \(h'(y) = \delta_{y^*}(y)\), which is equal to zero for all \(y \neq y^*\), and \(h\) is the step function \(h(y) = \mathbb{I}_{\{y \geq y^*\}}\).
that government revenue is now also impacted directly by the changes in wages, captured by the general equilibrium wage effects \( d \hat{w}(\theta, h) \). Summing these effects over all individuals, weighted by the density of incomes \( f_y(y) \), yields equation (25).

Equation (25) describes the effect of the tax reform \( h \) on social welfare, expressed in terms of government funds. It consists of the change in tax revenue \( d R(T, h) \), plus the effects on individual utilities of perturbing taxes, described in equation (23). Note that the marginal utilities are weighted by the shadow value of public funds to obtain a monetary measure of welfare, leading to the marginal social welfare weights \( g(y) \).

### 2.3 CES production function

The theory of integral equations guides us in finding specific functional forms for the production function that reduce the solution to the integral equation (20) to simpler closed form expressions. We show that for our two main examples of production functions, namely CES and Translog, sharp results can be obtained.

We obtain the simplest expressions when the kernel of the equation, \( K_1(\theta, \theta') = \tilde{E}_{l,w}(\theta) \tilde{\gamma}(\theta, \theta') \), is multiplicatively separable between \( \theta \) and \( \theta' \), i.e., of the form \( \kappa_1(\theta) \kappa_2(\theta') \). This is the case when the cross-wage elasticities \( \tilde{\gamma}(\theta, \theta') \) are themselves multiplicatively separable. In particular, this occurs for a CES production function, since in this case \( \tilde{\gamma}(\theta, \theta') \) depends only on \( \theta' \). We analyze the CES technology case in this section.

The second case where the solution to the integral equation can be simplified to a simple closed form is when its kernel is the sum of multiplicatively separable terms. We show that this allows us to characterize the solution to (19) for more general production functions, e.g., the Translog technology, in Section 2.4 below.

We start by considering a CES functional form for the production function as defined in Example 1. In this case, we can further characterize the solution to the integral equation (20) and obtain very sharp economic insights about the incidence of tax reforms.

**Proposition 2.** Suppose that the production function is CES, with constant elasticity of substitution \( \sigma > 0 \). The kernel \( K_1(\theta, \theta') = \tilde{E}_{l,w}(\theta) \tilde{\gamma}(\theta, \theta') \) of the integral equation (19) is multiplicatively separable between \( \theta \) and \( \theta' \). The solution to the integral equation (19) then reduces to

\[
\begin{align*}
\hat{d}_l(\theta, h) &= -\tilde{E}_{l,1}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta))} - \tilde{E}_{l,w}(\theta) \int_\Theta \tilde{\gamma}(\theta, \theta') \frac{\tilde{E}_{l,1}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))} d\theta'}{1 - \int_\Theta \tilde{E}_{l,w}(\theta') \tilde{\gamma}(\theta', \theta') d\theta'}. 
\end{align*}
\]

(26)

Suppose in particular, as in Saez (2001), that the tax reform \( h \) is the step function \( h(y) = \mathbb{I}_{\{y \geq y^*\}} \), so that \( h'(y) = \delta_{y^*}(y) \) is the Dirac delta function (i.e., marginal tax rates are perturbed at income...
We then obtain
\[
\tilde{d}\tilde{l}(\theta, h) = - \frac{\tilde{E}_{l,1-\tau}(y(\theta^*))}{1 - T'(y(\theta^*)))} \left[ \delta y^*(y) + \frac{\tilde{E}_{l,w}(y(\theta)) \tilde{\gamma}(\theta, \theta^*)}{1 - \int_\Theta \tilde{E}_{l,w}(y(\theta')) \tilde{\gamma}(\theta', \theta^*) d\theta'} \right].
\] (27)

Proof. See the Appendix for the full proof. Note here that the perturbation \( h = \mathbb{1}_{\{y \geq y^*\}} \) is not differentiable (\( h'(y) \) is a generalized function). Equation (27) is obtained by applying (26) to a sequence of perturbations \( \{h_n\}_{n \geq 1} \) that converges to \( \delta y^* \). Note also that equation (26) requires \( \int_\Theta \tilde{E}_{l,w}(\theta') \tilde{\gamma}(\theta', \theta') d\theta' \neq 1 \), which is generically satisfied. In fact, suppose for simplicity that the baseline tax schedule \( T \) is linear and the disutility of labor is isoelastic, so that \( \tilde{E}_{l,w}(\theta') = \frac{\varepsilon}{1 + \varepsilon/\sigma} \). We then have
\[
\int_\Theta \tilde{E}_{l,w}(\theta') \tilde{\gamma}(\theta', \theta') d\theta' = \frac{\varepsilon}{1 + \varepsilon/\sigma} \times \frac{1}{\sigma} = \frac{\varepsilon}{\sigma + \varepsilon},
\]
which is always strictly below 1. \( \square \)

We now sketch the proof of equation (26). As already discussed in equation (20), the first term in the r.h.s. is the direct effect of the tax reform \( h \), and the second term accounts for all of the feedback effects on type-\( \theta \) labor supply. The key property that leads to equation (26) is that when the production function is CES, \( \tilde{\gamma}(\theta, \theta') \) depends only on \( \theta' \), i.e., a change in the labor supply of \( \theta' \) affects the wage of all types \( \theta \) by the same amount (in percentage terms). This implies that the kernel of the integral equation (19) is multiplicatively separable, since it can be written as \( \kappa_1(\theta, \theta') = \kappa_1(\theta) \kappa_2(\theta') \), where \( \kappa_1(\theta) = \tilde{E}_{l,w}(\theta) \) and \( \kappa_2(\theta') = \tilde{\gamma}(\theta, \theta') \). It follows that the whole cumulative sum in (20) is equal to the first iterated kernel \( (n = 1) \) giving the direct impact of type-\( \theta' \) labor supply on type-\( \theta \) labor supply,
\[
\tilde{E}_{l,w}(\theta) \tilde{\gamma}(\theta, \theta') \times \tilde{E}_{l,1-\tau}(\theta') \frac{h'(y(\theta'))}{1 - T'(y(\theta'))},
\]
appropriately discounted by the denominator in (26). To see this, note that by using the integral equation (19) we can directly solve for the integral term on the right hand side (which accounts for all the general equilibrium effects), by multiplying each side of the equation by \( \kappa_2(\theta) \) and integrating:
\[
\left\{ \int_\Theta \kappa_2(\theta) d\tilde{l}(\theta, h) d\theta \right\} = - \int_\Theta \kappa_2(\theta) \tilde{E}_{l,1-\tau}(\theta) \frac{h'(y(\theta))}{1 - T'(y(\theta)))} d\theta
\]
\[
+ \left\{ \int_\Theta \kappa_1(\theta) \kappa_2(\theta) d\theta \right\} \left\{ \int_\Theta \kappa_2(\theta') d\tilde{l}(\theta', h) d\theta' \right\},
\]
from which we easily obtain \( \int_\Theta \kappa_2(\theta) d\tilde{l}(\theta, h) d\theta \). Substituting in equation (19) yields expression (26).

One fundamental benefit of the “tax reform” approach of equation (26) over the standard optimal

---

\(^{17}\)This special case is the quintessential nonlinear perturbation of the tax system, since it perturbs the tax rate of one individual independently of everyone else. In contrast, linear tax reforms of a linear baseline tax code would force the marginal tax rates of all individuals to be equal. We discuss the incidence effects of these various perturbations below.
i.e., both the marginal tax rate for \( p < 0 \) is progressive (resp., regressive), i.e., both the marginal tax rate \( T'(y) \) and the average tax rate \( T(y)/y \) are strictly increasing with income. Formally, the parameter \( p \) is the elasticity of the net-of-tax rate \( 1 - T'(y) \) with respect to income \( y \). This CRP tax schedule has two advantages. First, it is very tractable and leads to sharp economic insights about the tax incidence problem (Corollary 3). This is because, assuming that the disutility of labor is isoelastic (with coefficient \( \varepsilon \)), the elasticities \( \bar{\varepsilon}_{l,1-\tau} y \) and \( \bar{\varepsilon}_{l,w} y \) are then constant, respectively equal to \( \frac{\varepsilon}{1+\varepsilon} \) and \( \frac{(1-p)\varepsilon}{1+\varepsilon} \). Second it is a very accurate approximation of the actual US tax and transfer system (with \( p \approx 0.151 \), see Heathcote et al. (2014)), which thus allows us to formally analyze the incidence of policy-relevant tax reforms.

The following corollary gives us a simple theoretical expression for the full incidence effects of all the nonlinear perturbations, starting from the baseline tax schedule (28). This theoretical formula is thus directly operational to characterize the effects of any tax reform of the U.S. tax code in general equilibrium (assuming a CES production function). It moreover leads to one of our key results, namely, that the insights about the optimum tax schedule may be misleading when considering the effects of reforming a suboptimal tax code.

**Corollary 3.** Suppose that the production function is CES, that the disutility of labor is isoelastic, and that the baseline tax schedule is CRP. Then the effect on government revenue of the Saez (2001) perturbation at \( y^* \), \( h(y) = \mathbb{I}_{y \geq y^*} \), is given by:

\[
\frac{d\mathcal{R}(T,h)}{1 - F_y(y^*)} = \left\{ 1 - \frac{T'(y^*)}{1 - T'(y^*)} \bar{\varepsilon}_{l,1-\tau} \frac{y f_y(y^*)}{1 - F_y(y^*)} \right\} + \left\{ \bar{\varepsilon}_{l,w} \left( 1 + \bar{\varepsilon}_{l,w} \right) \right\} \int_{\mathbb{R}} \frac{T'(y') - T'(y)}{1 - T'(y^*)} y f_y(y) dy.
\]

(Note in particular that if the baseline tax schedule is linear \( p = 0 \), then the effect of any perturbation on government revenue is the same as in partial equilibrium.) The effect of the perturbation \( h \) on social welfare is given by

\[
\frac{d\mathcal{W}(T,h)}{1 - F_y(y^*)} = \left\{ \frac{d\mathcal{W}(T,h)}{1 - F_y(y^*)} - \int_{y^*}^{\infty} g(y) \frac{f_y(y)}{1 - F_y(y^*)} dy \right\} - \left\{ \bar{\varepsilon}_{l,1-\tau} \gamma(\theta,\theta^*) \int_{\mathbb{R}} g(y) \frac{1 - T'(y) - T'(y) y f_y(y)}{1 - T'(y^*)} dy \right\}.
\]

**Proof.** See the Appendix for the full proof. In the proof we first derive the incidence of the tax...
reform \( h(y) = \mathbb{I}_{\{y \geq y^*\}} \) on labor supply as:

\[
d\tilde{l}(\theta, h) = -\frac{\tilde{E}_{l,1-\tau}}{1 - T'(y^*)} \left[ \delta_{y^*}(y) + \tilde{\epsilon}_{l,w}(\theta) \gamma(\theta, \theta^*) \right],
\]

which leads to straightforward individual wage and welfare incidence formulas from (22) and (23).

Note that equations (29) and (30) are closed-form expressions in which all the variables (taxes, elasticities, density of incomes) are empirically observable, since they are evaluated given the current tax system. This is in contrast with the standard optimal taxation approach (see Section 3), where the formulas involve evaluating the corresponding variables at the optimum tax system.

