Annual Review of Economics

Dynamic Taxation

Stefanie Stantcheva[1,2,3]

[1]Department of Economics, Harvard University, Cambridge, Massachusetts 02138, USA; email: sstantcheva@fas.harvard.edu

[2]Center for Economic and Policy Research, Washington, DC 20009, USA

[3]National Bureau of Economic Research, Cambridge, Massachusetts 02138, USA

Annu. Rev. Econ. 2020. 12:801–31

First published as a Review in Advance on May 5, 2020

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-100119-013035

JEL codes: H21, H23, H24, H41

Keywords

taxation, dynamic taxes, capital taxes, labor taxes, life cycle, human capital, equity, fairness, Mirrlees, Ramsey

Abstract

This article reviews recent advances in the study of dynamic taxation, considering three main approaches: the dynamic Mirrlees, the parametric Ramsey, and the sufficient statistics approaches. In the first approach, agents' heterogeneous abilities to earn income are private information and evolve stochastically over time. Dynamic taxes are not restricted ex ante and are set for redistribution and insurance considerations. Capital is taxed only in order to improve incentives to work. Human capital is optimally subsidized if it reduces posttax inequality and risk on balance. The Ramsey approach specifies ex ante restricted tax instruments and adopts quantitative methods, which allow it to consider more complex and realistic economies. Capital taxes are optimal when age-dependent labor income taxes are not possible. The newer and tractable sufficient statistics approach derives robust tax formulas that depend on estimable elasticities and features of the income distributions. It simplifies the transitional dynamics thanks to a newly defined criterion, the utility-based steady-state approach, which prevents the government from exploiting sluggish responses in the short run. Capital taxes are here based on the standard equity-efficiency trade-off.

1. INTRODUCTION

This article reviews recent advances in the study of dynamic taxation. Dynamic taxation has been addressed using three approaches in the literature: the dynamic Mirrlees, the parametric Ramsey, and the sufficient statistics approaches. Considering each in turn, I describe some of their main methods, derivations, and results; review some key papers; and discuss various extensions.

The dynamic Mirrlees approach covered in Section 2 adopts the static Mirrlees's (1971) idea that agents' heterogeneous abilities to earn income are private information and lets productive ability evolve stochastically over time. The ability to earn income can be stochastic among other reasons because of health shocks, shocks to one's human capital, individual labor market idiosyncrasies, or luck. Thus, agents not only start with heterogeneous skills but also face uncertainty over their life cycles. The dynamic optimal tax problem is a mix between an optimal redistribution across initial heterogeneity and an optimal insurance problem to smooth consumption over time. The tax instruments are not restricted a priori; rather, the goal is to solve for the optimal constrained efficient allocations subject to the informational constraints, and then look for possible decentralized tax implementations. Section 3 describes recent extensions to the dynamic Mirrlees approach that endogenize wages through human capital investments.

The parametric and quantitative Ramsey approach covered in Section 4 parametrically specifies the tax instruments to use and quantitatively (or, more rarely, analytically) solves for the optimal policies. The environments considered are often more complex than in dynamic Mirrlees models, featuring overlapping generations, credit constraints, and incomplete markets, public goods, or human capital investments. Some questions answered are what forces quantitatively matter for the optimal levels and progressivity of taxes, how important tagging by age is, and whether positive capital taxes are optimal.

The newer sufficient statistics approach to dynamic taxation described in Section 5 derives robust tax formulas that depend on estimable factors such as the elasticity of supply of capital or labor income with respect to their tax rate, the shape of the capital and labor income distributions, and the social welfare weights at different levels of labor or capital income. This approach also simplifies the transitional dynamics thanks to a newly defined criterion, the utility-based steady-state approach, that essentially prevents the government from exploiting sluggish responses in the short run, which is normatively more appealing and circumvents commitment issues. This approach is very tractable and empirically applicable. It allows addressing policy-relevant questions, which are much harder to answer in more complex models: These are, for instance, nonlinear capital taxation, income shifting between the capital and labor tax bases, heterogeneity in types of capital assets and in individual returns to capital or preferences, or broader social fairness and equity concerns.

One of the main findings from the dynamic Mirrlees literature is that taxes will be optimally smoothed over the life cycle, featuring a persistent component that depends on last period's taxes and a drift term that captures the insurance motive. Their mean-reversion or persistence will closely mimic that of the underlying stochastic skill process. In many settings, although the full implementations are complex, age-dependent linear taxes appear to reap most of the welfare gains from the constrained efficient allocations, although it is not clear how robust this result would be to different stochastic processes, preferences, or social objectives than those typically studied in the literature. Savings are typically discouraged at the optimum relative to the free-savings case because higher levels of assets and lower work effort are complements. This is the inverse Euler logic that arises when labor effort is not observed and needs to be incentivized.

With human capital added to the dynamic Mirrlees model, it is optimal to subsidize human capital investments on net if and only if they do not benefit high-ability agents disproportionately;

if human capital investments disproportionately benefit already high-ability agents, they increase posttax inequality and tighten high-ability agents' incentive constraints and should be taxed on net. When human capital investments take the form of time (training) rather than resource (money) investments, the key parameter is how substitutable or complementary they are to labor effort, i.e., whether there is learning-and-doing or rather learning-or-doing.

The sufficient statistics approach delivers the key insight that the same linear or nonlinear formulas as in a static setting apply, but using the long-run elasticities in lieu of short-run ones. This greatly simplifies the study of dynamic taxation and permits making use of the many results already derived for income taxation in static settings. To give some examples, a restricted, comprehensive income tax that does not differentiate between capital and labor income is optimally set as in the static Mirrlees case, simply using the elasticities and distribution of total income. Such a tax system is optimal if there is a lot of income shifting between capital and labor income. Similarly, one can directly "plug in" different generalized social welfare weights that directly capture a broad set of justice and fairness considerations, as in the static framework by Saez & Stantcheva (2016). For instance, if the wealth distribution is considered fair, a zero capital tax rate will be optimal for equity reasons. If, on the other hand, wealth is a tag for parental background and equality of opportunity is valued, then a positive capital tax will be optimal.

These different approaches offer very distinct reasons for taxing or not taxing capital, among other things. With restricted instruments in the parametric Ramsey approach, capital taxation is often an imperfect substitute for missing age-dependent taxes and transfers. In the dynamic Mirrlees approach, capital is taxed in order to provide more efficient labor supply incentives when there is imperfect information and as part of the optimal insurance scheme against stochastic earnings abilities. In the sufficient statistics approach, it is made clear that capital income is taxed based on the standard equity–efficiency trade-off familiar from the static income tax literature. Capital taxes will be positive for redistribution (equity) reasons as long as capital income is concentrated among agents with relatively lower social welfare weights (typically, higher-income agents) and the elasticity of capital to taxes (the efficiency cost) is not infinite. This is the case, for instance, in a model with wealth in the utility function that generates finite elasticities of capital income to taxes and a nondegenerate steady state. In the latter, heterogeneous wealth holdings arise from heterogeneous returns and preferences for wealth.

2. THE DYNAMIC MIRRLEES APPROACH

This section is based on the core models and methods developed by Farhi & Werning (2013), Golosov et al. (2016), and Kapicka (2013) and characterizes the optimal labor and savings distortions over the life cycle. We start with a dynamic life cycle model that features a Markov skill process. The persistence of types poses particular challenges and also significantly affects the shape of the optimal policies.[1] The methodology is laid out in some detail for clarity.

The early papers in the dynamic Mirrlees literature focused on the savings distortion. Golosov et al. (2003) demonstrate that it is optimal to distort savings downward for a general class of economies with stochastic skills. This follows the core inverse Euler logic that is described below and is already apparent in the more specialized settings by Diamond & Mirrlees (1978) and Rogerson (1985). Kocherlakota (2005) shows that it also holds with aggregate shocks. Farhi & Werning (2010) highlight that the capital distortions prescribed by the inverse Euler logic play only a quantitatively small role in improving welfare once general equilibrium effects are taken

[1] All detailed derivations can be found in the appendix of Farhi & Werning (2013) or Stantcheva (2017).

into account. For labor taxes, Werning (2002) derives an optimal smoothing result for income taxes across time and states.

2.1. A Dynamic Life-Cycle Model

The economy consists of agents who live for T years, during which they work and consume.[2] Agents who work $l_t \geq 0$ hours in period t at a wage rate w_t make income $y_t = w_t l_t$. The disutility cost to an agent of supplying labor effort l_t is $\phi_t(l_t)$. ϕ_t is strictly increasing and convex. The wage rate w_t is simply equal to "ability" θ_t, a catch-all term to capture the ability to earn income, such as innate ability, skill, or labor market opportunities. For generality, I often maintain the notation $w_t(\theta_t)$, which facilitates the transition to extensions such as endogenous wages through human capital accumulation. There is a physical capital asset that yields a fixed gross interest rate R. Investments in this physical capital are called savings. The analysis is in partial equilibrium in that the wage function and the interest rate are taken as given.

Agents are born at time $t = 1$ with a heterogeneous earning ability θ_1 that has distribution $f^1(\theta_1)$. Earning ability evolves according to a Markov process with a possibly time-varying transition function $f^t(\theta_t|\theta_{t-1})$ over a fixed support $\Theta \equiv [\underline{\theta}, \bar{\theta}]$.

Agents' per-period utility is separable in consumption and labor; that is, we have

$$\tilde{u}_t\,(c_t, y_t; \theta_t) = u_t\,(c_t) - \phi_t\left(\frac{y_t}{\theta_t}\right). \tag{1}$$

In this equation u_t is increasing, twice continuously differentiable, and concave. More complicated tax formulas arise when utility is nonseparable in consumption and labor.

Denote by θ^t the history of ability shocks up to period t, by Θ^t the set of possible histories at t, and by $P(\theta^t)$ the probability of a history θ^t; that is, we have $P(\theta^t) \equiv f^t(\theta_t|\theta_{t-1})\ldots f^2(\theta_2|\theta_1) f^1(\theta_1)$. An allocation $\{x_t\}_t$ specifies consumption and output, for each period t, conditional on the history θ^t, i.e., $x_t = \{x\,(\theta^t)\}_{\Theta^t} = \{c\,(\theta^t), y\,(\theta^t)\}_{\Theta^t}$. The expected lifetime utility from an allocation, discounted by a factor β, is given by

$$U\big(\{c\,(\theta^t), y\,(\theta^t)\}\big) = \sum_{t=1}^{T} \int \beta^{t-1}\left[u_t\,(c\,(\theta^t)) - \phi_t\left(\frac{y\,(\theta^t)}{\theta_t}\right)\right] P\,(\theta^t)\,d\theta^t, \tag{2}$$

where $d\theta^t \equiv d\theta_t \ldots d\theta_1$.

2.2. The Planning Problem

Every period, the planner can observe agents' output y_t and consumption choice c_t, but ability θ_t is never observable and neither is labor supply $l_t = y_t/\theta_t$. Hence, if an agent produces low output, the planner does not know whether it was labor effort or ability that was low.

In this technical part, I walk through a typical dynamic asymmetric information problem. Starting from the full sequential problem with incentive compatibility constraints, two key steps are taken:

1. The problem is turned into a relaxed program using a first-order approach that replaces the (infinite) set of incentive compatibility constraints with agents' envelope conditions.
2. The relaxed program is made recursive, using as state variables the promised utility and its gradient.

[2] A retirement period during which agents only consume is typically also added. It only changes the results quantitatively, as consumption is simply smoothed during this period, with no uncertainty and no labor supply decisions.