The first line on the right hand side of equation (29) is the partial equilibrium effect of the perturbation on government revenue, that one would compute if one assumed (wrongly) that the wage distribution were exogenous. It consists of the mechanical effect of the tax increase, normalized to 1, and of the behavioral effect, which reduces revenue through the labor supply adjustments. The latter effect is determined by the labor supply elasticity \( \tilde{\epsilon}_{l,1-\tau} \) and the hazard rate of the income distribution, which measures the fraction of individuals whose labor supply is distorted, \( f_y(y^*) \), relative to those who pay the additional tax, \( 1 - F_y(y^*) \). We can interpret this behavioral effect (normalized by the mechanical effect of the perturbation) as the excess burden of the tax reform: out of a $1 statutory increase in government revenue, it computes how much the government effectively loses (in partial equilibrium) due to individual labor supply responses.

The second line of equation (29) describes the additional excess burden generated by the tax reform on government revenue, due to the general equilibrium forces. Similarly, in equation (30), the integral in the first line of the right hand side is the partial equilibrium welfare effect of the perturbation (which is given by a weighted sum of the marginal social welfare weights), while the second line gives the welfare adjustments in general equilibrium.

As a first step, suppose that the baseline tax schedule is linear, i.e. \( p = 0 \). In this case, we have \( T'(y) = T'(y^*) \) for all \( y \), so that (29) implies directly that the government revenue effects of the reform are identical to those in partial equilibrium. Note, however, that the general equilibrium effects are non-zero at the individual level; but they cancel out in the aggregate. As a result, the general equilibrium effects of any (linear or nonlinear) tax reform on aggregate revenue are equal to zero, if the baseline tax schedule is linear and the production function is CES.

Now suppose that the baseline tax schedule is nonlinear, i.e. \( p \neq 0 \). To understand the direction in which these novel forces drive the effects of tax reforms, consider the key term in the second line of equation (29):

\[
\int_R \frac{T'(y^*) - T'(y)}{1 - T'(y^*)} y f_y(y) \, dy.
\]

We show in Section 4 that this is the term that drives the shape of the second line of equation (29).\(^{18}\) Given that \( T'(y) = 1 - (1 - \tau) y^{-p} \), this term is strictly increasing in \( y^* \), first negative (for

\(^{18}\)The term \( \tilde{E}_{l,1-\tau} (1 + \tilde{\epsilon}_{l,w}) \) is constant, and \( \tilde{\epsilon}_{l,w}(\theta) \gamma(\theta, \theta^*) \) is almost constant (it is slightly inverse-U-shaped in the
$y < y^*$) then positive (for $y > y^*$) for a progressive baseline tax schedule ($p > 0$).

This implies that the government revenue gains\(^\text{19}\) of raising the marginal tax rate at a given income level $y^*$ are lower than in partial equilibrium for low $y^*$, and higher for high $y^*$. In other words, starting from the U.S. tax code (represented by the CRP tax schedule (28)), general equilibrium forces raise the benefits of increasing the progressivity of the tax schedule. At first sight this result may seem to be at odds with the familiar insight of Stiglitz (1982) in the two-income model, which we generalize to a continuum of incomes in Section 3. Indeed, these results say that the optimal tax rates should be lower at the top, and higher at the bottom, of the income distribution, relative to the partial equilibrium benchmark. That is, the optimal tax schedule should be more regressive when the general equilibrium forces are taken into account. We explain in greater detail this discrepancy in Section 4, but we can already understand its reason. This is because we consider here reforms of the current (suboptimal) U.S. tax code, with low marginal tax rates at the bottom. Instead, the results about the optimum tax schedule use as a benchmark the optimal partial equilibrium tax schedule, which features high marginal tax rates at the bottom (see Diamond (1998), Saez (2001)). Thus, starting from a regressive tax code (at the bottom at least) leads to the opposite sign for the terms $T'(y^*) - T'(y)$ in (29), which yields gains from raising tax rates at the bottom and lowering them at the top. The key take-away of this section is thus that insights about the optimum tax schedule may actually be reversed when considering reforms of the current U.S. tax code. We should therefore be cautious, in practice, when applying the results of optimal tax theory in general equilibrium. This is illustrated in Figure 3 in Section 4.

2.4 Translog production function

In this section, we show that the general solution (20) to the integral equation (19) can be simplified for other production functions, which have the feature that the change in labor supply of type $\theta'$ affects the wage of different types $\theta$ differently. Recall that in the CES case, we obtained further tractability thanks to the multiplicative separability of the kernel $K_1(\theta, \theta')$. More generally, simple closed forms for the solution to the integral equation can be obtained when its kernel is the sum of functions that are multiplicatively separable in $\theta$ and $\theta'$, i.e., of the form $K_1(\theta, \theta') = \sum_{i=1}^{n} \kappa_{i,1}(\theta) \kappa_{i,2}(\theta')$. We show here how to use this result to solve the integral equation (19) for more general production functions than the CES.

Suppose in particular that the production function is Translog, as defined in Example 2. We saw in equation 15 that the cross wage elasticities $\bar{\gamma}(\theta, \theta')$ inherit the separability properties of the coefficients $\bar{\beta}_{\theta, \theta'}$, so that it is not possible to simplify (20) at this level of generality. Now suppose as in Lemma 1 that the coefficients $\tilde{\beta}_{\theta, \theta'}$ have a bivariate lognormal distribution with correlation coefficient $\rho$. Expression (17) shows that a first-order Taylor expansion in $\rho$ leads to an elasticity calibrated version of the model of Section 4, but its variations are small in magnitude compared to those of the integral term).

\(^{19}\)The revenue gains are equivalently the welfare gains of a Rawlsian planner, if the lowest type is $\theta = 0$, and hence never works. We show numerically in Section 4 that the insights of this paragraph on government revenue carry on to welfare when the lowest type $\theta$ is small enough (and hence has a low wage) and the government’s social objective is close enough to Rawlsian (i.e., values mostly the welfare of the very low income agents).
that is linearly decreasing in the square-distance between types for small values of $\rho$. This expression can be easily written as the sum of multiplicatively separable functions, so that the integral equation (19) can be solved in closed form. We obtain:

**Proposition 3.** Suppose that the production function is Translog with $\beta_{\theta,\theta'} = \hat{\varphi}_{m,s,\rho}(\theta,\theta')$. To a first-order as $\rho \to 0$, the kernel $K_1(\theta,\theta') = \tilde{E}_{l,w}(\theta) \bar{\gamma}(\theta,\theta')$ of the integral equation (19) is then the sum of multiplicatively separable functions. Equation (20) then reduces to the solution:

$$dl(\theta, h) = -\frac{h'(y(\theta))}{1 - T'(y(\theta))} - \tilde{E}_{l,w}(\theta) \left\{ c_1 + c_2 \bar{\gamma}(\theta,\theta) + c_3 \bar{\gamma}(\theta, \tilde{\theta}) \right\},$$

where the constants $c_1, c_2, c_3, \bar{\theta}$, given in closed form in the Appendix. The incidence of tax reforms on wages, utilities, government revenue, and social welfare are then obtained from Corollaries 1 and 2.

**Proof.** See Appendix.

The interpretation of formula (32) is as follows. As usual, the first term is the partial equilibrium effect corrected to account for the own-wage GE effect.

The second term shows that, despite the significantly more complicated structure than in the case of a constant elasticity of substitution, the iterated general equilibrium effect has a particularly simple expression. Specifically, it is the product of the general equilibrium labor supply elasticity $\tilde{E}_{l,w}(\theta)$ with an affine function of just two variables – the own-wage elasticity $\bar{\gamma}(\theta,\theta)$ and the cross elasticity $\bar{\gamma}(\theta, \tilde{\theta})$ with an “average” type $\tilde{\theta}$. The intuition behind the affine structure is as follows. In the Cobb-Douglas case where $\beta_{\theta,\theta'} = 0$ for all $\theta, \theta'$, we already showed in equation (26) that the term in the curly brackets is a constant, $c_1$. If the correlation coefficient of the joint Gaussian distribution for $\beta_{\theta,\theta'}$ were exactly $\rho = 0$, we would have an additional second term $c_2 \bar{\gamma}(\theta,\theta)$ inside the brackets (“variable elasticity effect”). We show in the Appendix, by explicitly computing the first few feedback effects on type-$\theta$ labor supply in the general formula (20), that they are all proportional to the own-wage effect $\bar{\gamma}(\theta,\theta)$ (up to a constant $c_1$).

Finally, when $\rho \neq 0$, so that a higher distance between types lowers their substitutability, equation (32) shows that the total wage effect on $\theta$ is proportional to $\bar{\gamma}(\theta,\theta)$, for some “average” type $\tilde{\theta}$ (which depends on $\theta$) that satisfies

$$\ln \tilde{\theta} = \int_{\Theta} \ln \theta' \frac{\hat{\varphi}_{m,s,\rho}(\theta')}{\int_{\Theta} \hat{\varphi}_{m,s,\rho}(\theta'')} d\ln l_{\theta'}(T, h) \frac{d\ln l_{\theta''}(T, h)}{d\theta''} d\theta''.$$

Thus the whole sequence of cross wage effects on $\theta$ can be summarized by a single round of wage effect coming from one average type $\tilde{\theta}$, i.e., by the cross wage elasticity $\bar{\gamma}(\theta,\theta)$ (“distance dependence effect”). This simple aggregation result comes from the fact that (with our bivariate Gaussian functional form and a Taylor approximation at $\rho = 0$), the distance between types affects their substitutability linearly.
In the Appendix, we show that this technique can be fruitfully extended to other production functions as well.

3 Optimal income taxation

In this section, we study the optimal taxation problem. The government maximizes social welfare subject to a resource constraint, and given the fact that wages and labor supply form an equilibrium:

$$\begin{align*}
\max_{T(.)} \int_{\Theta} u \left[ w(\theta) l(\theta) - T(w(\theta) l(\theta)) - v(l(\theta)) \right] f_{\theta}(\theta) \, d\theta \\
\text{s.t.} \quad \int_{\Theta} [w(\theta) l(\theta) - T(w(\theta) l(\theta))] \, f_{\theta}(\theta) \, d\theta \leq F(L) \\
\text{and} \quad (55), \ (56),
\end{align*}$$

Equation (34) is the resource constraint of the government, which imposes that the total amount of resources consumed in equilibrium must be smaller than the total production (or equivalently, by constant returns to scale, that the total tax revenue levied must be non-negative).

In what follows, we characterize this problem in two ways. First, in Section 3.1, we use a mechanism-design approach to derive the optimal informationally-constrained efficient consumption and labor supply allocations. Second, in Section 3.2, we show that the same optimality conditions can be derived from a very different angle (following the analysis of Section 2), namely, by allowing the government directly optimize over tax schedules. Both approaches are complementary but provide different economic insights about the problem. The former allows us to interpret the optimality conditions in terms of incentive compatibility conditions, while the latter leads to an alternative interpretation in terms of labor supply and wage elasticities, allowing us to shed a new light on the economic forces underlying the design of optimal taxes.

In Section 3.3, we show that the optimality conditions simplify when the production function is CES. This assumption on the production function yields sharp economic insights about the design of optimal taxes, in particular regarding the value of the optimal top income tax rate and the shape of the optimal marginal tax schedule. We also show in the Appendix that we can derive the optimum for the Translog production function using again the theory of integral equations and similar techniques as in Section 2.4 above.

3.1 Mechanism design approach

We first study the government problem (33)-(35) using a mechanism-design approach, by optimizing over consumption and labor supply allocations \(\{c(\theta), l(\theta)\}_{\theta \in \Theta}\) subject to feasibility and incentive compatibility of these allocations. This analysis relies on the taxation principle (Hammond, 1979; Rochet, 1985) which states that any incentive-compatible allocation can be implemented with a nonlinear tax schedule.
3.1.1 Government’s problem

It is useful to make a change of variables and optimize over \( \{V(\theta), l(\theta)\}_{\theta \in \Theta} \) instead of \( \{c(\theta), l(\theta)\}_{\theta \in \Theta} \), where \( V(\theta) \equiv c(\theta) - v(l(\theta)) \). The mechanism-design problem then reads as

\[
\max_{V(\theta), l(\theta)} \int_{\Theta} u(V(\theta)) \tilde{f}_{\theta}(\theta) d\theta
\]

subject to the resource constraint

\[
\mathcal{F}(\mathcal{L}) - \int_{\Theta} [V(\theta) + v(l(\theta))] f_{\theta}(\theta) d\theta \geq 0
\]

and the incentive compatibility constraints

\[
c(\theta) - v(l(\theta)) \geq c(\theta') - v\left( l(\theta') \frac{w(\theta')}{w(\theta)} \right), \quad \forall (\theta, \theta') \in \Theta^2.
\]

The incentive constraint of type \( \theta \) can be expressed as a standard envelope condition:

\[
V'(\theta) = v' (l(\theta)) l(\theta) \frac{w'(\theta)}{w(\theta)}
\]

along with the monotonicity constraints \( w'(\theta) > 0 \) and \( y'(\theta) \geq 0 \).

An important issue in contrast with the partial equilibrium case about the envelope condition (39) is that \( w'(\theta) \) it is not only a function of \( l(\theta) \) but also of its derivative \( l'(\theta) \). Indeed, recall from (3) that \( w(\theta) = \omega(\theta, L(\theta), \mathcal{L}) \), so that

\[
w'(\theta) = \omega_1(\theta, L(\theta), \mathcal{L}) + \omega_2(\theta, L(\theta), \mathcal{L}) L'(\theta).
\]

Thus, the envelope condition writes:

\[
V'(\theta) = v' (l(\theta)) l(\theta) \frac{\omega_1(\theta, L(\theta), \mathcal{L}) + \omega_2(\theta, L(\theta), \mathcal{L}) L'(\theta)}{w(\theta)}
\]

Since (40) contains the derivative of the control function \( l'(\theta) \),\(^{20}\) we cannot solve the optimal control problem \( (36,37,40) \) using \( V(\theta) \) as a state and \( l(\theta) \) as a control variable. To overcome this issue, define \( b(\theta) = l'(\theta) \) and rewrite the government problem as follows:

\[
\max_{V(\theta), l(\theta), b(\theta)} \int_{\Theta} u(V(\theta)) \tilde{f}_{\theta}(\theta) d\theta
\]

subject to the resource constraint (37), the envelope condition

\[
V'(\theta) = v' (l(\theta)) l(\theta) \frac{\omega_1[\theta, l(\theta) f_{\theta}(\theta), \mathcal{L}] + [l(\theta) f'_{\theta}(\theta) + b(\theta) f_{\theta}(\theta)] \omega_2[\theta, l(\theta) f_{\theta}(\theta), \mathcal{L}]}{w(\theta)}
\]

\(^{20}\)Note that \( L'(\theta) = l'(\theta) f_{\theta}(\theta) + l(\theta) f'_{\theta}(\theta) \).
the requirement that
\[ l'(\theta) = b(\theta), \quad (43) \]
and the monotonicity constraints \( y'(\theta) \geq 0 \) and \( w'(\theta) > 0 \). As is standard in the literature, we assume that the monotonicity assumptions are satisfied and verify them ex-post in our numerical simulations. This is now a well-defined optimal control problem with two state variables, \( V(\theta) \) and \( l(\theta) \), and one control variable, \( b(\theta) \) (Seierstad and Sydsaeter, 1986).

3.1.2 Optimal tax schedule
Throughout this section we consider a general production function \( F(L) \).

**Proposition 4.** For any \( \theta \in \Theta \), the optimal marginal tax rate \( \tau(\theta) \) of type \( \theta \) satisfies
\[
\frac{\tau(\theta)}{1 - \tau(\theta)} = \left(1 + \frac{1}{\varepsilon} \right) \frac{\mu(\theta)}{f(w(\theta))w(\theta)} + \frac{\int_{\Theta} [\mu(x)v'(l(x))l(x)]' \gamma(x,\theta) dx}{f(\theta)\lambda (1 - \tau(\theta)) y(\theta)},
\]
where \( \mu(\theta) = \int_{\Theta} (1 - g(x))dF_\theta(x) \) is the Lagrange multiplier on the envelope condition (42) of type \( \theta \).

**Proof.** See Appendix.

The first term in the right hand side of (44) is the formula for optimal taxes in partial equilibrium, which can be easily written in the “ABC” form as in Diamond (1998). The second term in (44) captures how each incentive constraint is affected by a small variation in type-\( \theta \) labor supply \( l(\theta) \).

To gain some intuition for this term, consider first a model with two types, as in Stiglitz (1982). In this case, an decrease in the tax on the high type increases her labor supply, which in turn decreases her wage rate; conversely a higher tax on the low type raises her wage. This compression of the pre-tax wage distribution in general equilibrium is beneficial as it relaxes the downward incentive constraint (38) of the high type. Therefore optimal taxes should be more regressive than in partial equilibrium: the optimal marginal tax rate on the high type is negative (as compared to zero in the partial equilibrium setting) and it is higher for the low type than in the partial equilibrium model.

Now suppose that there is a discrete set of types, \( \Theta = \{\theta_i\}_{i=1,\ldots,N} \). In this case, the effects are more complicated because every incentive constraint will be affected by a higher tax on type \( \theta_i \), and it is no longer obvious whether they get relaxed or tightened by the perturbation. The incentive constraint of type \( \theta_i \) reads:
\[
V(\theta_i) \geq V(\theta_{i-1}) + v(l(\theta_{i-1})) - v \left( l(\theta_{i-1}) \frac{w(\theta_{i-1})}{w(\theta_i)} \right).
\]
Denote by \( \mu(\theta_i) \) the Lagrange multiplier on this incentive constraint, and consider the welfare impact of a small increase \( dw(\theta_i) \) in the wage \( w(\theta_i) \), through the incentive constraints. On the one hand, this perturbation reduces the gap between \( w(\theta_{i+1}) \) and \( w(\theta_i) \), and therefore relaxes the downward
incentive constraint of type \( \theta_{i+1} \). The impact on welfare that this force implies is given by

\[
\mu(\theta_{i+1}) v' \left( l(\theta_i) \frac{w(\theta_i)}{w(\theta_{i+1})} \right) \frac{l(\theta_i)}{w(\theta_{i+1})} dw(\theta_i) > 0.
\]  

(45)

On the other hand, the perturbation \( dw(\theta_i) > 0 \) increases the gap between \( w(\theta_i) \) and \( w(\theta_{i-1}) \), and therefore tightens the incentive constraint of type \( \theta_i \). The impact on welfare that this force implies is given by

\[
-\mu(\theta_i) v' \left( l(\theta_{i-1}) \frac{w(\theta_{i-1})}{w(\theta_i)} \right) l(\theta_{i-1}) \frac{w(\theta_{i-1})}{w(\theta_i)^2} dw(\theta_i) < 0.
\]  

(46)

As a consequence, whether a small increase in \( w(\theta_i) \) increases or decreases welfare through its impact on the incentive constraints, depends on whether (45) or (46) is larger in magnitude. First, this depends on the relative size of the Lagrange multipliers \( \mu(\theta_i) \) and \( \mu(\theta_{i+1}) \), i.e., on \( \mu(\theta_{i+1}) - \mu(\theta_i) \). Intuitively, this captures which incentive constraint binds more strongly. In the continuous-type limit \( (\Theta = [\bar{\theta}, \bar{\theta}] \) ), this difference becomes the derivative \( \mu'(x) \) which appears in expression (44). Second, it depends on the respective changes in the values of “lying” that the perturbation \( dw(\theta_i) \) induces. This is given by the difference between \( v' \left( l(\theta_{i-1}) \frac{w(\theta_{i-1})}{w(\theta_i)} \right) l(\theta_{i-1}) \frac{w(\theta_{i-1})}{w(\theta_i)^2} \) and \( v' \left( l(\theta_i) \frac{w(\theta_i)}{w(\theta_{i+1})} \right) \frac{l(\theta_i)}{w(\theta_{i+1})} \).

In the continuous-type limit, this difference becomes \( [v'(l(x))l(x)]' \) which appears in expression (44).

Summarizing, Proposition 4 derives the optimal tax formula in our continuous-type model, using a mechanism design approach. This formula (44) is similar to those derived by Rothschild and Scheuer (2016) and Ales et al. (2015), in their context. However, it is difficult to get more detailed insights for the general equilibrium term using the mechanism design approach either with continuous types or with discrete types (except for two types as in Stiglitz (1982)), as the results depend on the size of the Lagrange multipliers on the incentive constraints which are not easily interpretable or estimable. This is why, in the next section, we turn to the variational approach introduced in Section 2, which allows us to derive further insights about the design of optimal taxes in our economy, both for a general production function, and for specific (CES, Translog) technologies.

### 3.2 Variational approach

In this section we propose an alternative characterization of the optimal tax schedule using a variational, or “tax reform” approach. This method builds on the characterization of the welfare effects of tax reforms, which we analyzed in Section 2. We propose a specific tax reform that allows us not only to recover expression (44), but also to provide a more comprehensive economic characterization of the optimum tax schedule.

Specifically, consider as in Saez (2001) a tax reform \( h_1(y) \) that consists of an increase in the marginal tax rate at the income level \( y^\star \). This induces a uniform lump-sum tax increase for all incomes \( y \geq y^\star \). Formally, we let \( h_1'(y) = \delta_{y^\star}(y) \) be the Dirac delta function at \( y^\star \), so that \( h_1 \) is the step function \( h_1(y) = \mathbb{I}_{\{y \geq y^\star\}} \). Denote by \( \theta^\star \) the type such that \( y(\theta^\star) = y^\star \) in the baseline equilibrium.
We derived in Section 2 (equation (20)) the effects of this tax reform on individual labor supplies, as the fixed point of an infinite sequence of cross-effects mediated through wages. Since at the optimum tax schedule the aggregate welfare effects of any perturbation should be equal to zero, we propose a counteracting perturbation $h_2(y)$ such that the general equilibrium effects of $h_1 + h_2$ on labor supply are the same as the partial equilibrium effects of $h_1$. That is, in response to $h_1 + h_2$, the labor supply of all types $\theta \neq \theta^*$ stays constant, and that of type $\theta^*$ reduces to its partial equilibrium response to $h_1$ (as in Saez (2001)). We then derive an optimality condition by imposing that the tax reform $h_1 + h_2$ has no first-order effect on social welfare.