2.2.1. Incentive compatibility in a dynamic setting. Imagine a direct revelation mechanism in which, each period, agents report their current ability θ_t. Denote a reporting strategy, specifying a reported type r_t after each history, by $r = \{r_t(\theta^t)\}_{t=1}^T$, with $\mathcal{R}$ being the set of all possible reporting strategies and $r^t = \{r_1(\theta_1), \ldots, r_t(\theta^t)\}$ being the history of reports from this reporting strategy r. Allocations are specified as functions of the history of reports by the planner. Let the continuation value after history θ^t under a reporting strategy r, denoted by $\omega^r(\theta^t)$, be the solution to

$$\omega^r(\theta^t) = u_t(c(r^t(\theta^t))) - \phi_t\left(\frac{y(r^t(\theta^t))}{\theta_t}\right) + \beta \int \omega^r\left(\theta^{t+1}\right) f^{t+1}\left(\theta_{t+1}|\theta_t\right) d\theta_{t+1}.$$

The continuation value under truthful revelation, $\omega(\theta^t)$, is a solution to

$$\omega\left(\theta^t\right) = u_t(c\left(\theta^t\right)) - \phi_t\left(\frac{y(r^t(\theta^t))}{\theta_t}\right) + \beta \int \omega\left(\theta^{t+1}\right) f^{t+1}\left(\theta_{t+1}|\theta_t\right) d\theta_{t+1}.$$

Incentive compatibility means imposing that truth-telling yields at least weakly higher continuation utility than any other reporting strategy; that is, we have

$$(IC): \omega\left(\theta_1\right) \geq \omega^r\left(\theta_1\right) \quad \forall \theta_1, \forall r. \qquad 3.$$

Let X^{IC} be the set of incentive compatible allocations. To solve this dynamic problem, a version of the first-order approach is used (see the set of assumptions in Farhi & Werning 2013 and Stantcheva 2017).

Consider a history θ^t and one special deviation strategy $\tilde{r}_t$, under which the agent reports truthfully until period $t(\tilde{r}_s(\theta^s) = \theta_s \forall s \leq t-1)$ but reports $\tilde{r}_t(\theta^t) = \theta' \neq \theta_t$ in period t. Under this strategy, the continuation utility is the solution to

$$\omega^{\tilde{r}}(\theta^t) = u_t(c(\theta^{t-1}, \theta')) - \phi_t\left(\frac{y(\theta^{t-1}, \theta')}{\theta_t}\right) + \beta \int \omega^{\tilde{r}}\left(\theta^{t-1}, \theta', \theta_{t+1}\right) f^t\left(\theta_{t+1}|\theta_t\right) d\theta_{t+1}.$$

Incentive compatibility in Equation 3 implies that, after almost all θ^t, the temporal incentive constraint holds, and we obtain

$$\omega\left(\theta^t\right) = \max_{\theta'} \omega^{\tilde{r}}\left(\theta^t\right). \qquad 4.$$

Inversely, if Equation 4 holds after all θ^{t-1} and for almost al θ_t, then Equation 3 also holds (see Kapicka 2013, lemma 1). If we take the derivative of promised utility with respect to (true) ability, there are two direct effects, namely on the wage (higher types have higher wages) and on the Markov transition $f^t(\theta_t|\theta_{t-1})$, and indirect effects on the allocation through the report. By the first-order condition of the agent, all indirect effects are jointly zero, and only the two direct effects remain. This leads to the envelope condition of the agent, which is necessary for incentive compatibility:[3]

$$\dot{\omega}\left(\theta^t\right) := \frac{\partial \omega\left(\theta^t\right)}{\partial \theta_t} = \frac{w_{\theta,t}}{w_t} l\left(\theta^t\right) \phi_{l,t}\left(l\left(\theta^t\right)\right) + \beta \int \omega\left(\theta^{t+1}\right) \frac{\partial f^{t+1}\left(\theta_{t+1}|\theta_t\right)}{\partial \theta_t} d\theta_{t+1}. \qquad 5.$$

This envelope condition describes how the promised utility (or, equivalently, the informational rent) has to vary with respect to the true type for allocations to be incentive compatible. The

[3]This is an application of Milgrom & Segal (2002, theorem 2).

first term is the same as in the static screening model (Mirrlees 1971). It ensures that higher-ability types in the current period need to be compensated to reveal their information to the planner. The second is the future component of the information rent. An agent that reveals their type to the planner today needs to be compensated not only for the gain that they could extract by pretending to be another type today, but also for the gain they could extract in the future from doing so. This is due to the persistence in types: Today's type realization brings the agent information about their type tomorrow. In particular, this second term would disappear under independent and identically distributed (i.i.d.) shocks, when today's type realization carries no information about future realizations.

The planner's objective is to minimize the expected discounted cost of providing an allocation, subject to incentive compatibility as defined in Equation 3 and to the expected lifetime utility of each (initial) type θ being above a threshold $\underline{U}(\theta)$. Let $U(\{c, y\}; \theta)$ be lifetime utility as defined in Equation 2 for agents with initial type θ. The relaxed planning problem, denoted by P^{FOA}, replaces the incentive constraint by the envelope condition and is given by

$$[P^{\mathrm{FOA}}]: \quad \min_{\{c,y\}} \Pi\left(\{c,y\};\underline{U}(\theta)_{\Theta}\right) = \left[\sum_{t=1}^{T}\left(\frac{1}{R}\right)^{t-1}\int_{\Theta^t}\left(c\left(\theta^t\right) - y\left(\theta^t\right)\right)P\left(\theta^t\right)d\theta^t\right], \qquad 6.$$

$$\text{such that } U\left(\{c,y\};\theta\right) \geq \underline{U}(\theta),$$

$$y\left(\theta^t\right) \geq 0, c\left(\theta^t\right) \geq 0,$$

$$\{c,y\} \in X^{\mathrm{FOA}}.$$

The envelope condition is necessary, but not sufficient, for optimality. Unlike in the static Mirrlees model, it is not easy to find conditions on the primitives that guarantee that the first-order approach will deliver the optimum (Pavan et al. 2014). In general, incentive compatibility of the candidate allocation, as well as any omitted nonnegativity constraint, is checked numerically (see Farhi & Werning 2013 or Stantcheva 2017).

2.2.2. Recursive formulation of the relaxed program. Without persistent types, writing the planning problem recursively requires specifying the promised utility as a state variable. This ensures that in period t, the continuation value provided to the agent remains consistent with what was promised to the agent by the planner in period $t-1$ (and, working backwards, in earlier periods, too). Let $v(\theta^t)$ be the expected future continuation utility:

$$v\left(\theta^t\right) \equiv \int \omega\left(\theta^{t+1}\right) f^{t+1}\left(\theta_{t+1}|\theta_t\right)d\theta_{t+1}. \qquad 7.$$

Continuation utility $\omega(\theta^t)$ can hence be rewritten as

$$\omega\left(\theta^t\right) = u_t(c\left(\theta^t\right)) - \phi_t\left(\frac{y\left(\theta^t\right)}{\theta_t}\right) + \beta v\left(\theta^t\right). \qquad 8.$$

With persistence in types this is not sufficient, since the agent is promised not only a level of continuation value but also a variation in that continuation value with their type. Thus, we need to define the future marginal rent (the second term in the envelope condition) as

$$\Delta\left(\theta^t\right) \equiv \int \omega\left(\theta^{t+1}\right)\frac{\partial f^{t+1}\left(\theta_{t+1}|\theta_t\right)}{\partial\theta_t}d\theta_{t+1}. \qquad 9.$$

The envelope condition can then be rewritten as

$$\dot{\omega}(\theta^t) = \frac{w_{\theta,t}}{w_t} l(\theta^t)\phi_{l,t}(l(\theta^t)) + \beta\Delta(\theta^t). \tag{10.}$$

The state variables in any given period t are then the promised utility, the promised marginal utility, and the previous period's type realization—respectively, v_{t-1}, Δ_{t-1}, and θ_{t-1}. Because the shock process is Markov, θ_{t-1} is all that needs to be known for the continuation of the problem. The full history θ^{t-1} is not needed. The expected continuation cost of the planner at time t given these states is

$$K(v_{t-1}, \Delta_{t-1}, \theta_{t-1}, t) = \min\left[\sum_{\tau=t}^{T}\left(\frac{1}{R}\right)^{\tau-t}\int (c_\tau(\theta^\tau) - y_\tau(\theta^\tau))P\left(\theta^{\tau-t}\right)d\theta^{\tau-t}\right],$$

where, with some abuse of notation, we have $d\theta^{\tau-t} = d\theta_\tau d\theta_{\tau-1}\ldots d\theta_t$, and $P(\theta^{\tau-t}) = f^\tau(\theta_\tau|\theta_{\tau-1})\ldots f^t(\theta_t|\theta_{t-1})$. A recursive formulation of the relaxed program is then, for $t \geq 2$,

$$K(v, \Delta, \theta_-, t) = \min\int (c(\theta) - w_t(\theta)l(\theta) + \frac{1}{R}K(v(\theta), \Delta(\theta), \theta, t+1))f^t(\theta|\theta_-)d\theta, \tag{11.}$$

subject to

$$\omega(\theta) = u_t(c(\theta)) - \phi_t(l(\theta)) + \beta v(\theta),$$

$$\dot{\omega}(\theta) = \frac{w_{\theta,t}}{w_t} l(\theta)\phi_{l,t}(l(\theta)) + \beta\Delta(\theta),$$

$$v = \int \omega(\theta) f^t(\theta|\theta_-)d\theta,$$

$$\Delta = \int \omega(\theta)\frac{\partial f^t(\theta|\theta_-)}{\partial\theta_-}d\theta,$$

where the maximization is over the functions $(c(\theta), l(\theta), \omega(\theta), v(\theta), \Delta(\theta))$.

To incorporate redistribution concerns, we can interpret the initial type θ_1 as an arbitrary heterogeneity that agents start with. For each type θ_1, the planner specifies a target utility to be reached, $(\underline{U}(\theta))_\Theta$. In period $t = 1$, the planner's problem takes these target utilities as constraints.

2.2.3. Wedges. To characterize the optimal allocations, obtained as solutions to the relaxed program above, the literature has relied on wedges, which are implicit taxes and subsidies. For any allocation, define the intratemporal wedge on labor, $\tau_L(\theta^t)$, and the intertemporal wedge on savings (also called capital wedge), $\tau_K(\theta^t)$, as follows:

$$\tau_L(\theta^t) \equiv 1 - \frac{\phi_{l,t}(l_t)}{w_t u_t'(c_t)}, \tag{12.}$$

$$\tau_K(\theta^t) \equiv 1 - \frac{1}{R\beta}\frac{u_t'(c_t)}{E_t\left(u_t'(c_{t+1})\right)}. \tag{13.}$$

These wedges can be thought of as locally linear taxes. Absent government intervention, they would be equal to zero. They are thus a measure of the distortion at an allocation relative to the laissez-faire allocation. For instance, the labor wedge is defined as the gap between the marginal rate of substitution and the marginal rate of transformation between consumption and labor, which would be zero in the laissez-faire allocation when an agent optimizes their labor supply. Imagine the planner imposing a linear tax equal to $\tau_L(\theta^t)$ and letting an agent of type θ^t choose their labor supply locally around $l(\theta^t)$. Equation 12 is a necessary condition for the agent's labor supply choice. A positive labor wedge means that labor supply is distorted downwards. Similarly, the savings wedge τ_K measures the difference between the expected marginal rate of intertemporal substitution and the return on savings.

For the exposition, denote by $\varepsilon_{x,y,t}$ the elasticity of a variable x_t to another variable y_t, so that $\varepsilon_{x,y,t} \equiv d\log(x_t)/d\log(y_t)$. Let ε_t^u be the uncompensated and ε_t^c the compensated labor supply elasticities to the net wage, holding savings fixed.

2.3. Labor Income Taxation: Tax Smoothing, Persistence, and Age Patterns

The following proposition highlights the insurance and redistribution forces that drive the labor wedge.