Note that our tax reform approach in Section 2 immediately delivers a characterization of the optimum tax schedule, by equating the welfare effects of any tax reform to zero in (25). The reason we analyze the combination of two perturbations $h_1 + h_2$ in this section is that it leads to a substantially simpler formula for the optimum than the one we would obtain with the simple analysis of the perturbation $h_1$. In particular, the formula we obtain using only $h_1$ would feature a large number of objects on the right hand side, e.g. the tax rates, elasticities, and density functions evaluated at the optimum (see (20) and (25)). Setting to zero the sum of the effects of $h_1$ and the counteracting perturbation $h_2$ allows us to derive a simpler optimality condition that features a much smaller amount of such endogenous variables.

The key step consists in using our key integral equation (19) derived in Section 2 to derive the counteracting perturbation $h_2$ that cancels out the general equilibrium effects on labor supply, i.e., that ensures that $d\hat{l}(\theta, h_1 + h_2) = 0$ for all $\theta \neq \theta^*$. We do so formally in the Appendix. Here we provide a sketch of the main steps of the derivation. In order to leave the labor supply of an individual $\theta$ unchanged, we must ensure that the net-of-tax-wage $(1 - T'(y(\theta)))$ remains unchanged, i.e., that

$$d\ln (1 - T'(y(\theta))) = -d\ln w(\theta).$$

That is, the change in the log-net-of-tax rate must exactly compensate the change in the log-wage. Now the change in the wage $w(\theta)$ due to the perturbation $h'_1 = \delta_{\theta^*}$ is given by

$$d\ln w(\theta) = \int_{\Theta} \gamma(\theta, \theta') d\ln l(\theta') d\theta' = -\gamma(\theta, \theta^*) \frac{\hat{\varepsilon}_{l,1-\tau}(\theta^*)}{1 - T'(y(\theta^*))},$$

(47)

where the second equality follows from the fact that the labor supply of all the other agents ($\theta' \neq \theta^*$), and hence all the feedback effects in equation (19), cancel out by construction of the counteracting perturbation $h_2$. If the baseline tax schedule $T$ were linear, so that $T'' = 0$ and $\hat{\varepsilon}_{l,1-\tau} = \varepsilon_{l,1-\tau}$, the change in the log net-of-tax rate would simply be equal to

$$d\ln (1 - T'(y(\theta))) = -\frac{dT'(y(\theta))}{1 - T'(y(\theta))} = -\frac{h'_1(y(\theta))}{1 - T'(y(\theta))},$$

21This also implies that the computation of the optimum is much simpler as we do not have to solve for the integral equation for labor supply for each $h_1$-reform in each iteration.

22Note that this is an equality of generalized functions.
which along the expression for \(d \ln w(\theta)\) immediately yields \(h'_2(y(\theta)) = -\gamma(\theta, \theta^*) \tilde{\varepsilon}_{1,1-\tau}(\theta^*)\). Instead, for a nonlinear tax schedule \(T\), the required counteracting change in the log net-of-tax rate must take in addition into account the feedback effect described above equation (6), i.e., the fact that the change in income induced by the tax perturbation triggers in turn an endogenous marginal tax rate adjustment. Note that since the labor supply responses are equal to zero, by construction of our counteracting perturbation, this endogenous income change is driven only by wages, that is, \(dy(\theta) = l(\theta) \, dw(\theta)\). We thus obtain:

\[
d \ln \left(1 - T'(y(\theta))\right) = -\frac{dT'(y(\theta))}{1 - T'(y(\theta))} = -\frac{h'_2(y(\theta)) + T''(y(\theta)) \, l(\theta) \, dw(\theta)}{1 - T'(y(\theta))}.
\]

The counteracting perturbation is therefore defined by:

\[
h'_2(y) = -\frac{\tilde{\varepsilon}_{1,1-\tau}(\theta^*)}{1 - T'(y^*)} \left(1 - T'(y) - yT''(y)\right) \gamma(y, y^*),
\]

where we make a slight abuse of notation and denote by \(\gamma(y(\theta), y(\theta^*)) \equiv \gamma(\theta, \theta^*)\).

This combination of perturbations \(h_1 + h_2\) is particularly convenient to derive the optimal tax schedule, and in particular isolate the novel effects that arise purely from general equilibrium forces. As any other small perturbation, it cannot have a first-order impact on social welfare if the initial tax schedule is optimal. The following proposition derives the implied optimal marginal tax rate formula.

**Proposition 5.** The optimal tax schedule satisfies

\[
\frac{T'(y^*)}{1 - T'(y^*)} = \frac{1}{\tilde{\varepsilon}_{1,1-\tau}(y^*)} \left(1 - \bar{g}(y^*)\right) \left(1 - \frac{F_y(y^*)}{y^* f_y(y^*)}\right) \left(1 - \frac{T'(y)}{1 - T'(y^*)}\right) \gamma(y, y^*) \, dy,
\]

where \(\bar{g}(y) \equiv \int_{y}^{\infty} g(y') \frac{f_y(y')}{1 - F_y(y')} \, dy'\) is the average marginal social welfare weight above income \(y\).

**Proof.** See Appendix. We also prove that this formula is equal to the expression (44) derived in the mechanism design problem. \(\square\)

The first line in formula (49) is the partial equilibrium formula for optimal tax rates, which is the same as in Diamond (1998). The optimal marginal tax rate at income \(y^*\) is proportional to the inverse elasticity of labor supply \(\tilde{\varepsilon}_{1,1-\tau}(y^*)\), which measures the disincentive effects of taxes on work effort. It is proportional to the hazard rate of the income distribution, \(\frac{1 - F_y(y^*)}{y^* f_y(y^*)}\), which measures the tax revenue gains coming from individuals \(y \geq y^*\) who pay a higher tax liability, relative to the loss coming from those whose labor supply is distorted (\(y = y^*\)). Finally it is decreasing in the average marginal social welfare weight above \(y^*\), \(\bar{g}(y^*)\), which measures the welfare loss of levying a higher lump-sum tax above \(y^*\). Note, however, that the density of incomes \(f_y\) (as well as the labor supply elasticity) that appears in this expression is evaluated at the general equilibrium optimum
tax code, and is therefore not observed given the current (suboptimal) tax code and resulting income distribution (as opposed to the results we had obtained in our general study of tax incidence).

To understand the intuition for the general equilibrium effects in the second line of (49), we sketch the proof of the derivation of the welfare effects of the combination of perturbations \( h_1 + h_2 \) and then equate them to zero to obtain a characterization of the optimum. Recall that the labor supply response to this combination of perturbations is given by \( d \ln l(\theta) = 0 \) for all types \( \theta \neq \theta^* \), and by the partial equilibrium effect of \( h_1 \), \( d \ln l(\theta^*) = -\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \), for \( \theta = \theta^* \).

The combination of perturbations \( h_1(\cdot) + h_2(\cdot) \) has two distinct effects on individuals with income \( y(\theta) \neq y^* \). First, the initial perturbation \( h'_1 = \delta y^* \) induces a direct change in the wage of individual \( \theta \) through cross-wage effects, given by the amount (47); second, the counteracting perturbation \( h'_2 \) raises the marginal tax rate faced by individual \( \theta \) by the same amount as this wage increase (see (48)). There is a crucial difference between these two identical effects on the wage and the marginal tax rate, however. Indeed, a change in the wage \( w(\theta) \) affects only the revenue levied at this income level. In contrast, a change in the marginal tax rate at \( y(\theta) \) increases the total tax liability paid by everyone above \( y(\theta) \) by a lump-sum amount. We now describe these two effects in detail.

**Effects of the wage change.** The change (say, the decrease) in \( \theta \)'s wage by \( dw(\theta) \) induces a change in income \( y(\theta) \) by \( l(\theta) dw(\theta) \), which triggers in turn a change in individual welfare and a change in government revenue. First, the decrease in each individual \( \theta \)'s income, \( l(\theta) dw(\theta) \), affects her utility. Since she only loses a share \( 1 - T'(y(\theta)) \) of her income change, aggregate welfare (in monetary units) changes by

\[
\int_{\Theta} g(y(\theta)) \left[ (1 - T'(y(\theta))) y(\theta) \right] \times d \ln w(\theta) f_{\theta}(\theta) d\theta
\]

\[
= -\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} g(y) \left( 1 - T'(y) \right) \gamma(y, y^*) y f_y(y) dy,
\]

by construction of the marginal social welfare weights. The other share \( T'(y(\theta)) \) is borne by the government, whose tax revenue changes by

\[
\int_{\Theta} \left[ T'(y(\theta)) y(\theta) \times d \ln w(\theta) \right] f_{\theta}(\theta) d\theta = -\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} T'(y) \gamma(y, y^*) y f_y(y) dy.
\]

But since the production function has constant returns to scale, we can use Euler’s homogeneous function theorem to write: for all \( y^* \),

\[
\int_{\mathbb{R}_+} \gamma(y, y^*) y f_y(y) dy = 0.
\]

\footnote{Recall that the perturbation \( h'_2 \) cancels out the labor supply changes of type \( \theta \neq \theta^* \).}
Thus the change in government revenue can be rewritten as

\[
\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} \left(1 - T'(y)\right) \gamma(y, y^*) y f_y(y) \, dy.
\]

Collecting the aggregate welfare and revenue terms, we obtain that social welfare changes by

\[
\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} \left(1 - g(y)\right) \left(1 - T'(y)\right) \gamma(y, y^*) y f_y(y) \, dy.
\]

**Effects of the counteracting marginal tax rate change.** The compensating marginal tax rate increase \( h'_2(y(\theta)) \), which ensures that the labor supply of individual \( \theta \) does not change, raises the total tax liability of all individuals above income \( y(\theta) \) by an equal amount. Therefore this change in marginal tax rate at income \( y(\theta) \) induces an increase in government revenue equal to

\[
h'_2(y(\theta)) \left(1 - F(y(\theta))\right) = -\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \left(1 - T'(y) - yT''(y)\right) \gamma(y, y^*) \left(1 - F(y(\theta))\right).
\]

Moreover, because of the redistributive preferences of the planner, this increase in revenue is only valued \( (1 - \bar{g}(y)) h'_2(y(\theta)) \left(1 - F(y(\theta))\right) \) by the government, where \( \bar{g}(y) \) is the average marginal social welfare weight above income \( y \). Note that the relevant welfare weight for this counteracting increase in the marginal tax rate at income \( y \) is \( \bar{g}(y) \) because, as explained above, this perturbation affects the disposable income of all the agents with income above \( y \). Summing this term over incomes \( y(\theta) \) leads to the change in social welfare coming from this counteracting perturbation.

Collecting these two terms, we obtain that the change in social welfare must be adjusted, due to the general equilibrium effects on the wage, by

\[
\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} \left[ \left(1 - g(y)\right) \left(1 - T'(y)\right) y f_y(y) \right.
\]

\[
\left. - (1 - \bar{g}(y)) \left(1 - T'(y) - yT''(y)\right) \left(1 - F(y)\right) \right] \gamma(y, y^*) \, dy.
\]

Note finally that the term in brackets is a derivative, so that this expression can be rewritten as

\[
-\frac{\tilde{\varepsilon}_{l,1-\tau}(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} \left[ \left(1 - \bar{g}(y)\right) \left(1 - T'(y)\right) y \left(1 - F(y)\right) \right]' \gamma(y, y^*) \, dy,
\]

which leads formula (49).