Proposition 1. At the optimum, the labor wedge is equal to

$$\frac{\tau_{Lt}^*(\theta^t)}{1-\tau_L^*(\theta^t)} = \frac{\mu(\theta^t)\,u_t'(c(\theta^t))}{f^t(\theta_t|\theta_{t-1})}\frac{\varepsilon_{w\theta,t}}{\theta_t}\frac{1+\varepsilon_t^u}{\varepsilon_t^c} \tag{14.}$$

with $\mu(\theta^t) = \eta(\theta^t) + \kappa(\theta^t)$, where

$$\eta(\theta^t) = \frac{\tau_{Lt-1}^*(\theta^{t-1})}{1-\tau_{L,t-1}^*(\theta^{t-1})}\left[\frac{R\beta}{u_{t-1}'(c(\theta^{t-1}))}\frac{\varepsilon_{t-1}^c}{1+\varepsilon_{t-1}^u}\frac{\theta_{t-1}}{\varepsilon_{w\theta,t-1}}\int_{\theta_t}^{\bar{\theta}}\frac{\partial f(\theta_s|\theta_{t-1})}{\partial\theta_{t-1}}d\theta_s\right]$$

and

$$\kappa(\theta^t) = \int_{\theta_t}^{\bar{\theta}}(1-g_s)\frac{1}{u_t'(c(\theta^{t-1},\theta_s))}f(\theta_s|\theta_{t-1})\,d\theta_s \tag{15.}$$

$$\text{with } g_s = u_t'\left(c\left(\theta^{t-1},\theta_s\right)\right)\lambda_{t-1} \text{ and } \lambda_{t-1} = \int_{\underline{\theta}}^{\bar{\theta}}\frac{1}{u_t'(c(\theta^{t-1},\theta_m))}f(\theta_m|\theta_{t-1})\,d\theta_m.$$

The insurance motive is captured in $\kappa(\theta^t)$. g_s is the marginal social welfare weight on an agent of type θ_s, measuring the social value of one more dollar transferred to that individual, and $1/\lambda_{t-1}$ is the social cost of public funds at time t. The insurance motive would be zero with linear utility. The redistributive term $\eta(\theta^t)$ can be written recursively in terms of the previous period's labor wedge weighted by a measure of ability persistence. Recall that there can be a redistributive motive in the first period if there is initial heterogeneity.[4] This motive persists through $\eta(\theta^t)$, and the more so if types are more persistent. In one polar case, if θ_t is i.i.d., only the contemporaneous insurance motive $\kappa(\theta_t)$ plays a role. If, in addition to i.i.d. shocks, utility is linear in consumption, the optimal labor wedge is zero in all periods, except in the first one if there are different utility thresholds (i.e., different social welfare weights) for different agents.

[4]In the first period, heterogeneity in θ_1 leads to

$$\mu(\theta_1) = \int_{\theta_1}^{\bar{\theta}}\frac{1}{u_1'(c_1(\theta_s))}\left(1-\lambda_0(\theta_s)\,u_1'(c_1(\theta_s))\right)f(\theta_s),$$

where $\lambda_0(\theta_s)$ is the multiplier (scaled by $f(\theta_s)$) on type θ_s target utility. With linear utility we obtain $1 = \int_{\underline{\theta}}^{\bar{\theta}}\lambda_0(\theta_s)f(\theta_s)$.

We can also immediately see the counterpart to the classic "zero distortions at the top and bottom" result in the static literature, which here holds in every period; that is, we have $\tau^*_{Lt}(\theta^{t-1}, \bar{\theta}) = \tau^*_{Lt}(\theta^{t-1}, \underline{\theta}) = 0$, $\forall t$. This result no longer holds if there is moving support, i.e., if the upper and lower bounds $\bar{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$ depend on the past type realization (Farhi & Werning 2013).

A special case is the log autoregressive process with persistence p such that

$$\log(\theta_t) = p \log(\theta_{t-1}) + \psi_t, \qquad 16.$$

where ψ_t has density $f^{\psi}(\psi|\theta_{t-1})$, with $E(\psi|\theta_{t-1}) = 0$. In this case, we can rewrite the evolution of the labor wedge over time as

$$E_{t-1}\left(\frac{\tau_{Lt}}{(1-\tau_{Lt})}\frac{\varepsilon_{w\theta,t-1}}{\varepsilon_{w\theta,t}}\frac{\varepsilon^c_t}{1+\varepsilon^u_t}\frac{1+\varepsilon^u_{t-1}}{\varepsilon^c_{t-1}}\left(\frac{1}{R\beta}\frac{u'_{t-1}}{u'_t}\right)\right)$$
$$= \varepsilon_{w\theta,t-1}\frac{1+\varepsilon^u_{t-1}}{\varepsilon^c_{t-1}}Cov\left(\frac{1}{R\beta}\frac{u'_{t-1}}{u'_t}, \log(\theta_t)\right) + p\frac{\tau_{Lt-1}}{(1-\tau_{Lt-1})}. \qquad 17.$$

The risk-adjusted expectation for the labor wedge (on the left-hand side) depends on the past period's wedge weighted by the persistence in types p. Thus, the labor wedge's persistence or mean reversion reflects that of the stochastic ability process. In addition, the labor wedge evolves over the life of the agents according to a drift term that captures the insurance motive; if there is no risk in consumption, i.e., there is perfect consumption smoothing over time, the drift term is zero. Dynamic incentive compatibility implies a positive covariance between the growth of consumption of an agent and their productivity, because the government induces high-productivity agents to reveal their type by promising them higher consumption growth. As a result, the value of insurance increases and so does the labor wedge.

2.4. Capital Taxation: The Inverse Euler Equation Logic

The dynamic Mirrlees framework has strong implications for how capital and savings should be treated. The uncertainty in types over time and the inability to control labor supply force the planner to impose a distortion on savings to improve the provision of incentives to work.

Proposition 2. At the optimum, the inverse Euler equation holds:

$$\frac{R\beta}{u'_t(c(\theta^t))} = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'_{t+1}(c(\theta^{t+1}))} f^{t+1}(\theta_{t+1}|\theta_t)\, d\theta_{t+1}. \qquad 18.$$

Thus, at the optimum, the inverse marginal utility in the current period is equal to the expected inverse marginal utility in the next period. By the concavity of marginal utility and Jensen's inequality, it is thus the case that

$$u'_t(c(\theta^t)) < \beta R \int_{\underline{\theta}}^{\bar{\theta}} u'_{t+1}\left(c\left(\theta^{t+1}\right)\right) f^{t+1}(\theta_{t+1}|\theta_t)\, d\theta_{t+1}, \qquad 19.$$

i.e., agents are off their Euler equation and, in particular, would like to save more than they were made to save at the optimal allocation. Savings are downward distorted and there is a positive savings wedge τ_K. Another way to see this is that the desired labor supply at the optimal allocation is incompatible with free savings. Increasing savings in period t increases disposable income in period $t+1$. Unless utility is quasilinear, this implies an income effect on labor supply, and the agent is then tempted to work less. More savings in period t and lower labor supply in period

$t+1$ are complements. Ruling out such a deviation requires discouraging savings below the level that would occur at the free-market rate.

Another way to understand the inverse Euler equation is to consider the cost of providing utility—rather than consumption—in every period and state. Providing different utilities in different (reported) states is how incentives are provided. Switching to this utility metric, consider the change in the planner's expected resource cost above of moving resources from one period to the next. Pick a history θ^t and leave all allocations unchanged, except at node θ^t, where we perturb the utility by providing $\beta \cdot \triangle$ less utility in period t, and $\triangle$ more utility for all θ_{t+1} after history (θ^t); that is, let perturbed utilities be $u_t(c(\theta^t)) - \beta\triangle$ and $u_{t+1}(c(\theta^t, \theta_{t+1})) + \triangle$. If the original allocation is optimal, it has to minimize the resource cost of providing utility across time at $\triangle = 0$. That cost is

$$c\left(u_t - \beta\triangle\right) + \frac{1}{R}\int_{\underline{\theta}}^{\bar{\theta}} \left(c\left(u_{t+1}(\theta^t, \theta_{t+1}) + \triangle\right)\right) f^t\left(\theta_{t+1}|\theta^t\right) d\theta_{t+1}.$$

Its first-order condition (FOC) evaluated at zero equates the inverse marginal utility in period t (which is the resource cost of providing utility in period t) to the expected inverse marginal utility in period $t+1$ (which is the expected resource cost of providing utility in period $t+1$). This is exactly the inverse Euler equation. The standard Euler equation, on the other hand, equates the marginal utility of allocating resources across time, which is what a consumption-smoothing agent would do in the laissez-faire condition but not what a planner who seeks to efficiently allocate incentives would do under incentive constraints.

Note that in the special case in which there is uncertainty in the first period but not thereafter, we recover the result by Atkinson & Stiglitz (1976) that capital should not be taxed. As soon as shocks are introduced from the second period (or some period t) onward, we obtain the inverse Euler equation.

2.5. Implementation and Approximation

We now turn to the implementation of the optimal constrained efficient allocations using actual taxes and subsidies and discuss how the policies can be approximated using simpler tools.

2.5.1. Implementation: concepts and examples. The issue of implementation can be formulated as follows: Given the structure of the private market, are there (possibly complicated) tax functions such that, if agents were left in the decentralized economy facing these taxes, they would choose the allocations of the planner's problem? In the static Mirrlees problem, the answer is immediate under some regularity conditions: A nonlinear income tax can implement the constrained efficient allocation.

In the dynamic problem, although it is very helpful to think about wedges as implicit taxes and subsidies, the link between these implicit taxes and explicit tax functions is not immediate. This is because each wedge characterizes the distortion in one of the actions of the agent, holding the others fixed at the optimum; they are distortions along a single dimension. When an agent has several actions to take, there is scope for joint deviations.

In the case of the savings wedge, it may be intuitive to think that a savings tax equal to the wedge at each history as defined in Equation 13 would implement the allocation. However, the savings wedge characterizes the marginal intertemporal distortion, holding labor supply constant at its optimal level. It would implement the right amount of savings if the agent was choosing the optimal level of labor in period $t+1$. If the capital or savings tax was truly set to $\tau_K(\theta^t)$, the agent could jointly choose another level of savings and another level of labor supply in period $t + 1$, according to the logic described for the inverse Euler equation.

Albanesi & Sleet (2006) propose a simple implementation that works only in the case in which types are i.i.d., namely a nonlinear income tax that depends on the stock of wealth accumulated to date. With i.i.d. shocks, wealth acts as a sufficient statistic for the past history, providing all the information needed to give an agent their planned allocation.

Kocherlakota (2005) proposes an implementation with a nonlinear, fully history-contingent labor income tax and a linear capital tax rate in period t that conditions not only on the past history of income, but also on next period's $t + 1$ income. This rules out a joint deviation in savings and income. Although this savings tax raises on average no revenues, as Kocherlakota shows, it acts by making the return to savings stochastic. In fact, the return to savings is made higher for good realizations of the type. Thus, savings are turned into a worse hedge by the tax system, which discourages savings, as the inverse Euler equation requires.

With more structure on the type process, other implementations can be found. In Golosov & Tsyvinski's (2007) work, agents face the risk of permanent disability, which is an absorbing state. The optimal allocation is implemented with a transfer to agents with assets below a certain threshold, which is essentially a disability insurance scheme that has an asset test for benefits.

The implementation of a given allocation is in general not unique; however, all possible implementations generate the same marginal distortions as characterized by the wedges. Moreover, an implementation is always relative to a given market structure in the laissez-faire economy. For instance, in the planning problem, it can be ignored that agents face credit constraints because the incentive-compatible allocations directly specify private borrowing and saving. However, if the laissez-faire economy features credit constraints, the taxes and transfers that implement the planner's allocation will be different from the ones that implement it in the absence of credit constraints.

2.5.2. Approximations. Because the decentralized policies that implement an allocation are in general quite complex, the literature has sought to find simpler parametric approximations to the optimal policies. A recurring theme has been that linear, age-dependent tax rates can reap a large share of the welfare gain from the constrained efficient allocations. Weinzierl (2011) shows that moving from age-independent to age-dependent policies generates sizable welfare gains.

In practice, tax and transfer policies feature a nontrivial amount of age dependence. Sometimes this dependence is explicit, as is the case for old-age pensions; in other cases, it is implicit and due to factors that naturally evolve over the life cycle, such as marital status, children, and the life-cycle shape of income. **Supplemental Table 1** summarizes the many age-dependent features in tax policy in several OECD countries.

Supplemental Material >

Farhi & Werning (2013) go further and show that linear age-dependent policies approximate the optimal policies well in their life-cycle model; Stantcheva (2017) shows that this holds true even when there is endogenous human capital investment (in which case, a linear age-dependent human capital subsidy is also required). The linearity result is, however, likely not valid for all types of skill processes and may depend on the log-autoregressive processes assumed in the literature. While this has not been shown formally, the approximation with linear policies is likely to do more poorly if the variance of shocks is larger, if the planner is more redistributive, and if the agents are more risk averse. In a new model applying dynamic tax methods to the design of corporate taxation and innovation policies, with spillovers between firms, a linear corporate profit tax combined with a nonlinear research and development (R&D) subsidy also does very well (Akcigit et al. 2016).