To sum up the reasoning: in response to a (say, lower) marginal tax rate, \( \theta^* \) increases her labor supply, which affects (say, raises) the wage of \( \theta \). This wage effect directly impacts government revenue. An equal increase in marginal tax rates is then necessary to cancel out the induced change in the labor supply of \( \theta \), and this tax increase impacts the taxes paid by all individuals above \( \theta \). This allows us to understand the key difference between, on the one hand, Stiglitz (1982), Rothschild and Scheuer (2013), Scheuer (2014), Rothschild and Scheuer (2016), Ales et al. (2015), and this
paper, versus, on the other hand, Saez (2004) and Scheuer and Werning (2016). The former set of papers have a limited set of tax instruments, in the following sense. With a finite set of types, or equivalently sectors / occupations, this restriction consists of imposing that the same tax schedule applies to each sector. In our context with a continuum of types, this restriction on the available tax instruments becomes the most natural: it simply means that the government has access to a standard nonlinear income tax schedule, and hence, as we just explained, that a change in the marginal tax rate at some income level affects everyone above that level. Instead, the latter set of papers assume an extremely rich tax structure, where each type faces its own tax rate, so that an increase in the tax rate in sector \( \theta \) does not affect the tax liability paid by individuals in sectors \( \theta' > \theta \). In that situation, we obtain the result that optimal tax rates are the same as in partial equilibrium independently of the production function (this follows from the analysis of Diamond and Mirrlees (1971)).

To conclude this section, notice that one difficulty with equation (49) is that the unknown function \( T'( \cdot ) \) appears both on the left hand side (through its value at \( y^* \)) and on the right hand side (under the integral sign). This makes it a priori difficult to solve (49). However, a simple transformation of this equation allows us to rewrite it as an integral equation, which can be analyzed, both theoretically and numerically, using the same techniques as in Section 2 for the general analysis of incidence.

Specifically, assume that the function \( y \mapsto \bar{\gamma}(y, y^*) \) is continuously differentiable for each \( y^* \), and denote by \( \bar{\gamma}'(y, y^*) \) its derivative evaluated at \( y \). Using the definition (10) to disentangle the wage elasticities \( \gamma(y, y^*) \) into the own-effects \( \bar{\gamma}(y^*, y^*) \) and the cross-effects \( \bar{\gamma}(y, y^*) \), we show in the Appendix that we can rewrite equation (49) as:

\[
\frac{T'(y^*)}{1 - T'(y^*)} = \frac{1}{\hat{E}_{1,\tau}(y^*)} \left( 1 - \bar{g}(y^*) \right) \left( \frac{1 - F_y(y^*)}{y^* f_y(y^*)} \right) - (g(y^*) - 1) \bar{\gamma}(y^*, y^*) + \int_{R_+} \left( 1 - \bar{g}(y) \right) \left( \frac{1 - F_y(y)}{y^* f_y(y^*)} \right) \left( \frac{1 - T'(y)}{1 - T'(y^*)} \right) y\bar{\gamma}'(y, y^*) \, dy,
\]

where we used an integration by parts to transform the integral term. Denoting by \( \frac{T'(y^*)}{1 - T'(y^*)} \) the first line on the right hand side of this expression (which we analyze in depth in Section 3.3 below), and reorganizing the terms of this expression, we obtain that the optimal tax schedule satisfies

\[
\left\{ 1 - T'(y^*) \right\} = 1 - T'(y^*) - \int_{R_+} \left[ (1 - \bar{g}(y)) \left( \frac{1 - F_y(y)}{y^* f_y(y^*)} \right) \left( 1 - T'(y^*) \right) y\bar{\gamma}'(y, y^*) \right] \left\{ 1 - T'(y) \right\} \, dy,
\]

which is now a well-defined integral equation in the net-of-tax rates \( 1 - T'(y) \). We can then use the mathematical apparatus introduced in Section 2 to characterize the solution to this equation, i.e., the optimal marginal tax schedule. This equation becomes particularly simple if the production function is CES, because we then have \( \bar{\gamma}'(y, y^*) = 0 \), so that the integral term is equal to zero. We focus on this case in Section 3.3. If the production function is Translog, as in Section 2.4, we can use Lemma 1 to reduce the kernel of this integral equation to the sum of multiplicatively
separable functions, allowing to solve for the optimal tax rates. We show how this can be done in the Appendix.

3.3 CES production function

3.3.1 Optimal marginal tax rates

The formula in Proposition 4 dramatically simplifies for the special case of a CES production function (defined in Example 1). As we discussed in the previous paragraph, the kernel of the integral equation that defines the optimal tax schedule is equal to zero, so that the formula for optimal tax rates can be significantly simplified. Formula (50) then directly implies:

**Corollary 4.** Assume that the production function is CES. Then the optimal marginal tax rate at income $y^*$ is given by

$$
\frac{1}{1 - T'(y^*)} \frac{1}{\tilde{E}_{l,1-\tau}(y^*)} \left( 1 - g(y^*) \right) \left( 1 - F_y(y^*) \right) + \frac{g(y^*) - 1}{\sigma}.
$$

(51)

*Proof.* See Appendix.

This equation highlights the two main differences between the partial equilibrium and the general equilibrium optimal taxes. These two differences come from the two types of wage effects that occur in general equilibrium: the own-wage effects (of a given type’s labor supply on their own wage) and the cross wage-effects (of a given type’s labor supply on other types’ wages).

The first of these general equilibrium forces is reflected in the inverse of the elasticity $\tilde{E}_{l,1-\tau}(y^*)$, defined (11), which replaces the inverse of the elasticity $\tilde{\varepsilon}_{l,1-\tau}(y^*)$ that appears in the corresponding partial equilibrium formula. This “general equilibrium-corrected elasticity” accounts for the effects on the wage of $y^*$ of a change in her own labor supply. Note that this correction implies a smaller elasticity $\tilde{E}_{l,1-\tau}(y^*) < \tilde{\varepsilon}_{l,1-\tau}(y^*)$, which therefore tends to raise optimal marginal tax rates. Intuitively, this is because an increase in the marginal tax rate at $y^*$ reduces the labor supply at type $\theta^*$, which lowers the wage of all the other types $\theta \neq \theta^*$ (by the same amount in percentage terms, due to the CES technology), and hence their labor supply. Hence the elasticity $\tilde{E}$ that takes into account those own-wage effects is lower than the corresponding partial equilibrium elasticity $\tilde{\varepsilon}$.

The second novel effect is captured by the term $\frac{1-g(y^*)}{\sigma}$ in the equation, which is equal to zero in partial equilibrium (i.e., when $\sigma \to \infty$). While the general equilibrium effect discussed in the previous paragraph is driven by the own-wage effects and tends to raise optimal taxes, this term is driven by the cross-wage effects and works in the opposite direction. Intuitively, an increase in marginal tax rates at $y^*$ reduces the labor supply at type $\theta^*$, which lowers the wage of all the other types $\theta \neq \theta^*$ (by the same amount in percentage terms, due to the CES technology), and hence their labor supply. If the government does not value the welfare of individuals with type $\theta^*$, i.e. $g(y^*) = 0$, then this effect implies that the cost of raising the tax rate at $y^*$ is higher than in partial equilibrium, which tends to lower the optimal tax rate. More generally, if the government values the welfare of individual $y^*$ less than average (i.e., $g(y^*) < 1$, where 1 is the average marginal social
welfare weight in the economy), this negative effect induced by the behavior of \( y^* \) implies that the marginal tax rate at \( y^* \) should be lower than in partial equilibrium. Conversely, if \( g(y^*) > 1 \), the government raises optimal tax rates on \( y^* \); this makes this agent work less and earn a higher wage, which makes him strictly better off, at the expense of the other individuals in the economy, who earn a lower wage.

Formula (51) generalizes the partial equilibrium optimal tax formula of Diamond (1998) to a CES production function, and the two-type general equilibrium analysis of Stiglitz (1982) to a continuum and a nonlinear tax schedule. Note that the case of a CES production function is particularly interesting, as it is the most parsimonious extension to general equilibrium of the standard optimal taxation analysis. In particular, the partial equilibrium model is nested by this case as a strict special case, where the production function is CES with perfect (constant) substitutability between skills, i.e. \( \sigma = \infty \). Formula (51) generalizes the partial equilibrium optimum formula in terms of a single additional parameter, \( \sigma \), which is no longer restricted to be infinite. Thus formula (51) allows us to go beyond the qualitative insights obtained in a model with two types by Stiglitz (1982) and make optimal tax theory in general equilibrium operational. In particular, we can show how the general equilibrium considerations affect the familiar and influential results obtained in partial equilibrium, namely, the characterization of the top tax rate and the U-shape of marginal tax rates. We do so in the following subsections.

### 3.3.2 Top tax rate

We now derive the implications of equation (51) for the optimal top tax rate. We derive a simple expression that shows how the familiar top tax rate formula obtained by Saez (2001) changes in the general equilibrium setting with CES production.

**Corollary 5.** Assume that the production function is CES with elasticity of substitution \( \sigma > 0 \). Assume also that in the data, incomes are Pareto distributed at the tail with Pareto coefficient \( \alpha \), and that the top marginal tax rate that applies to these incomes is constant. Assume moreover that the marginal social welfare weights at the top converge to \( \bar{g} \) and that the elasticity of labor supply is constant and equal to \( \varepsilon \). Then the optimal top tax rate satisfies

\[
\frac{\tau_{top}}{1 - \tau_{top}} = \frac{1 - \bar{g}}{\alpha \varepsilon} + \frac{1 - \bar{g}}{\alpha \sigma} - \frac{1 - \bar{g}}{\sigma} < 0.
\]

**Proof.** See Appendix. The non-trivial part of the proof consists in showing that for a CES production function, if the income distribution has a Pareto tail in the data, then it has the same Pareto tail at the optimum tax schedule if the tax schedule has a constant marginal tax rate at the top (even though the wage distribution is endogenous).

The first term in the right hand side of equation (52), \( (1 - \bar{g}) / \alpha \varepsilon \), is the standard top tax rate formula in partial equilibrium (Saez (2001)). Formula (52) generalizes this well-known result to a general CES production function (where the parameter \( \sigma \) is no longer restricted to be infinite as in
partial equilibrium), and gives in simple closed form the tax rate that should apply to top incomes, as a function of one additional parameter that can be easily estimated in the data, namely, the elasticity of substitution between skills in production $\sigma$. Since $\alpha > 1$, the second and third (general equilibrium) terms lead to a strictly lower top marginal tax rate than in partial equilibrium.

Moreover, and perhaps surprisingly, this general equilibrium correction is stronger (so that the top tax rate should be reduced by a larger amount, relative to the partial equilibrium benchmark) when inequality at the top is lower, i.e. when $\alpha$ is higher, and when the average welfare weight $\bar{g}$ at the top is lower. This is because $\alpha$ appears negatively in the term $\frac{1-\bar{g}}{\alpha \sigma}$, which reflects the elasticity correction due to the own-wage effects discussed after Corollary 4 that tends to raise the optimal top tax rate. Note finally that since the solution for $\tau_{\text{top}}$ is concavely increasing in the r.h.s. of (52), this formula implies that the optimal top tax rate is more sensitive to the labor supply elasticity $\varepsilon$ than in partial equilibrium.

Corollary 5, which generalizes the insight from the two-income model of Stiglitz (1982) to the workhorse framework of taxation (with a continuum of types and arbitrary nonlinear taxes), makes operational the theory of optimal tax design in general equilibrium. Immediate back-of-the envelope calculations of the optimal top tax rate illustrate the power of formula (52). Suppose that the marginal social welfare weights at the top are equal to zero. Suppose that the tail of the Pareto coefficient of incomes is equal to $\alpha = 2$, that the structural elasticity of labor supply is equal to $\varepsilon = 1/2$, and that the elasticity of substitution is equal to $\sigma = 1.5$ (versus $\sigma = \infty$ in the partial equilibrium model).\footnote{These values are meant to be only illustrative, though they are in the ballpark of the parameters estimated in the empirical literature. See our calibration in Section 4 below.} Formula (52) immediately implies that the optimal tax rate on top incomes is equal to $\tau_{\text{top}} = 50\%$ in partial equilibrium, and drops to $\tau_{\text{top}} = 40\%$ once general equilibrium forces are taken into account.

We provide more comprehensive comparative statics for the optimal top tax rate in Figure 1. The left panel shows how the difference between the partial and general equilibrium top tax rates varies with the elasticity of substitution $\sigma$. The horizontal bold lines give the value of the optimal top tax rate in partial equilibrium. The dashed curves show the general equilibrium counterparts. We set two values for the Pareto coefficient $\alpha$, which measures the thinness of the Pareto tail of the income distribution: $\alpha = 1.5$ (red curves) and $\alpha = 2$ (black curves). This figure shows that the difference between the general and partial equilibrium tax rates is convexely increasing in $-\sigma$, and that it increases with the thinness parameter $\alpha$. 

24These values are meant to be only illustrative, though they are in the ballpark of the parameters estimated in the empirical literature. See our calibration in Section 4 below.
The right panel illustrates the sensitivity of the optimal top tax rate to the structural elasticity parameter $\varepsilon$. Again we assume two values for the Pareto coefficient: $\alpha = 1.5$ (red curves) and $\alpha = 2$ (black curves). Moreover we set $\sigma = 1.4$, which is the baseline value for our calibration in Section 4. As can be seen in the figure, the difference between the general and the partial equilibrium tax rate is increasing in the structural elasticity parameter $\varepsilon$, so that the optimal general equilibrium top tax rate is more sensitive to the elasticity than in partial equilibrium. We generalize this insight to the whole tax schedule below.

### 3.3.3 U-shape of optimal marginal tax rates

We finally analyze the impact of the general equilibrium forces (in particular, of the novel term $(g(\theta) - 1)/\sigma$ in (51)) on the familiar U-shape of optimal tax rates derived by Diamond (1998) in the partial equilibrium model.

**Corollary 6.** Suppose that the social planner is Rawlsian, so that $g(\theta) = 0$ for all $\theta > \theta$. If the partial equilibrium optimal tax formula\(^{25}\)

$$
\frac{1 - F_y(y(\theta))}{E_T(\theta) y(\theta) f_y(y(\theta))}
$$

implies a U-shaped pattern of marginal tax rates, then the additional term $-1/\sigma$ leads to a general equilibrium correction for $T'(\cdot)$ that is also U-shaped.

**Proof.** See Appendix.

Corollary 6 is a simple but economically important result as it implies that the general equilibrium forces tend to reinforce the U-shape of the optimal marginal tax rate schedule. This leads in particular to a more pronounced dip, with higher marginal tax rates at the bottom and lower tax rates in the bulk and at the top of the distribution. Therefore, while the two-income model of

\(^{25}\)Note that these partial equilibrium effect is computed given the wage distribution at the optimum general equilibrium tax schedule.
Stiglitz (1982) suggests that optimal tax rates should be more regressive (i.e., lower at the top and higher at the bottom) than in partial equilibrium, we see here that this insight must be qualified once we extend it to the canonical taxation framework with a U-shaped optimal marginal tax schedule: taxes should be more regressive for incomes below the bottom of the U, but more progressive in the region where optimal tax rates were already progressive. We illustrate this result in Section 4 (Figure 1).

4 Numerical simulations

In this section we provide a quantitative analysis of our results. We start in Section 4.1 by describing the calibration of our model. In Section 4.2, we study numerically the tax incidence results of Section 2. In Section 4.3, we study numerically the optimal income taxation results of Section 3. In particular, in Section 4.4, we construct a policy-relevant partial equilibrium benchmark in order to focus on the quantitative departure of our optimal general equilibrium tax policy recommendations from those of Diamond (1998). Finally, in Section 4.4, we show how to reconcile our tax incidence and optimal taxation results, and emphasize that the policy recommendations that would be true at the optimum may be reversed when contemplating reforms of the current tax system.

4.1 Calibration

We first calibrate the wage distribution from the income distribution as in Saez (2001), and then infer the parameters of the CES production function that are consistent with these wages in a second step. We assume that preferences are quasilinear with isoelastic disutility of labor, \( U(c, l) = c - \frac{l^{1+\frac{1}{\varepsilon}}}{(1 + \frac{1}{\varepsilon})} \) and set \( \varepsilon = 0.33 \) in our benchmark calibration (Chetty et al., 2011) (we show comparative statics for \( \varepsilon \) in Section Section 4.3).

We assume that incomes are log-normally distributed apart from the top, where we append a Pareto distribution for incomes above $150,000. We set the mean and variance of the lognormal distribution at 10 and 0.95, respectively. The mean parameter is chosen such that the resulting income distribution has a mean of $64,000, i.e., approximately the average US yearly earnings. The variance parameter was chosen such that the hazard ratio at level $150,000 is equal to that reported by Diamond and Saez (2011, Fig.2). The resulting hazard ratio is illustrated in Figure 2.

We then calibrate the resulting wage distribution from the agents’ first-order conditions (Saez, 2001) and the current marginal tax rates. We assume a CRP tax schedule as defined in (28) and set the parameters \( p = 0.151 \) and \( \tau = -3 \) (Heathcote et al., 2014).

We now back out the production function that is consistent with this wage distribution and the associated effort levels. We assume a CES production function for most of the analysis, and relegate the analysis of the Translog technology to the Appendix. An assumption has to be made about the elasticity of substitution. Our approach here is to look at a variety of values that are in line with

\[ \frac{1 - F_y(y)}{yf_y(y)} \] we decrease the thinness parameter of the Pareto distribution linearly between $150,000 and $350,000 and let it be constant at 1.5 afterwards (Diamond and Saez, 2011). In the last step we use a standard kernel smoother to ensure differentiability of the hazard ratios at $150,000 and $350,000.
empirical evidence and show to what extent our results are sensitive with respect to that choice. The benchmark values that we consider are $\sigma \in \{0.6, 1.4, 3.1\}$, taken respectively from Dustmann et al. (2013), Katz and Murphy (1992) and Borjas (2003), and Card and Lemieux (2001). Having made an assumption about the elasticity of substitution $\sigma$, we can back out the values of $a(\theta)$ using the values of $l(\theta)$ and $w(\theta)$ at each income level.

### 4.2 Tax incidence

We first study the incidence of tax reforms on government revenue, based on the theoretical analysis of Section 2. The reforms that we consider are those that increase the marginal tax rate at a single income level $y^*$. We thus plot the values of $dR(y^*)/(1 - F_y(y^*))$ (defined in (24)), as a function of the income $y^*$ where tax rates are perturbed. The interpretation of this variable is as follows. A value 0.7, say, at a given income level $y^*$, means that for each additional dollar of tax revenue raised from the statutory increase in the marginal tax rate at $y^*$, the government effectively gains only 70 cents, while 30 cents are lost through the behavioral responses of individuals to the tax reform. That is, the marginal excess burden of this tax reform is 30%.

Figure 3 illustrates the result for two values of the elasticity of substitution $\sigma = 0.6$ and $\sigma = 3.1$. The blacked dashed curves give the results. The bold red curves plot the same variable but under the assumption that there are no general equilibrium effects. Unsurprisingly, the impact of general equilibrium effects is larger the lower the elasticity of substitution.

In line with our theoretical result in Corollary 3, we observe that increasing the marginal tax rates for intermediate and high incomes (starting from about $100,000) becomes more desirable due to general equilibrium effects. The opposite holds for low income levels. For example, for $\sigma = 0.6$ we find that a higher marginal tax rate at income level $200,000 implies a marginal excess burden of 35% if general equilibrium effects are ignored: for each mechanically raised dollar the

\[ 27 \text{Dustmann et al. (2013) is the most relevant study for our analysis, as they do not classify workers according to their education level (as opposed to Katz and Murphy (1992), Card and Lemieux (2001), Borjas (2003)) but according to their position in the wage distribution, i.e., by type } \theta. \]
government receives 65 cents and loses 35 cents from the induced labor supply distortion. Taking into account general equilibrium effects considerably decreases the marginal excess burden to 22 cents (and increases revenue to 78 cents). This is a striking result that contrasts the famous regressivity result of Stiglitz (1982): general equilibrium effects make higher taxes on the rich more favorable as measured by the marginal excess burden of taxation. We therefore conclude that the insights about the optimal tax schedule may be misleading when considering the effects of reforming a suboptimal tax code. We show the implications for optimal taxes in Section 4.3 and reconcile the two sets of results in Section 4.4.

Figure 3: Tax incidence: $\sigma = 0.6$ (Dustmann et al. (2013), left panel) and $\sigma = 3.1$ (Card and Lemieux (2001), right panel)

4.3 Optimal taxation

4.3.1 Partial equilibrium benchmark

In this section we construct a policy-relevant partial equilibrium benchmark to ask how the normative prescriptions for optimal taxes in general equilibrium differ from those one would obtain from the optimal tax formula of Diamond (1998). Specifically, we define the marginal tax rates that a partial equilibrium planner would set from Diamond (1998), using the same data to calibrate the model and making the same assumptions about the utility function, but wrongly assuming that the wage distribution is exogenous. That is, we compute the optimal tax schedule in partial equilibrium as:

$$
\frac{\tau_{\text{PE}}(\theta)}{1 - \tau_{\text{PE}}(\theta)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1 - F_w^d(w_d(\theta))}{f_w^d(w_d(\theta))w_d(\theta)} (1 - \bar{g}(\theta)),
$$

where $F_w^d(w_d(\theta))$ is the wage distribution in the data (wrongly assumed to be exogenous), i.e., backed out from the observed income distribution.

---

28 In unreported simulations, we also explored a Gouveia-Strauss approximation of Guner et al. (2014) for the U.S. tax schedule, where the top tax rate is constant. The results are equivalent.
4.3.2 Simulation results

We consider the optimal Rawlsian policy of maximizing the lump-sum element of the tax schedule, i.e., the utility of agents that earn zero income.