2.6. Extensions of the Core Model

We now turn to several key extensions of the core model.

2.6.1. Further life-cycle considerations. Some papers have tried to incorporate a more realistic and complex life cycle in the dynamic Mirrlees problem, while simplifying the model along other dimensions, for example, by focusing on a two-period model only or on simpler shock structures. Best & Kleven (2013) incorporate career effects, whereby current work effort also affects future wages. Without age-dependent taxes, this tends to make the optimal tax schedule much less progressive; with age-dependent taxes, taxes at older ages should be lowered. In a model with endogenous retirement age, Ndiaye (2017) shows that retirement benefits that are increasing with age are needed in addition to the age-dependent linear taxes in order to achieve the welfare gain from the constrained efficient allocations.

2.6.2. Hidden savings. Ábrahám et al. (2016) propose a pure insurance model, in which agents are ex ante identical and in each period the type is realized after the labor choice has been made. In this case, labor and capital taxes are complements: A higher capital tax, by the inverse Euler logic, helps provide incentives for working more efficiently. Thus, if private savings cannot be taxed because they are unobservable, optimal labor taxes become less progressive (see also Ábrahám & Pavoni 2008 and Ábrahám et al. 2011).

2.6.3. Restrictions on asset taxes. A less stark case of restriction on capital taxes is to rule out nonlinear taxes, which happens if individual borrowing and lending are not observable but it is possible to (linearly) tax the observable overall capital stock, as proposed by da Costa (2009), da Costa & Werning (2002), and Golosov & Tsyvinski (2007). In Golosov & Tsyvinski's (2007) model, agents can privately trade in assets markets in order to self-insure. Whether capital should be taxed a (linear) positive rate depends on the shock process: With i.i.d shocks or absorbing disability shocks this is the case, but with more general skill processes, where the current skill realization grants the agent significant advance information about their future skill, it may be optimal to subsidize capital. Chang & Park (2018) derive a fully nonlinear income tax schedule in the presence of private endogenous insurance and find that the optimal nonlinear tax rates can be very different from those with no private insurance.

2.6.4. Innovation and externalities. A very recent development in this literature is to consider the taxation of firms. Akcigit et al. (2016) build a new and general framework to study the taxation of firms that captures key elements such as market power, investments, production, heterogeneity in productivity, intellectual property, and asymmetric information. In their model, firms invest in innovation, which has spillovers on other firms. Firms have heterogeneous research productivities, i.e., abilities to convert a given set of research inputs into innovation. Productivities are stochastic and are private information. Theoretically, they show how the Pigouvian subsidies on R&D and taxes on corporate income should be increased or decreased due to screening considerations and depending on the relative complementarities between observable R&D investments, unobservable R&D inputs, and firm productivity. Quantitatively, the model is estimated on the US Census's longitudinal business data matched to patent data, and it shows that implementing the constrained efficient allocation only requires using very simple policies. In particular, a nonlinear, separable Heathcote-Storesletten-Violante (HSV)-type subsidy (as described in Section 4) combined with an HSV-type profit tax reaps almost all of the benefits of the full optimum. It features lower marginal taxes for more profitable firms and lower marginal subsidies at higher R&D investment levels. In fact, making the profit tax linear only generates a small additional welfare loss.

3. ADDING HUMAN CAPITAL TO THE DYNAMIC MIRRLEES MODEL

As in the static Mirrlees model, the work presented in the previous section assumes that agents' wages are exogenous. A major development in dynamic life-cycle tax models has been to make the wages endogenous by modeling human capital accumulation. Investments in human capital—through money, time, or a mix of both—play a key role in shaping the skill and income distributions that ultimately drive the revenue-raising, redistribution, and insurance motives behind taxation. In turn, investments in human capital depend on their net return, which is affected by the tax and transfer system. In this section, I describe recent work that incorporates human capital and skill acquisition into the dynamic life-cycle tax model.

3.1. Literature

A first strand of the literature has focused on dynamic models with persistent heterogeneity across agents and without uncertainty or risk (Bohacek & Kapicka 2008, Kapicka 2013). Other papers have rather focused on the risk aspects of human capital, without incorporating heterogeneity (Anderberg 2009, Grochulski & Piskorski 2010). Findeisen & Sachs (2016) include both heterogeneity and uncertainty and focus on a one-shot investment during college, before the work life of the agent starts, with a one-time realization of uncertainty. Stantcheva (2017) extends the core dynamic Mirrlees model above, with heterogeneity and uncertainty, to include monetary (resource) investments in human capital over the life cycle. Stantcheva (2015a) instead considers time investment in human capital. Kapicka & Neira (2014) propose a human capital accumulation process with time investments and a fixed ability, and they consider the case in which the effort spent to acquire human capital is unobservable. Perrault (2015) considers loss of human capital from unemployment. Kapicka (2015) studies a Ben-Porath economy in which both ability and human capital are unobservable. Stantcheva (2015b) considers an intergenerational setting in which parents invest in children's human capital and can choose to transfer resources in the form of bequests and human capital. Koeniger & Prat (2018) study a very similar setting and provide quantitative results. Makris & Pavan (2019) consider income taxation when there is learning-by-doing, i.e., human capital acquisition happens as a by-product of working.

3.2. Model

The general setup is as in Section 2, but the wage that an agent receives is now a function not only of their stochastic type but also of their human capital. Each period, agents can build their stock of human capital s_t by spending money; below we also consider time investments. A monetary investment $M_t(e_t)$ increases human capital by $e_t \geq 0$. The cost function is increasing and convex, and we have $M_t'(0) = 0$. Human capital s_t evolves according to $s_t = s_{t-1} + e_t$. The wage w_t is determined by the stock of human capital built until time t and stochastic ability θ_t:

$$w_t = w_t(\theta_t, s_t).$$

Here, w_t is strictly increasing and concave in each of its arguments, but no restrictions are placed on the cross-partials. As the wage can depend on age, human capital could yield different returns at different ages.

Let $w_{m,t}$ denote the partial of the wage function with respect to argument m ($m \in \{\theta, s\}$), and $w_{mn,t}$ denote the second-order partial with respect to arguments $m, n \in \{\theta, s\} \times \{\theta, s\}$. A key parameter turns out to be the Hicksian coefficient of complementarity between ability and human capital in the wage function at time t (Hicks 1970, Samuelson 1974), denoted by $\rho_{\theta s,t}$. It is defined as

$$\rho_{\theta s} \equiv \frac{w_{\theta s} w}{w_s w_\theta}. \tag{20.}$$

A positive $\rho_{\theta s}$ means that higher-ability agents reap higher marginal benefits from human capital, but it also means that human capital increases the exposure of the agent to stochastic ability and risk. A value of $\rho_{\theta s} > 1$ means that higher-ability agents reap proportionally higher returns from human capital, i.e., the wage elasticity with respect to human capital is increasing in ability. A separable wage function of the form $w_t = \theta_t + h_t(s_t)$ for some function h_t implies that $\rho_{\theta s,t} = 0$. A multiplicative form $w_t = \theta_t h_t(s)$, the one typically used in the taxation literature, implies that $\rho_{\theta s,t} = 1$. A constant elasticity of substitution (CES) wage function $w_t = [\alpha_{1t}\theta^{1-\rho_t} + \alpha_{2t}s_t^{1-\rho_t}]^{\frac{1}{1-\rho_t}}$ has $\rho_{\theta s,t} = \rho_t$.

The planning problem is as above, except that an allocation now also specifies a desired level of human capital per period, in addition to consumption and output. Ability θ_t and labor supply are unobservable in any period. The envelope condition above was written more generally in terms of the wage function, and hence it is still valid, but the wage is now explicitly a function of human capital as well. In addition, the recursive problem now has an extra state variable, which is human capital in the last period.

The recursive formulation of the relaxed objective is then, for $t \geq 2$,

$$K(v, \Delta, \theta_-, s_-, t) = \min \int (c(\theta) + M_t(s(\theta) - s_-) - w_t(\theta, s(\theta)) l(\theta) + \frac{1}{R} K(v(\theta), \Delta(\theta), \theta, s(\theta), t+1)) f^t(\theta|\theta_-) d\theta,$$

where the maximization is over the functions $(c(\theta), l(\theta), s(\theta), \omega(\theta), v(\theta), \Delta(\theta))$, subject to the same constraints as in Equation 11. As above, the formulation is modified for period $t = 1$ to capture redistributive concerns.

3.3. Optimal Human Capital Policies

Similarly to the labor and the savings wedges above, we can define the human capital wedge,

$$\tau_S(\theta^t) \equiv -(1 - \tau_L(\theta^t)) w_{s,t} l_t + M_t'(e_t) - \beta E_t\left(\left(\frac{u_{t+1}'(c_{t+1})}{u_t'(c_t)}\right)(M_{t+1}'(e_{t+1}) - \tau_{St+1})\right), \tag{21.}$$

as the gap between marginal costs and marginal benefits from human capital investments. Implicitly, the agent's net marginal cost of investing in human capital is locally reduced to $M_t'(e_t) - \tau_{St}$.[5]

There are many simultaneous distortions here, and thus a zero human capital wedge does not mean that human capital is not distorted: Part of this subsidy is simply undoing some of the effects of the labor and capital distortions on human capital investments. A useful object is the net human capital subsidy, which ensures that the tax system is neutral with respect to human capital—i.e.,

[5] Human capital yields flows of returns in all future periods. It is written recursively here, replacing the latter stream by the next period's marginal cost.

that, conditional on the labor choice, the human capital decision is chosen as in the first best with perfect consumption smoothing. Formally, we have $l_t w_{st}(s_t, \theta_t) = M'_t(s_t - s_{t-1}) - \frac{1}{R}E_t(M'_{t+1}(s_{t+1} - s_t))$.

To grasp this concept, consider a one-period version of the model, with $s = e$ and linear taxes and subsidies. An agent of type θ solves $\max_{s,l} u(w(s,\theta)l(1-\tau_L) - M(s) + \tau_S s) - \phi(l)$. If we set the subsidy to be $\tau_S = \tau_L M'(s)$, the agent chooses human capital as in the first best, $(1-\tau_L)(w_s l - M'(s)) = 0$. This is equivalent to making human capital expenses fully tax deductible, i.e., taxable income is only $wl - M(s)$. The net subsidy on human capital is (appropriately scaled)

$$t_{st} \equiv \frac{\tau_S - \tau_L M'(s)}{(\tau_S - M'(s))(1-\tau_L)}.$$

It is zero when there is full deductibility, and it is positive when human capital is encouraged more than at full deductibility.

In this more complex multi-period model, I introduce a similar concept of full dynamic, risk-adjusted deductibility, taking into account that (*a*) marginal utility varies across states due to imperfect insurance, (*b*) there is a stream of benefits from human capital, (*c*) savings are distorted, and (*d*) there are also human capital subsidies in future periods. Hence, I define the net wedge as the gross wedge from which I filter out all the parts just listed that only go toward compensating for the other distortions.

First, for any variable x, define the "insurance factor" of x,

$$\xi_{x,t+1} \equiv -Cov\left(\beta \frac{u'_{t+1}}{u'_t}, x_{t+1}\right) / \left(E_t\left(\beta\frac{u'_{t+1}}{u'_t}\right) E_t(x_{t+1})\right),$$

with $\xi_{x,t+1} \in [-1, 1]$. If x is a flow to the agent, it is a good hedge if $\xi < 0$ and a bad hedge otherwise. With some abuse of notation, define also

$$\xi'_{x,t+1} \equiv -Cov\left(\frac{\beta u'_{t+1}}{u'_t} - \frac{1}{R}, x_{t+1}\right) / \left(E_t\left(\frac{\beta u'_{t+1}}{u'_t} - \frac{1}{R}\right) E_t(x_{t+1})\right),$$

which, up to an additive constant, captures the same risk properties as $\xi_{x,t+1}$.