**The role of the elasticity of substitution.** We first show the optimal tax schedules for a fixed value $\varepsilon = 0.33$ of the labor supply elasticity, and for three different values of the elasticity of substitution $\sigma \in \{0.6, 1.4, 3.1\}$. We start by plotting the optimal marginal tax rates as a function of types in Figure 4. The scale in the horizontal axis is measured in (current) incomes; e.g., the value of the optimal marginal tax rate at the notch $100,000 is that of a type $\theta$ who earns an income $y(\theta) = 100,000 in the calibration to the U.S. data – the income that this type earns in the optimal allocation is generally different (see below).

![Figure 4: Optimal marginal tax rates as a function of type](image)

The black dashed curve shows the marginal tax rates set by the partial equilibrium planner defined above (equation (53)), i.e., for an elasticity of substitution $\sigma \to \infty$. It follows a U-shape in line with previous results in this literature (Diamond, 1998; Saez, 2001).

We now show how these results change due to general equilibrium effects. The blue, red and yellow curves show the optimal general equilibrium marginal tax rates for the three values of $\sigma$. The results are in line with our theoretical derivations in Section 3.3. First of all, the top tax rate is lower, more so when the the elasticity of substitution is lower (see Corollary 5). Second, the decrease in marginal tax rates is even larger for income levels around $100,000, i.e., in the lower part of the U where the marginal tax rates in partial equilibrium were already the lowest. Third, for low income levels (below about $40,000 which is slightly below the median household income
in the U.S.), by contrast, partial equilibrium forces lead to very high marginal tax rates for low incomes, which implies that the general equilibrium corrections here are quantitatively very small (at most 1.8 percentage points) but they imply higher marginal tax rates for very low income levels. We further examine this result in Section 4.3.3. To sum up, Figure 4 shows that the well-known optimal U-shaped pattern of marginal tax rates is not only preserved but becomes more pronounced in general equilibrium. At the same time, the optimal taxes on those below the median U.S. income are practically unaffected.

We further examine this result in Section 4.3.3. To sum up, Figure 4 shows that the well-known optimal U-shaped pattern of marginal tax rates is not only preserved but becomes more pronounced in general equilibrium. At the same time, the optimal taxes on those below the median U.S. income are practically unaffected.

Alternatively, in Figure 5, we plot marginal tax rates as a function of incomes in the optimal allocation rather than as a function of skill types. Marginal tax rates in this graph reflect the policy recommendations of the optimal tax exercise which is to set marginal tax rates at each income (rather than unobservable productivity) level. A general pattern is that the marginal tax rate schedule is shifted to the left because individuals work less for optimal taxes than current taxes. This is visible most clearly for the top bracket and the bottom of the U that start earlier in Figure 5 than in Figure 4.

**Welfare gains.** The welfare gains of moving from the optimal partial equilibrium taxes to the fully optimal (general equilibrium) taxes can be large. Figure 6 plots the welfare gains in consumption equivalent – which can also be interpreted as a uniform increase in the lump sum transfer. Naturally these gains are lower for higher values of $\sigma$ (recall that the partial equilibrium corresponds to $\sigma \to \infty$, where by definition these gains are zero). For low values of $\sigma$, the gains can be as high as 4 percent.

\[^{29}\text{Consistent with our construction above, for the partial equilibrium planner we calculate incomes for that tax schedule under the naïve assumption that wages are constant.}\]
and they remain nontrivial for the whole range of plausible parameters.

![Graph showing welfare gains as a function of the elasticity of substitution.](image)

**Figure 6:** Welfare gains as a function of the elasticity of substitution

The role of the elasticity of labor supply. In this paragraph we fix the constant elasticity of substitution to $\sigma = 1.4$ and vary the value of the structural labor supply elasticity $\varepsilon \in \{0.2, 0.4, 0.6\}$. All the results are illustrated in Figure 7. The bold curves illustrate the respective optimal tax rates in general equilibrium, and the dashed curves illustrate the partial equilibrium counterparts. Consistent with our theoretical results in Sections 3.3.2 and 3.3.3, we find that (i) the magnitude of the correction to optimal taxes due to general equilibrium forces is increasing in the structural labor supply elasticity; and (ii) optimal marginal tax rates in general equilibrium are more sensitive to the value of the elasticity.

### 4.3.3 General equilibrium wedge accounting

We now address the question of what drives the differences between partial and general equilibrium optimal taxes, highlighted in Figures 4, 5 and 7. Recall that the partial equilibrium tax rates are based on (53) and the optimal general equilibrium taxes are based on equation (51). There are three differences: (i) the additional general equilibrium term $(g(y^*) - 1)/\sigma$ in the general equilibrium tax formula (51); (ii) the use of general equilibrium rather than partial equilibrium elasticities ($\tilde{E}$ vs. $\tilde{\varepsilon}$); and (iii) the fact that the formulas are evaluated at different wage distributions. We now present a wedge accounting analysis that cleanly decomposes the gap between the partial and the general equilibrium optima into these different components.
Proposition 6. The optimal marginal tax rate of type $\theta$ can be expressed as a function of $\tau_{PE}(\theta)$ and three additional terms:

\[
\frac{\tau(\theta)}{1-\tau(\theta)} = \frac{\tau_{PE}(\theta)}{1-\tau_{PE}(\theta)} + \left(\frac{g(\theta) - 1}{\sigma}\right) + \left(1 + \frac{1}{\varepsilon}\right) \left(1 - \bar{g}(\theta)\right) \left(\frac{\tilde{E}_{l,1-\tau}(\theta)}{E_{l,1-\tau}(\theta)} - 1\right)
\]

\[
+ \left(1 + \frac{1}{\varepsilon}\right) \left(1 - \bar{g}(\theta)\right) \left(\frac{1 - F_w(w(\theta))}{f_w(w(\theta))w(\theta)} - \frac{1 - F_{w_d}(w_d(\theta))}{f_{w_d}(w_d(\theta))w_d(\theta)}\right)
\]

\[
(54)
\]

Proof. See Appendix. \qed

Figure 7: The role of the structural labor supply elasticity

Formula (54) shows that the difference between the partial and general equilibrium optimum taxes can be decomposed into three terms: (i) the general equilibrium cross-wage effect, analyzed in Corollary 4 (note that this correction is always positive for a Rawlsian planner, i.e., leads to lower tax rates); (ii) the general equilibrium own-wage effect, also analyzed in Corollary 4, which affects the relevant labor supply elasticity, $\tilde{E}_{l,1-\tau}(\theta)$ vs. $\tilde{\varepsilon}_{l,1-\tau}(\theta)$ (note that this correction is always positive, i.e., leads to higher tax rates); and (iii) the adjustment in the wage distribution (the distribution at the GE optimum, $f_{w}(w(\theta))$, differs from that inferred from the data, $f_{w_d}(w_d(\theta))$). The latter adjustment is expressed as a function of the difference in the hazard rates of the two distributions.\(^{30}\)

\(^{30}\)With endogenous marginal social welfare weights, there would be an additional correction term that would account for the fact that the welfare weights are different in the optimal partial and general equilibrium tax schedules.
We now quantitatively decompose the relative importance of each cause of departure of the
general equilibrium optimum from the partial equilibrium counterpart. We look at our intermediate
case $\varepsilon = 0.33$ and $\sigma = 1.4$. The accounting analysis is depicted in Figure 8. The black dashed curve
is the partial equilibrium benchmark. The bold red curve illustrates optimal general equilibrium
tax rates. The yellow, green, and purple curves illustrate suboptimal general equilibrium tax rates
where the cross wage term, the elasticity correction term and the hazard rate correction term are
ignored, respectively. As anticipated from the theoretical result in Proposition 6, ignoring the
general equilibrium cross wage term leads to marginal tax rates that are too high. The opposite is
true for the elasticity correction term. Both effects are quantitatively important, but the Stiglitz
term dominates. Finally, this graph shows that the hazard rate correction term ($iii$) is quantitatively
of minor importance.

![Figure 8: General equilibrium wedge accounting for $\sigma = 1.4$](image)

### 4.4 Reconciling the tax incidence and optimal taxation results

The results of Sections 4.2 and 4.3 may seem contradictory at first sight. On the one hand, Corol-
lary 3 and Figure 3 show that the general equilibrium effects make it desirable (in terms of tax
revenue) to *increase* the marginal tax rates on high incomes, relative to the partial equilibrium
benchmark, if we start from the *current* U.S. tax system. On the other hand, Corollaries 4 and 5
and Figures 4, 5, and 7 imply that *optimal* tax rates on high incomes are strictly *lower* than the
partial equilibrium optimum (as defined in equation (53)).

This is a striking result: whether taking into consideration the general equilibrium forces calls
for a more or less progressive tax schedule crucially hinges on the baseline tax schedule we start
with. To reconcile the results for the optimum with those from the incidence analysis, we apply a tax incidence analysis to the optimal partial equilibrium tax schedule according to (53). The tax rate that a partial equilibrium planner would set is illustrated in Figure 5 (black dashed curve). Taking this tax schedule as the baseline and plotting the counterpart of Figure 3 yields the left panel. Comparing these two graphs shows that the incidence results are completely overturned depending on whether the low-income marginal tax rates are low (as in the U.S. tax code) or high (as in the partial equilibrium optimum), consistent with our discussion of Figure 3.

First of all note that the red bold curve is exactly at zero in the left panel. Given that the tax schedule to which we apply the incidence analysis is set to maximize tax revenue under the assumption that there are no general equilibrium effects, the red bold line must by definition be equal to zero. There is no possibility to raise further tax revenue according to the partial equilibrium considerations. The black dashed line shows that taking into account general equilibrium effects calls for lower tax rates for intermediate and high incomes and higher marginal tax rates for low incomes. This analysis also helps to understand our result that the U-shape is more pronounced with general equilibrium effects from a tax incidence perspective.

In the right panel, we also illustrate the tax incidence analysis starting from the optimal general equilibrium tax schedule, i.e. the red bold curve in Figure 5. Here, by construction the black dashed curve is zero at each income level. Further the red bold curve has an inverted shape compared to the black dashed curve in the left panel. To sum up, the effects of reforming marginal tax rates crucially depend on the benchmark tax schedule we use, and the results that hold at the optimum tax code may actually be reversed if the goal is to reform the current tax system.

5 Conclusion

In this paper we bring together two strands in the taxation literature: the study of tax incidence in general equilibrium, and that of nonlinear income tax design. The difficulty in the analysis of the
nonlinear income tax in general equilibrium is that reforms of the tax schedule lead to an infinite sequence of feedback effects that affect all of the agents. Thus solving for the key element of the tax incidence problem – the change in labor supplies of all agents – becomes very challenging. We show that one can mathematically formalize this tax incidence problem as an integral equation. We then use the theory of integral equations to provide a comprehensive characterization of both the incidence effects of tax reforms, and of the optimal tax schedule. We show that the general equilibrium forces may substantially affect the results obtained in partial equilibrium. Moreover, we show that analyzing tax reform of a suboptimal (e.g., the U.S.) tax code may reverse the insights that are valid at the optimum.

References


CHEN, L. AND C. ROTHSCILD (2015): “Screening with endogenous preferences,” *Available at SSRN 2617834*.


A primer on tax incidence

In this section we summarize some of the classical results on tax incidence in a framework with one good, two factors of production (typically labor and capital, see e.g. (Kotlikoff and Summers, 1987; Salanie, 2011)), and linear taxation of these two factors. We consider instead the case where the two factors are high-skilled and low-skilled labor, as our primary goal is to study a model with a continuum of labor inputs.