Definition 1. Define the net wedge on human capital expenses, t_{st}, as

$$t_{st} \equiv \frac{\tau^d_{St} - \tau_{Lt} M'^d_t + P_t}{\left(M'^d_t - \tau^d_{St}\right)(1-\tau_{Lt})}, \qquad 22.$$

where $\tau^d_{St} \equiv \tau_{St} - \frac{(1-\xi_{\tau_S})}{R(1-\tau_K)}E_t(\tau_{St+1})$ is the dynamic risk-adjusted subsidy, and $M'^d_t \equiv M'_t - \frac{(1-\xi_{M'})}{R(1-\tau_K)}E_t\left(M'_{t+1}\right)$ denotes the dynamic, risk-adjusted cost. $P_t \equiv \frac{\tau_K}{R(1-\tau_K)}(1-\tau_{Lt})\left(1-\xi'_{M'}\right)E_t\left(M'_{t+1}\right)$ captures the risk-adjusted savings distortion.

If $\tau_{St} = \tau_{Lt}M'^d_t - P_t + \frac{(1-\xi_{\tau_S})}{R(1-\tau_K)}E_t(\tau_{St+1})$, such that for every marginal investment e_t a locally linear subsidy $\tau_{St}e_t$ is received, there is full dynamic risk-adjusted deductibility.

Proposition 3. At the optimum and at each history, the labor and human capital wedges need to satisfy the following relation:

$$t^*_{st} = \left(\frac{\tau^*_{Lt}}{1-\tau^*_{Lt}}\right)\frac{\varepsilon^c_t}{1+\varepsilon^u_t}(1-\rho_{\theta s,t}). \qquad 23.$$

Despite the complexity of the model, Equation 23 gives us a clear link between the labor wedge and the net human capital wedge. This relation can be used to simply check for the optimality of a given existing tax and subsidy system. The two wedges need to comove if and only if $\rho_{\theta s} < 1$.[6]

If the wage is a CES function as above, with ρ_t constant, and disutility is separable and isoelastic $\phi(l) = \frac{1}{\gamma} l^{\gamma}$ $(\gamma > 1)$, the ratio of the net human capital and labor wedges is constant cross-sectionally and over time:

$$t^*_{st} / \left(\frac{\tau^*_{Lt}}{1 - \tau^*_{Lt}} \right) = \frac{(1 - \rho)}{\gamma}.$$

The sign of the net human capital wedge is determined by the Hicksian coefficient of complementarity, $\rho_{\theta s}$: The net human capital wedge is positive if and only if $\rho_{\theta s} < 1$.

Proposition 4. When there is a positive labor wedge, $\tau^*_{Lt}(\theta^t) \geq 0$, we have

$$t^*_{st}(\theta^t) \geq 0 \Leftrightarrow \rho_{\theta s,t} \leq 1.$$

The optimal net wedge results from the balance of two effects. First, it increases human capital and the returns to work, thereby encouraging labor supply, which is a beneficial labor supply effect, given that there is a positive labor wedge. Second, it affects the pretax income distribution. If $\rho_{\theta s} > 0$—i.e., if ability is complementary to human capital in the wage—human capital mostly benefits already able agents, and hence it compounds existing pretax inequality due to intrinsic differences in θ_t. The opposite occurs if $\rho_{\theta s} < 0$, in which case human capital reduces inequality. This effect will be labeled the inequality effect. What happens on net is determined by the gap between $\rho_{\theta s}$ and 1. If $\rho_{\theta s} < 1$, the positive labor supply effect outweighs the inequality effect, i.e., on net, posttax inequality is reduced. Human capital then has a positive insurance and redistributive effect on after-tax income inequality.

Intuitively, the inequality effect comes from agents' incentive compatibility constraints. If high-productivity agents benefit more from a marginal increase in human capital ($\rho_{\theta s} > 0$), an increase in their human capital tightens their incentive constraints. What is relevant for social welfare is whether the overall increase in resources from more labor is completely canceled out by the information rent forfeited to high-productivity agents or not. When $\rho_{\theta s} < 1$, human capital investments generate positive net resources to be used for redistribution and insurance of all agents.

With a multiplicatively separable wage $w = \theta h(s)$ for some function h, for which $\rho_{\theta s} = 1$, a null net wedge is optimal. This is an application of Atkinson & Stiglitz's (1976) result on the nonoptimality of differential commodity taxation under preference separability.

If there are several types of human capital, $s_1, \ldots, s_J$, with different Hicksian coefficients of complementarity $\rho_{\theta s_j}$, with $j = 1, \ldots, J$, Equation 23 applies to each, so that at the optimum we obtain

$$\frac{t^*_{s_j t}}{1 - \rho_{\theta s_j,t}} = \frac{t^*_{s_i t}}{1 - \rho_{\theta s_i,t}} \ \forall (i, j). \qquad 24.$$

It is thus optimal to subsidize at higher rates the human capital types that have the highest redistributive and insurance effects.

[6]Note that the zero distortion at the bottom and top result, familiar for the labor wedge, holds here for the net human capital wedge. It does not hold for the gross wedge τ_{St}, underscoring again that the true incentive effects are captured by t_{st}, not τ_{St}.

It is also possible to rewrite the net human capital wedge recursively to show that it inherits the persistence of the shock process and also features a drift term, the sign of which is driven by $(1 - \rho_{\theta s})$. When human capital has a positive redistributive or insurance value, it is optimally increasing with age as it provides valuable insurance against the compounding skill shocks (see Stantcheva 2017).

3.3.1. Implementation. The optimal policies can be implemented with a system of income-contingent loans, whereby agents receive loans throughout life in order to invest in their human capital and repay the loan in an income-contingent way, paying more in periods and after histories in which their income realizations are better.

3.3.2. Unobservable human capital investments. If human capital investments are unobservable, as shown by Stantcheva (2014), the labor wedge is used to indirectly incentivize the right amount of human capital investments, but the labor wedge need not be smaller with unobservable human capital than it is when human capital is observable. When the desired net wedge on human capital (were it observable) is negative, the labor wedge could be higher with unobservable human capital. In addition, as hidden human capital investments are an alternative to physical capital (i.e., savings) for transferring resources to the future, their presence invalidates the standard inverse Euler equation. While the planner does control the total financial resources allocated per period, they do not control how these resources are allocated by the agent between consumption and human capital expenses. The agent's standard Euler equation in human capital holds, which imposes a restriction on the marginal utilities from consumption in different periods and modifies the inverse Euler equation for physical capital.

3.4. Training

In addition to spending resources, people spend a lot of time acquiring human capital throughout their lives, whether through formal college, continuing education, online degrees, on-the-job training, or vocational training. The peculiarity of time investments in human capital is that they are immutably linked to a given agent: Not only do their returns depend on an agent's ability, but also their costs depend on the agent's labor supply. In this section, I present a model based on previous work (Stantcheva 2015a) in which agents can invest in training and work every period, and the disutility from training depends on labor supply. I introduce two concepts: learning-or-doing and learning-and-doing. The former indicates that labor and training time are substitutes, because time spent working cannot be spent training. The limit case of this is the standard opportunity cost of time model by Ben-Porath (1967), in which agents have a given set of hours to allocate between training and working. Learning-and-doing is the case in which labor and training are complements, the limit case of which is the canonical learning-by-doing model of Arrow (1962), in which training is a direct by-product of labor.

Let z_t denote the stock of training time, or the stock of human capital of an agent, at time t, and let i_t denote the incremental training time acquired in period t. Human capital, z_t, evolves according to

$$z_t = z_{t-1} + i_t.$$

The disutility cost to an agent who provides $l_t \geq 0$ units of work and spends $i_t \geq 0$ units in training is $\phi_t(l_t, i_t)$, strictly increasing and convex in each of its arguments. The wage w_t is determined by the training acquired as of time t and by a stochastic ability θ_t: $w_t = w_t(\theta_t, z_t)$. Define ρ^{ϕ}_{lz} to be the Hicksian complementarity coefficient between labor and training in the disutility function ϕ: $\rho^{\phi}_{lz} \equiv \frac{\phi_{lz}\phi}{\phi_l \phi_z}$.

Let me illustrate the optimal policies in a simple one-time investment framework before generalizing the results. For any type θ, define the training wedge $\tau_Z(\theta)$, or implicit subsidy on training, as follows:

$$\tau_Z(\theta) \equiv \frac{\phi_z}{u'(c)} - (1 - \tau_L(\theta))w_z l. \quad 25.$$

The implicit subsidy on training τ_Z can be thought of as the incremental pay received by an agent for training for one more unit of time. Following the same logic as above, it is useful to find a measure of the net distortion on human capital, the one that goes beyond just compensating for the presence of a labor distortion. The net wedge on training time t_z is defined as

$$t_z \equiv \frac{\tau_Z - \tau_L(\frac{\phi_z}{u'(c)})}{((\frac{\phi_z}{u'(c)}) - \tau_Z)(1 - \tau_L)}. \quad 26.$$

Proposition 5. At the optimum, the net subsidy for training and the labor wedge are set according to

$$t^*_z(\theta) = \frac{\tau^*_L(\theta)}{1 - \tau^*_L(\theta)} \frac{\varepsilon^c}{1 + \varepsilon^u} \left(1 - \rho_{\theta z} - \frac{\varepsilon_{\phi z}}{\varepsilon_{wz}} \rho^{\phi}_{lz}\right), \quad 27.$$

where $\varepsilon_{\phi z} \equiv d \log(\phi)/d \log(z)$ is the elasticity of disutility, and $\varepsilon_{wz} \equiv d \log(w)/d \log(z)$ is the elasticity of the wage with respect to training.

In this simple one-period model, there are three effects from subsidizing training that are balanced at the optimum. The labor supply effect and the inequality effect are the same as described for resource investments in human capital in the previous section. The third and new effect is the direct interaction with labor supply through the disutility function, i.e., either learning-or-doing or learning-and-doing. Rearranging this, we can see that the total effect of a training subsidy on labor is positive if and only if

$$\varepsilon_{wz} > \varepsilon_{\phi_l z} \text{ with } \varepsilon_{\phi_l z} \equiv \partial \log(\phi_l) / \partial \log(z),$$

i.e., if and only if the wage is more sensitive to training than the marginal disutility of work is.[7]

The question, then, is whether the increase in total resources from the total labor effect of training more than compensates for the increased rent transfers (the inequality effect). The net subsidy on training will be positive if and only if the answer to this question is yes, i.e., if and only if we have

$$1 - \frac{\varepsilon_{\phi z}}{\varepsilon_{wz}} \rho^{\phi}_{lz} > \rho_{\theta z}. \quad 28.$$

With learning-and-doing ($\rho^{\phi}_{lz} < 0$), as long as ability and human capital are not too complementary (say, $\rho_{\theta z} < 1$), the net subsidy on training is positive. Intuitively, training does not distract from labor, and so it is good to encourage it as long as high-ability types do not disproportionately benefit from it. However, if there is learning-or-doing, training makes labor supply more costly. In this case, even if the coefficient of complementarity $\rho_{\theta z}$ is small, it might not be sufficiently small to compensate for the lost work effort.

[7]Note that $(\varepsilon_{\phi z,t}/\varepsilon_{wz,t})\rho^{\phi}_{lz,t} = \varepsilon_{\phi_l z,t}/\varepsilon_{wz,t}$.

In the special case in which the wage is multiplicatively separable and ϕ is additively separable, with $\phi(l, z) = \phi^1(l) + \phi^2(z)$, an application of the Atkinson-Stiglitz theorem (Atkinson & Stiglitz 1976) to the two commodities c and z yields $t_z = 0$ (through a simple variational argument).

I point out some additional special cases, which also hold in the dynamic model if human capital fully depreciates between periods.

1. If the wage is multiplicatively separable, $w(\theta, z) = \theta z$, and the disutility is Cobb-Douglas, $\phi(l, z) = \frac{1}{\gamma\alpha} l^\gamma z^\alpha$, we obtain a simple negative relation between the optimal labor wedge and the optimal training wedge at any point in the skill distribution, such that $t_z^*(\theta) = -\frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} \frac{\alpha}{\gamma}$ and $t_z(\theta) < 0$ at any interior type.
2. If the wage takes a CES form $w(\theta, z) = (\theta^{1-\rho} + z^{1-\rho})^{\frac{1}{1-\rho}}$ and the disutility is separable, with $\phi(l, z) = \phi^1(l) + \phi^2(z)$, then we have $t_z^*(\theta) = \frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} \frac{\varepsilon^c}{1+\varepsilon^u}(1 - \rho)$. If, in addition, disutility is isoelastic in labor, with $\phi^1(l) = \frac{1}{\gamma} l^\gamma$, then we obtain $t_z^*(\theta) = \frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} \frac{(1-\rho)}{\gamma}$. Hence, the net wedge on training is positive if and only if $\rho_{\theta z} < 1$, i.e., if ability and training are not too complementary in generating earnings.
3. If the wage is separable, with $w(\theta, z) = \theta + z$, and the disutility is again Cobb-Douglas as above, we have $t_z^*(\theta) = \frac{\tau_L^*(\theta)}{1-\tau_L^*(\theta)} \frac{(1-\alpha)}{\gamma}$. In this case, the optimal wedge is again negative, since $\alpha > 1$.

Note that, in general, with asymmetric information, money and time investments are not equivalent: One cannot be perfectly converted into the other because training time interacts with unobservable labor supply. It is only when the disutility is separable in labor and training that we exactly recover the same formulas and results as for the monetary investments above.

In the full-fledged dynamic model, the subsidy on training time has an additional direct interaction with future labor supply through the disutility function (in addition to all aforementioned static effects). Even if training diverts time away from contemporaneous labor supply, the effects on future labor supply can motivate a positive net subsidy. If contemporaneous labor supply and training are complements, then current training and future labor supply are substitutes and vice versa, because investing in human capital today means having to invest less tomorrow to reach any given level of it.

With learning-or-doing, the net wedge comoves positively with the future income tax rate τ_{Lt+1} but negatively with the current tax rate τ_{Lt}. When there is a higher current wedge on labor, training that is a substitute for labor will be fostered indirectly, with less need to subsidize it directly. The opposite holds for the future labor wedge.

3.5. Intergenerational Concerns

We can also consider human capital policies in an intergenerational model, where each t is a generation, as is done by Stantcheva (2015b). Parent i in generation t can buy an education amount s_{t+1} for their child of generation $t + 1$. This setup reflects the fact that most investments in human capital occur before and during college and are in large part paid by parents.

The wage w_{ti} of agent i in generation t is determined by their stock of human capital and their stochastic ability θ_{ti}, that is,

$$w_{ti}(s) \equiv w(s, \theta_{ti}).$$

Ability θ_{ti} is drawn from a stationary ergodic distribution that allows for correlation between generations. Unless there is perfect persistence, parents face some uncertainty regarding their children's ability realizations at the time they are making education investment decisions. In addition to financing their education, parents can also leave financial bequests to their children. Bequests left by generation t are denoted by b_{t+1i} and earn a generational gross rate of interest R. Thus, generation t inherits a pretax bequest of Rb_{ti} from their parents. The initial generation 1 has an exogenously given distribution of bequests b_{1i}.

First, by the inverse Euler logic described above, there is optimally a positive wedge on bequests, i.e., $\tau_B > 0$. In addition, the relation between tax treatment of bequests and human capital at the optimum is as follows.

Proposition 6. At the optimum, the following relation needs to be satisfied:

$$R = E\left(w_{s,t+1} l_{t+1}\left(1 + \tau_{Lt+1}\frac{\varepsilon^c_{t+1}}{1+\varepsilon^u_{t+1}}\left(1 - \rho_{\theta s,t+1}\right)\right)\right).$$

The left-hand side is simply the (social) return on bequests and it is equated to the right-hand side, which is the social return to education. The first part of the social return to education is simply the wage increase of the next generation from education. The second part captures the incentive implications of education for the next generation. Education has two effects on the incentives of children. First, it encourages their work effort, which relaxes their incentive constraints. This is the so-called labor supply effect. Second, depending on the sign of the complementarity between human capital and ability, education may increase or decrease pretax inequality. If $\rho_{\theta s} > 0$, education increases pretax inequality and benefits mostly able kids. This tends to reduce the effective incentive-adjusted benefit of education, and it is called the inequality effect. The net effect on children's incentives depends on the sign of $(1 - \rho_{\theta s})$, i.e., the redistributive and insurance effect of human capital. This is scaled by a multiple of the labor wedge, which captures the efficiency cost of taxation, i.e., the value of relaxing children's incentive constraints.

At the optimum, the return on bequests is not equated to the return on human capital investments; instead, it needs to be equated to the expected, incentive-adjusted return on education that takes into account the direct increase in earnings as well as the labor supply effect and the inequality effect on the incentive constraint. Whereas bequests benefit all types uniformly in marginal terms, human capital investments have redistributive incentive effects.[8]

If education is highly complementary to ability (i.e., $\rho_{\theta s} > 1$), which means that high-ability children benefit more in proportional terms from their parents' education investments, then the return to education investments will be reduced below the return on bequests. Put differently, education investments by parents will be taxed relative to bequests. The opposite happens when education is not too complementary to children's ability ($\rho_{\theta s} < 1$), in which case parental education investments should be subsidized relative to bequests. With the separable wage function that has $\rho_{\theta s} = 1$, parental education investments and bequest choices should not be distorted relative to each other, i.e., $R = E(w_{s,t+1} l_{t+1})$.

4. THE QUANTITATIVE RAMSEY APPROACH

A middle-of-the road approach between the fully unrestricted dynamic Mirrlees approach in the previous section and the sufficient statistics one in the next section is the parametric and quantitative Ramsey-style approach. This framework ex ante parametrically specifies the type of tax instruments to be used and quantitatively (more rarely, analytically) assesses optimal policies. The

[8] Bequests would have income effects that would interact with agents' types if utility were not separable in consumption and labor.

key advantage is that, thanks to the restrictions on instruments and the use of quantitative methods, more complex and realistic economies can be studied.

It is impossible to do justice to the very long-standing Ramsey tax literature and the very large number of papers studying either tax reform or optimal tax policy in very different settings. In the interest of space, I only present some very recent studies and findings using a quantitative Ramsey approach. The key foci of this recent quantitative literature have been on the optimality of age-dependent taxes, on when a positive capital tax is optimal, and on what shapes the progressivity and level of income taxes.

Heathcote et al. (2017) propose a parsimonious two-parameter tax function—already used by Benabou (2002)—that captures the level and progressivity of taxes: $T(y) = y - \lambda \cdot y^{(1-\tau)}$. In a life-cycle model with skill investment, heterogeneous tastes for work, a public good, and wealth in zero net supply, they study what factors shape the progressivity of the tax and transfer system. Quantitatively, the disincentives that taxes have on endogenous skill investment and labor supply, as well as the desire to finance government purchases, matter to similar extents. The progressivity in the actual US tax system can be obtained as optimal in a version of their model where credit constraints at low-income levels prevent efficient investments in skills. Karabarbounis (2016) shows that tagging the level of taxes by age, household assets, and filing status (married versus single) improves the efficiency of the tax system. Heathcote et al. (2019) also study age-dependent taxation in their parametric setting, allowing both the level and the progressivity of taxes to vary by age.

Using a quite general parametric tax function, Conesa & Krueger (2006) show that the optimal income tax system for the United States (which does not differentiate between labor and capital income) is well approximated by a flat tax of 17.2% combined with a fixed deduction of $9,400. Allowing for a distinction between capital and labor income, Conesa et al. (2009) develop an overlapping generations (OLG) life-cycle model with uninsurable idiosyncratic labor income risk, and they show that, in the absence of age-dependent labor income taxes, positive capital taxes are optimal as they imperfectly mimic the age dependency that would be needed. They thus generalize the same finding of Erosa & Gervais (2002) to a case with idiosyncratic risk. Findeisen & Sachs (2017) focus on a life-cycle model with heterogeneity and risk in skills and solve for the optimal nonlinear labor income tax that only depends on current income and the optimal linear capital tax. Insurance against idiosyncratic skill shocks drives the labor income tax, and the capital tax is optimally positive. Guvenen et al. (2019) show that with heterogeneous returns a wealth tax targets the unproductive entrepreneurs and increases the savings rate of productive ones relative to a capital income tax. Quantitatively, it can raise productivity while also reducing consumption inequality.

Some recent papers also adopt this parametric approach to study human capital policies. Benabou (2002) concludes that financing education produces more growth than using taxes and transfers to alleviate credit constraints that prevent efficient investments in education, but at the cost of providing less consumption insurance. Krueger & Ludwig (2013) incorporate education investments in the form of college into a large-scale OLG model with uninsurable idiosyncratic income risk, borrowing constraints, intergenerational transmission of wealth and ability, and incomplete financial markets. They show that the optimal tax and transfer system features substantial progressivity in labor income taxes complemented with a large subsidy for college education.

5. THE SUFFICIENT STATISTICS APPROACH

The goal of the sufficient statistics approach to dynamic taxation is to better connect the theory of optimal capital taxation to the policy debate by providing a tractable framework to address many policy questions. The goal is to derive robust optimal capital tax formulas expressed in terms of elasticities of capital and labor supply with respect to the net-of-tax rates that can be estimated in the data and to distributional considerations that society may have.

When it comes to studying capital taxation more specifically, the aim is also to build a model that generates an empirically realistic response of capital to taxes (unlike the infinite elasticities obtained in Chamley-Judd), is sufficiently tractable to yield results for a variety of policy topics related to capital taxation, and is general enough for these results to be robust to a broader set of models.

Some examples of the policy topics that have traditionally been hard to deal with in dynamic optimal capital tax models and that can be dealt with here include, among others, income shifting between capital and labor, economic growth, heterogeneous returns to capital across individuals, and different types of capital assets and heterogeneous tastes for each of them. This approach is also very amenable to incorporating a broader range of justice and fairness principles related to capital and labor taxation through the use of generalized social welfare weights, as done by Saez & Stantcheva (2016). In short, the optimal tax for different justice and fairness principles can be obtained by simply plugging into the formulas the corresponding generalized social welfare weights.

The analysis in this section is based on work by Saez & Stantcheva (2018). Golosov et al. (2014) provide a more general and formal analysis of dynamic taxation using perturbation and sufficient statistics methods.

5.1. Setup

Individual i has instantaneous utility with functional form $u_i(c, k, z) = c + a_i(k) - h_i(z)$, linear in consumption c, increasing in wealth k with $a_i(k)$ increasing and concave, and with a disutility cost $h_i(z)$ of earning income z that is increasing and convex in z. One strength of this framework is that the index i can represent any arbitrary heterogeneity in the preferences for work and wealth or in the discount rate δ_i. The discounted utility of i from an allocation $\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}$ is

$$V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = \delta_i \cdot \int_0^{\infty} [c_i(t) + a_i(k_i(t)) - h_i(z_i(t))]e^{-\delta_i t}\, dt. \qquad 29.$$

The net return on capital is r. The initial wealth of individual i is k_i^{init}. Consider a given time-invariant tax schedule $T(z, rk)$ based on labor and capital incomes. The budget constraint of individual i is

$$\frac{dk_i(t)}{dt} = rk_i(t) + z_i(t) - T(z_i(t), rk_i(t)) - c_i(t). \qquad 30.$$

$T'_L(z, rk) \equiv \partial T(z, rk)/\partial z$ denotes the marginal tax with respect to labor income, and $T'_K(z, rk) \equiv \partial T(z, rk)/\partial(rk)$ denotes the marginal tax with respect to capital income.

Because utility is linear in consumption, $(c_i(t), k_i(t), z_i(t))$ jumps immediately to its steady-state value (c_i, k_i, z_i) characterized by $h'_i(z_i) = 1 - T'_L, a'_i(k_i) = \delta_i - r(1 - T'_K), c_i = rk_i + z_i - T(z_i, rk_i)$. Lifetime utility can be rewritten as

$$V_i(\{c_i(t), k_i(t), z_i(t)\}_{t\geq 0}) = U_i(c_i, k_i, z_i) = [c_i + a_i(k_i) - h_i(z_i)] + \delta_i \cdot (k_i^{\text{init}} - k_i). \qquad 31.$$

The last term represents the utility cost of going from wealth k_i^{init} to wealth k_i at instant 0. The dynamic model of Equation 29 is thus mathematically equivalent to a static representation where the agent simply chooses (c_i, k_i, z_i) to maximize the static utility equivalent in Equation 31 subject to the static budget constraint $c_i = rk_i + z_i - T(z_i, rk_i)$. Anticipated and unanticipated reforms have the same effect.