A.1 Equilibrium

Individuals have preferences over consumption \( c \) and labor supply \( l \) given by

\[
U(c, l) = u(c - v(l)),
\]

where the functions \( u \) and \( v \) are twice continuously differentiable and strictly increasing, \( u \) is concave, and \( v \) is strictly convex. In particular, note that there are no income effects on labor supply. There are two skill levels \( \theta_1 < \theta_2 \), with respective masses \( F_1 \) and \( F_2 = 1 - F_1 \).

An individual of type \( \theta_i \) earns a wage \( w(\theta_i) \), which she takes as given. She chooses her labor supply \( l(\theta_i) \) and earns taxable income \( y(\theta_i) = w(\theta_i)l(\theta_i) \). The government levies linear income taxes \( \tau_i \) on income of type \( i \in \{1, 2\} \). Individual \( \theta_i \) therefore solves the following problem:

\[
l(\theta_i) = \arg \max_{l \in \mathbb{R}_+} u[(1 - \tau_i)w(\theta_i)l - v(l)].
\]

The optimal labor supply \( l(\theta_i) \) chosen by an individual \( \theta_i \) is thus the solution to the first-order condition:

\[
w(\theta_i) = \frac{v'(l(\theta_i))}{(1 - \tau_i)}.
\]

We denote by \( U(\theta_i) \) the indirect utility function attained by individual \( \theta_i \). Finally, the total amount of labor supplied by individuals of type \( \theta_i \) is denoted by \( L(\theta_i) \equiv l(\theta_i)F_i \). Equation (55) defines a decreasing "supply curve" in the plane \((w, l)\).

There is a continuum of identical firms that produce output using both types of labor \( \theta_1, \theta_2 \). The resulting aggregate production function \( \mathcal{F} \) is defined as:

\[
\mathcal{Y} = \mathcal{F}(L(\theta_1), L(\theta_2)).
\]

We assume that the production function has constant returns to scale. The representative firm chooses the (relative) demand of inputs (labor of each type), taking as given the wages \( w(\theta_i) \), to maximize its profit

\[
\max_{L_1, L_2} \left[ \mathcal{F}(L_1, L_2) - \sum_{i=1}^{2} w(\theta_i)L_i \right].
\]

As a result, in equilibrium the firm earns no profits and the wage \( w(\theta) \) is equal to the marginal productivity of the type-\( \theta \) labor, i.e.,

\[
w(\theta_i) = \mathcal{F}'_i(L(\theta_1), L(\theta_2)),
\]

(56)
for all $i \in \{1, 2\}$, where $\mathcal{F}_i'$ denotes the partial derivative of the production function with respect to its $i^{th}$ variable.

The equilibrium wages and quantities are derived by equating (56), which are infinitely elastic demand curves and (55) which are increasing supply curves.

A.2 Elasticity concepts

We first define the structural labor supply elasticity $\varepsilon_i = \varepsilon_{l,1-\tau}(\theta_i)$ as the change in the labor supply of individuals of type $\theta_i$ when the tax rate on their income, $\tau_i$, is increased. We let, for $i \in \{1, 2\}$,

$$\varepsilon_i = \left. \frac{\partial \ln l(\theta_i)}{\partial \ln (1 - \tau_i)} \right|_{w(\theta)} = \frac{v'(l(\theta_i))}{l(\theta_i) v''(l(\theta_i))},$$

(57)

where the second equality is proved in the Appendix. Note that this is a elasticity in partial equilibrium, since the change in labor supply is computed for a constant individual wage.

Next, we define the wage elasticity $\gamma_{ij} = \gamma(\theta_i, \theta_j)$ as the effect of a marginal increase in the labor supply of type $\theta_j$, $L(\theta_j)$, on the wage of type $\theta_i$, $w(\theta_i)$, keeping the labor supply of type $k \neq j$, $L(\theta_k)$, constant. That is, for $(i, j) \in \{1, 2\}^2$,

$$\gamma_{ij} = \left. \frac{\partial \ln w(\theta_i)}{\partial \ln L(\theta_j)} \right|_{w(\theta)} = \frac{L(\theta_i) \mathcal{F}_{i,j}''(L(\theta_1), L(\theta_2))}{\mathcal{F}_i'(L(\theta_1), L(\theta_2))}. $$

(58)

A.3 Tax incidence

To analyze the tax incidence problem in the simple two-type model laid out in this section, we start by deriving the general effects of arbitrary infinitesimal perturbations (“tax reforms”) $(d\tau_1, d\tau_2)$ of the baseline tax system $(\tau_1, \tau_2)$. Denote by $d\hat{l}_i = \frac{dl_i}{l_i}$ and $d\hat{w}_i = \frac{dw_i}{w_i}$ the implied percentage changes in labor supplies and wages induced by the tax reform. We show in the Appendix that we can express the labor supply effects as

$$
\begin{pmatrix}
  d\hat{l}_1 \\
  d\hat{l}_2
\end{pmatrix} = -\begin{pmatrix}
  \frac{1}{\varepsilon_1} & 0 \\
  0 & \frac{1}{\varepsilon_2}
\end{pmatrix} - \begin{pmatrix}
  \gamma_{11} & \gamma_{12} \\
  \gamma_{21} & \gamma_{22}
\end{pmatrix}^{-1} \begin{pmatrix}
  \frac{d\tau_1}{1 - \tau_1} \\
  \frac{d\tau_2}{1 - \tau_2}
\end{pmatrix}
$$

(58)

and the wage effects as

$$
\begin{pmatrix}
  d\hat{w}_1 \\
  d\hat{w}_2
\end{pmatrix} = \begin{pmatrix}
  \gamma_{11} & \gamma_{12} \\
  \gamma_{21} & \gamma_{22}
\end{pmatrix} \begin{pmatrix}
  d\hat{l}_1 \\
  d\hat{l}_2
\end{pmatrix}.
$$

(59)

The impact on individual welfare is given by

$$
dU_i = (1 - \tau_i) y_i U'_i \left[ -\frac{d\tau_i}{1 - \tau_i} + d\hat{w}_i \right].
$$

(60)

\[31\] Note that with the utility function $U$ that we consider, the Marshallian (uncompensated) and Hicksian (compensated) elasticities are identical.
Finally the impact on government revenue is given by

$$d\mathcal{R} = \sum_{i=1}^{2} (F_i y_i) d\tau_i + \sum_{i=1}^{2} (F_i \tau_i y_i) d\hat{l}_i + \sum_{i=1}^{2} (F_i \tau_i y_i) d\hat{w}_i.$$  \hfill (61)

Equation (58) shows that the changes in labor supplies \((d\hat{l}_1, d\hat{l}_2)\), induced by the changes in marginal tax rates \((d\tau_1, d\tau_2)\), are given by the sum of: (i) the partial equilibrium effects, captured by the diagonal matrix of labor supply elasticities \(\varepsilon_1\) and \(\varepsilon_2\) in equation (58); and (ii) the general equilibrium effects, coming from the fact that the initial labor supply changes trigger own- and cross-changes in wages, which in turn affect labor supplies, etc. This infinite sequence of feedback effects between equilibrium wages and labor supplies is captured by the inverse of the matrix of wage elasticities \((\gamma_{ij})_{i,j\in\{1,2\}}\) in equations (58) and (59).

Equation (60) shows that individual utilities are affected in two ways. First, their income (and hence utility) is directly affected by the changes in taxes \(d\tau_i\), holding wages and labor supply fixed. Second, their income is indirectly affected, holding labor supply fixed. Note that the endogenous change in labor supply has no first-order impact on utility by the envelope theorem.

The first term in (61) is the mechanical effect of the perturbation, i.e. the change in government revenue due to the change in tax rates, assuming that both labor supply behavior and wages remain constant. The second term is the behavioral effect, due to the change in labor supplies which induces a change in government revenue proportional to the marginal tax rate \(\tau_i\). The third term is the general equilibrium effect, coming from the fact that perturbing the marginal tax rates impacts individual wages and hence government revenue directly. The analysis so far thus shows that the size and the sign of the impacts of a particular tax reform \((d\tau_1, d\tau_2)\) depends generally on the baseline tax rates \((\tau_1, \tau_2)\), individual preferences (through the elasticities \((\varepsilon_1, \varepsilon_2)\)), and the production technology (through the elasticities \((\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})\)).

We now focus on the particularly simple case where one of the factors, say the labor supply of type \(\theta_2\), is in fixed supply, i.e. \(\varepsilon_2 = 0\). More precisely, we consider the limit of the previous expression as the elasticity of labor supply of type \(\theta_2\) is positive but small, so that \(l_2\) is always paid its marginal productivity. Suppose in addition that we perturb only the tax on the elastic factor, \(l(\theta_1)\). This amounts to reducing the model we have analyzed so far to its partial equilibrium limit, where the effects of the tax on \(\theta_1\) affect only the quantity of that factor (since \(\theta_2\) is in fixed supply).

In this case, we show in the Appendix by directly inverting the matrix in (58) and letting \(\varepsilon_2 \to 0\) that we obtain the following familiar result (see equation (2.6) in Kotlikoff and Summers (1987) and Section 1.1.1 in Salanié (2003)): the impact on the wage of types \(\theta_1\) and \(\theta_2\) is given by

$$d\hat{w}_1 = -\frac{w_2 L_2}{w_1 L_1} \quad \text{and} \quad d\hat{w}_2 = \frac{\varepsilon_1}{\varepsilon_1 - \gamma_{11}^{-1}} \frac{d\tau_1}{1 - \tau_1}. \hfill (62)$$

\(^{32}\)Using the matrix equality \((I - A)^{-1} = \sum_{n=0}^{\infty} A^n\) (assuming that the norm of the matrix is smaller than 1), we can solve equation (58) to express \((d\hat{l}_1, d\hat{l}_2)\) as an infinite sum, the economic interpretation of which is the sequence of feedback effects on labor supply coming from the general equilibrium wage effects. This expression parallels (20) in Section 2 for the case of a continuum of labor inputs and general nonlinear tax reforms.
Moreover, the impact on the labor supply of type $\theta_1$ is given by

$$d\hat{l}_1 = \frac{\varepsilon_1 \gamma_{11}^{-1}}{\varepsilon_1 - \gamma_{11}^{-1}} \frac{d\tau_1}{1 - \tau_1}.$$  

Formula (62) shows that the effect on the own wage $w_1$ of an increase in the tax rate $\tau_1$ on factor $\theta_1$, depends on the elasticity of supply of factor 1, $\varepsilon_1$, and the elasticity of the demand of factor 1, $-1/\gamma_{11} > 0.33$ This expression implies that the gross wage increases all the more that demand is less elastic relative to supply. In other words, when labor supply of a given type is much less elastic than labor demand ($\varepsilon_1 \ll |\gamma_{11}^{-1}|$), the cost of labor hardly changes and workers of type $\theta_1$ bear almost the full burden.

**B Proofs**

To be added.

---

33$1/\gamma_{11}$ is the standard elasticity of demand when the supply of factor $\theta_2$ is fixed, as it is constructed as the change in the labor supply $l(\theta_1)$ induced by a change in the wage $w(\theta_1)$, keeping $l(\theta_2)$ fixed.