Individuals in this model accumulate different levels of wealth based on their heterogeneous tastes for wealth and impatience levels as well as on the net-of-tax return $\bar{r} = r(1 - T_K'(z, rk))$. As a result, steady-state wealth levels are heterogeneous, even conditional on labor earnings. Because of this, the zero tax result of Atkinson & Stiglitz (1976) does not apply.

The wealth-in-the-utility feature puts a limit on individuals' impatience to consume; there is value in keeping some wealth. At the optimum, the value lost in delaying consumption $\delta_i - \bar{r}$ is equal to the marginal value of holding wealth $a_i'(k)$, and the optimum for capital holding is interior. Capital hence exhibits a smooth behavior in the steady state, with a finite elasticity of capital supply with respect to the net-of-tax return.

Having wealth in the utility makes sense for conceptual and empirical reasons. Conceptually, wealth can be held for reasons other than just smoothing consumption and can bring other benefits than just the future consumption flows. Empirically, it is difficult to rationalize the very large wealth holdings purely based on a consumption-smoothing motive. In addition, wealth holdings are heterogeneous even conditional on labor earnings, and the wealth distribution is much more skewed than could be explained by a model in which wealth is purely the result of different-ability people saving their labor income. This shows that there is additional heterogeneity in preferences related to wealth, over and beyond heterogeneous abilities to work. More precisely, wealth in the utility can be microfounded by a bequest motive, a utility cost or disutility benefit from entrepreneurship, services provided by wealth (such as liquidity), or social status concerns.

5.2. Optimal Tax Formulas

The social welfare function (SWF) is defined as

$$\text{SWF} = \int_i \omega_i \cdot U_i(c_i, k_i, z_i) di, \tag{32.}$$

where $\omega_i \geq 0$ is the Pareto weight on individual i. We denote by $g_i = \omega_i \cdot U_{ic} = \omega_i$ the social marginal welfare weight on individual i and normalize $\int_i \omega_i di = 1$. The government sets the time-invariant tax $T(z, rk)$, subject to budget balance, to maximize this social welfare objective. We start with linear taxes and then derive nonlinear taxes.

5.2.1. Optimal linear capital and labor taxation.

With linear taxes, the government rebates tax revenue lump-sum, and the transfer to each individual is $G = \tau_K \cdot rk^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$, where $z^m(1 - \tau_L) = \int_i z_i di$ is the aggregate labor income that depends on $1 - \tau_L$, and $k^m(\bar{r}) = \int_i k_i di$ is the aggregate capital stock, which depends on $\bar{r} = r(1 - \tau_K)$. τ_K and τ_L are chosen to maximize SWF in Equation 32, with $c_i = (1 - \tau_K) \cdot rk_i + (1 - \tau_L) \cdot z_i + \tau_K \cdot rk^m(\bar{r}) + \tau_L \cdot z^m(1 - \tau_L)$.

Let e_K be the elasticity of aggregate capital k^m with respect to $\bar{r}$, and let e_L be the elasticity of aggregate labor income z^m with respect to the net of tax rate $1 - \tau_L$. There are no income effects with the utility assumed. Hence, we have $e_L > 0$ and $e_K > 0$. Applying the individuals' envelope theorems for the choice k_i, we can obtain the optimal linear capital tax.

Proposition 7 (optimal linear capital tax). The optimal linear capital tax is given by

$$\tau_K = \frac{1 - \bar{g}_K}{1 - \bar{g}_K + e_K}, \quad \text{with} \quad \bar{g}_K = \frac{\int_i g_i \cdot k_i}{\int_i k_i} \quad \text{and} \quad e_K = \frac{\bar{r}}{k^m} \cdot \frac{dk^m}{d\bar{r}} > 0. \tag{33.}$$

The optimal labor tax can be derived exactly symmetrically:

$$\tau_L = \frac{1-\bar{g}_L}{1-\bar{g}_L+e_L}, \quad \text{with} \quad \bar{g}_L = \frac{\int_i g_i \cdot z_i}{\int_i z_i} \quad \text{and} \quad e_L = \frac{1-\tau_L}{z^m} \cdot \frac{dz^m}{d(1-\tau_L)} > 0. \qquad 34.$$

We can now see that the optimal capital tax will be zero only if $\bar{g}_K = 1$ or $e_K = \infty$. The former case occurs when there are no redistributive concerns regarding capital income (i.e., g_i is uncorrelated with k_i). However, as long as capital is concentrated among individuals with lower social marginal welfare weights (i.e., g_i is decreasing in k_i), we have $\bar{g}_K < 1$, and the optimal capital tax is strictly positive. The revenue-maximizing tax rates are obtained by setting $\bar{g}_K = 0$ and $\bar{g}_L = 0$. If there is no wealth in the utility, capital responses are no longer smooth; the elasticity e_K is infinite, which drives the Chamley-Judd zero optimal capital tax result.

Supplemental Material >

Supplemental Table 2 shows a summary of the empirical literature on the tax elasticities of capital and distinguishes between capital gains, dividends, ordinary capital income, and bequests and inheritances. Although there is substantial variation in the estimates based on the settings and tax considered, and although the elasticities are typically larger than those known for labor income, they are not as large as the standard tax model of Chamley-Judd would predict. This justifies the need for a theory that generates steady states with nondegenerate wealth distributions that feature very large wealth holdings and finite smooth responses of capital to taxation.

5.2.2. Optimal nonlinear separable taxes. Consider now nonlinear, separable, and time-invariant tax schedules $T_L(z)$ and $T_K(rk)$. Let $\bar{G}_K(rk)$ [symmetrically, $\bar{G}_L(z)$] be the average relative welfare weight on individuals with capital income higher than rk (symmetrically, labor income higher than z). We obtain

$$\bar{G}_K(rk) = \frac{\int_{\{i:rk_i \geq rk\}} g_i di}{P(rk_i \geq rk)} \quad \text{and} \quad \bar{G}_L(z) = \frac{\int_{\{i:z_i \geq z\}} g_i di}{P(z_i \geq z)}. \qquad 35.$$

The cumulative distributions of capital and labor income are $H_K(rk)$ and $H_L(z)$, and $h_K(rk)$ and $h_L(z)$ are the corresponding densities when the tax system is linearized at points rk and z. The local Pareto parameters of the capital and labor income distributions (which depend on the tax system) are

$$\alpha_K(rk) \equiv \frac{rk \cdot h_K(rk)}{1-H_K(rk)} \quad \text{and} \quad \alpha_L(z) \equiv \frac{z \cdot h_Z(z)}{1-H_Z(z)}.$$

The local elasticity of k with respect to the net of tax return $r(1-T'_K(rk))$ at income level rk is denoted by $e_K(rk)$, and that of z with respect to $1-T'_L(z)$ is denoted by $e_L(z)$.

Proposition 8 (optimal nonlinear capital and labor income taxes). The optimal nonlinear capital and labor income taxes are

$$T'_K(rk) = \frac{1-\bar{G}_K(rk)}{1-\bar{G}_K(rk)+\alpha_K(rk)\cdot e_K(rk)} \quad \text{and} \quad T'_L(z) = \frac{1-\bar{G}_L(z)}{1-\bar{G}_L(z)+\alpha_L(z)\cdot e_L(z)}. \qquad 36.$$

Most ordinary capital income in many countries is taxed jointly with labor income by the individual income tax (e.g., interest earned from a standard savings account). Within this framework, the optimal nonlinear tax on comprehensive income $y \equiv rk + z$ of the form $T_Y(y)$ takes the same form as found by Mirrlees (1971) and Saez (2001).

Let $\bar{G}_Y(y) = \frac{\int_{\{i:y_i \geq y\}} g_i di}{P(y_i \geq y)}$ be the average welfare weight on individuals with total income higher than y, $H_Y(y)$ be the cumulative distribution of the total income distribution, and $h_Y(y)$ be the corresponding density, assuming again a linearized tax system at point y. The local Pareto parameter for the distribution of total income y is $\alpha_Y(y) \equiv \frac{y h_Y(y)}{1-H_Y(y)}$, and the elasticity of total income to the net of tax rate $1 - T_Y'(y)$ at point y is $e_Y(y)$.

Proposition 9 (optimal tax on comprehensive income).

1. The optimal nonlinear tax on comprehensive income $y = rk + z$ is given by

$$T_Y'(y) = \frac{1 - \bar{G}_Y(y)}{1 - \bar{G}_{Y(y)} + \alpha_{Y(y)} \cdot e_{Y(y)}}.$$

2. The optimal linear tax on comprehensive income is

$$\tau_Y = \frac{1 - \bar{g}_Y}{1 - \bar{g}_Y + e_Y}, \qquad 37.$$

$$\text{with } \bar{g}_Y \equiv \frac{\int_i g_i y_i}{y^m} = \frac{z^m \bar{g}_L + rk^m \bar{g}_K}{z^m + rk^m} \text{ and } e_Y \equiv \frac{dy^m}{d(1-\tau_Y)} \frac{(1-\tau_Y)}{y^m} = \frac{z^m e_L + rk^m e_K}{z^m + rk^m}. \qquad 38.$$

A tax system based on comprehensive income may be optimal for equity reasons if society considers it unfair to discriminate income based on its source, or for efficiency reasons, if there are stark income-shifting responses between the capital and labor income bases.

5.2.3. Extensions.

In this framework, it is easy to incorporate, among others, the following extensions: (*a*) jointness in preferences between work and wealth, which introduces an additional cross-elasticity term in the formula for capital $\tau_K = (1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e_{L,(1-\tau_K)})/(1 - \bar{g}_K + e_K)$ (and symmetrically for the labor tax); (*b*) heterogenous returns to capital, which can be captured by just plugging into the tax formula $\bar{g}_{rK} = \frac{\int_i g_i \cdot r_i k_i}{\int_i r_i k_i}$ and $e_{rK} = d\log(r_i k_i)/d\log(1-\tau_K)$; and (*c*) different capital assets, such as financial assets or real estate, which have different returns and for which agents have different tastes. If there are no cross-elasticities, for asset j the tax formula is simply $\tau_K^j = \frac{1-\bar{g}_K^j}{1-\bar{g}_K^j + e_K^j}$, where $\bar{g}_K^j = \frac{\int_i g_i \cdot k_i^j}{\int_i k_i^j}$ and $e_K^j = \frac{\bar{r}^j}{k^{m,j}} \cdot \frac{dk^{m,j}}{d\bar{r}^j} > 0$ are the average social welfare weight and the tax elasticity of that particular asset's income, respectively. The formula can be extended easily to the case with nonseparabilities in preferences for different assets (see Saez & Stantcheva 2018).

5.3. Generalization and a New Steady-State Approach

In the generalized model with concave utility for consumption and wealth in the utility, individual i chooses $(c_i(t), k_i(t), z_i(t))_{t \geq 0}$ to maximize

$$V_i = \delta \int_{t=0}^{\infty} u_i(c_i(t), k_i(t), z_i(t)) e^{-\delta t} dt \quad \text{such that} \quad \frac{dk_i(t)}{dt} = rk_i(t) + z_i(t) - T_t(z_i(t), rk_i(t)) - c_i(t).$$

The steady state (c_i, k_i, z_i) is characterized by

$$u_{ik}/u_{ic} = \delta - r(1 - T_K'), \qquad u_{ic} \cdot (1 - T_L') = -u_{iz}, \quad \text{and} \quad c_i = rk_i + z_i - T(z_i, rk_i), \qquad 39.$$

where u_{ic}, u_{ik}, and u_{iz} denote the partial derivatives of $u_i(c, k, z)$, and T_K' and T_L' denote the marginal tax rates on capital income and labor income, all evaluated at the steady state (c_i, k_i, z_i).

First, assume that the government always chooses a period-by-period neutral budget constraint so that for all t we have

$$\int_i T_t(z_i(t), rk_i(t))di = 0.$$

Second, assume time-invariant tax rates τ_K and τ_L with a budget-balancing lump-sum rebate $G(t)$. Hence, we obtain $T_t(z, rk) = \tau_L z + \tau_K rk - G(t)$. From the per period budget balancing assumption, we derive $G(t) = \tau_L z^m(t) + \tau_K rk^m(t)$, where $k^m(t)$ and $z^m(t)$ are average wealth and earnings at time t.

Third, assume that at time 0, the economy is already in steady state with its initial tax system, which means that G, $c_i(t)$, $z_i(t)$, $k_i(t)$, $k^m(t)$, and $z^m(t)$ are all equal to their (time-invariant) steady-state values. In the steady state, average capital $k^m = \int_i k_i$ and average earnings $z^m = \int_i z_i$ will be functions of $1 - \tau_K$ and $1 - \tau_L$ (since the lump-sum rebate G is also a function of τ_K, τ_L through budget balance). Steady-state capital k^m has a finite elasticity with respect to the net-of-tax return $\bar{r} = r(1 - \tau_K)$. Steady-state elasticity, denoted by e_K, is given by

$$e_K = \frac{\bar{r}}{k^m}\frac{dk^m}{d\bar{r}}. \qquad 40.$$

Note that e_K mixes substitution and income effects and changes in G. The presence of utility for wealth remains crucial for having a finite elasticity e_K. The cross-elasticity of k^m with respect to $1 - \tau_L$, denoted by $e_{K,1-\tau_L} = ((1 - \tau_L)/k^m) \cdot dk^m/d(1 - \tau_L)$, is also finite, as are the elasticities of aggregate labor earnings z^m with respect to both $1 - \tau_L$ and $\bar{r}$. The cross-elasticity $e_{L,1-\tau_K} = (z_m/\bar{r})(dz^m/d\bar{r})$ measures how labor earnings respond to changes in $\bar{r}$. Note that such cross-elasticities arise not only if there is jointness of (k, z) in utility but also through income effects, as the marginal utility of consumption u_{ic} affects labor supply decisions.

There are typically two approaches that can be taken to determine the optimal tax rate: The first considers unanticipated reforms, the second anticipated reforms. Both have issues. The unanticipated reform approach makes it very tempting to exploit sluggish responses and aggressively tax the existing capital stock, thus creating commitment issues. The anticipated reform approach puts very low social welfare weight on impatient agents (which gets discounted heavily in the social objective), assumes infinite foresight and anticipation of policy by agents, and generates extremely large (in the limit, infinite) responses of capital.

Saez & Stantcheva (2018) propose an alternative, new, and nonstandard solution to the optimal capital tax problem: the utility-based steady-state approach. Their goal is to neutralize the ability of the government to exploit sluggish responses. However, this is not done by using the anticipated reform approach, which has undesirable features. Instead, it is achieved by (*a*) letting the government explicitly recognize the long-run steady-state behavioral responses as the normatively relevant ones and (*b*) imposing that the government should also respect individual savings choices. The solution arising based on this approach is to use the standard optimal tax formulas with the steady-state elasticities.

Consider a small reform $d\tau_K$ at time 0. The actual response to this tax change is sluggish, so that the real change in taxes collected at time t is given by $dG(t) = rk^m d\tau_K + \tau_K r dk^m(t) + \tau_L dz^m(t)$. To formalize that the government does not want to exploit sluggish responses, assume that the government in fact considers that the budgetary effect at time t is $dG = rk^m d\tau_K + \tau_K r dk^m + \tau_L dz^m$ and absorbs the difference between $dG(t)$ and dG. For instance, for a tax increase $d\tau_K > 0$, responses are smaller at first, so that $dG(t) > dG$, but the government dissipates this surplus. From a normative perspective, then, the government ignores the gains it can make by exploiting slow responses. Formally the goal is to find the tax system (τ_K, τ_L) that maximizes SWF but assuming that the lump-sum grant $G(t)$ is equal to the steady-state lump-sum grant, $G = r\tau_K k^m + \tau_L z^m$, instead of the actual lump-sum grant $r\tau_K k^m(t) + \tau_L z^m(t)$.

Proposition 10 (optimal linear capital tax in the utility-based steady-state approach). The optimal linear capital tax is

$$\tau_K = \frac{1 - \bar{g}_K - \tau_L \frac{z^m}{rk^m} e_{L,1-\tau_K}}{1 - \bar{g}_K + e_K} \quad \text{with} \quad K = \int_i g_i \cdot k_i / k^m. \qquad 41.$$

A symmetric equation holds for the optimal labor income tax rate τ_L. Hence, the same tax formulas hold by simply using the steady-state elasticities. One advantage of this approach is its robustness to introducing heterogeneity in discount rates across individuals, as heterogeneity in discount rates is normatively irrelevant in the steady state. Importantly, all of the applications from the linear utility model as derived by Saez & Stantcheva (2018) carry over to capital taxation with small modifications related to the fact that a concave utility introduces cross-elasticities between capital and labor (but there are no transitional dynamics). Another advantage of the utility-based steady-state approach is that all the work done in the literature to incorporate more realistic features of labor taxation in static settings can carry over here to the taxation of capital in a dynamic, general setting.

Note that this approach is related to, but not identical to, choosing the budget-balanced tax system that maximizes steady-state welfare SWF $= \int_i \omega_i \cdot u_i(c_i, k_i, z_i) di$, because the steady-state maximization objective is paternalistic. Intuitively, increasing wealth looks good in the steady state because it "forgets" that accumulating wealth required sacrificing consumption in the past, which artificially creates a positive welfare effect of wealth accumulation that tends to lower the optimal capital income tax. The simplest way to resolve this issue of paternalism is to intentionally ignore the effect of dk_i on individual welfare by stating that any behavioral response triggered by a tax reform should have zero first-order effect on individual welfare through the envelope theorem. This amounts to saying that the government respects individual savings decisions. With this "forced" envelope theorem assumption, the optimal tax can be derived entirely in the steady state without dynamic considerations. The optimal steady-state tax formula is given exactly as in Proposition 10. This is thus an alternative way to obtain the same results.

6. CONCLUSION

I conclude here with what appear to be productive avenues for future research. In the dynamic Mirrlees approach, a general theory of approximation of the optimal, often complicated policies would represent a big advance. It is, for instance, not known how well the linear age-dependent approximations do outside of the parameterizations currently used in the literature. In addition, more realistic features of the economy—such as general equilibrium effects—should be added, which will require more quantitative analysis. This will go hand in hand with fuller estimations of the underlying parameters that go beyond calibrations. The parametric Ramsey approach, on the other hand, would gain from the exploration of less standard and more complex tax instruments. In that sense, a middle-ground combination of the strengths of the dynamic Mirrlees approach (i.e., less restricted instruments) and the Ramsey approach (i.e., more realistic quantitative features) would probably be very fruitful.

In the sufficient statistics approach, the key challenge is to obtain credible empirical estimates of the relevant longer-run elasticities. As better individual-level tax data have become available over time, credible estimates of the short- or medium-run responses based on policy variation can be obtained. The best way to go may then be to embed these reduced-form estimates into a structural model to estimate the long-run responses.

SUMMARY POINTS

1. Three main approaches are used to study dynamic taxation: (*a*) the dynamic Mirrlees approach, (*b*) the parametric dynamic Ramsey approach, and (*c*) the dynamic sufficient statistics approach. These give different answers to the question of how capital should be taxed.
2. The dynamic Mirrlees approach assumes that agents' abilities to earn income are heterogeneous, stochastic, and private information. Tax instruments ex ante are unrestricted. The model solves for the optimal allocations using dynamic mechanism design (subject only to incentive compatibility constraints) and then considers how to implement these allocations using decentralized tax systems.
3. Taxes are set for redistribution and insurance reasons. Capital is taxed only in order to improve incentives to work. Human capital is optimally subsidized if it reduces posttax inequality and risk on balance.
4. The dynamic Ramsey approach specifies restricted parametric forms for tax instruments and adopts quantitative methods, which allow it to consider more realistic and complex economies.
5. In the Ramsey approach, capital taxes are typically optimal when age-dependent labor income taxes are not feasible.
6. The newer and tractable sufficient statistics approach derives robust tax formulas that depend on estimable elasticities and features of the income distributions. It simplifies the transitional dynamics thanks to a newly defined criterion, the utility-based steady-state approach, that prevents the government from exploiting sluggish responses in the short run.
7. Capital is taxed here for the same reasons labor income would be taxed, based on the equity–efficiency trade-off. Capital taxes are higher when capital income is more unequally distributed or when redistributive concerns are stronger. They are lower when the elasticity of capital to taxes is higher.

FUTURE ISSUES

1. The dynamic Mirrlees approach needs a general theory of approximation of the optimal, often complicated, policies. How does the best approximation depend on the primitives of the economy?
2. More realistic features of the economy—such as general equilibrium effects—should be added to the Mirrlees approach, which will require more quantitative analysis, going beyond calibration and toward full estimation of the underlying parameters.
3. The parametric Ramsey approach, on the other hand, would gain from the exploration of less standard and more complex tax instruments.

4. A middle-ground combination of the strengths of the dynamic Mirrlees approach (i.e., less restricted instruments) and the Ramsey approach (i.e., more realistic quantitative features) would probably be very fruitful.
5. In the sufficient statistics approach, the key challenge is to obtain credible empirical estimates of the relevant longer-run elasticities.
6. As better individual-level tax data have become available over time, credible estimates of the short- or medium-run responses based on policy variation can be obtained.
7. The best way to go may then be to embed these reduced-form estimates into a structural model to estimate the long-run responses.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I thank Clément Herman, Julia Paul-Venturine, and Yannic Rehm for outstanding research assistance.

LITERATURE CITED

Ábrahám Á, Koehne S, Pavoni N. 2011. On the first-order approach in principal-agent models with hidden borrowing and lending. *J. Econ. Theory* 146:1331–61

Ábrahám Á, Koehne S, Pavoni N. 2016. Optimal income taxation when asset taxation is limited. *J. Public Econ.* 136:14–29

Ábrahám Á, Pavoni N. 2008. Efficient allocations with moral hazard and hidden borrowing and lending. *Rev. Econ. Dyn.* 11:781–803

Akcigit U, Hanley D, Stantcheva S. 2016. *Optimal taxation and R&D policies*. NBER Work. Pap. 22908

Albanesi S, Sleet C. 2006. Dynamic optimal taxation with private information. *Rev. Econ. Stud.* 73:1–30

Anderberg D. 2009. Optimal policy and the risk properties of human capital reconsidered. *J. Public Econ.* 93:1017–26

Arrow KJ. 1962. The economic implications of learning by doing. *Rev. Econ. Stud.* 29:155–73

Atkinson AB, Stiglitz JE. 1976. The design of tax structure: direct versus indirect taxation. *J. Public Econ.* 6:55–75

Benabou R. 2002. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica* 70:481–517

Ben-Porath Y. 1967. The production of human capital and the life cycle of earnings. *J. Political Econ.* 75:352–65

Best M, Kleven HJ. 2013. *Optimal income taxation with career effects of work effort*. Work. Pap., Dep. Econ., Lond. Sch. Econ., London

Bohacek R, Kapicka M. 2008. Optimal human capital policies. *J. Monet. Econ.* 55:1–16

Chang Y, Park Y. 2018. *Optimal taxation with private insurance*. Work. Pap., Seoul Natl. Univ., Seoul

Conesa JC, Kitao S, Krueger D. 2009. Taxing capital? Not a bad idea after all! *Am. Econ. Rev.* 99:25–48

Conesa JC, Krueger D. 2006. On the optimal progressivity of the income tax code. *J. Monet. Econ.* 53:1425–50

da Costa CE. 2009. Yet another reason to tax goods. *Rev. Econ. Dyn.* 12:363–76

da Costa CE, Werning I. 2002. *Commodity taxation and social insurance*. Work. Pap., Mass. Inst. Technol., Cambridge, MA

Diamond P, Mirrlees J. 1978. A model of social insurance with variable retirement. *J. Public Econ.* 10:295–336
Erosa A, Gervais M. 2002. Optimal taxation in life-cycle economies. *J. Econ. Theory* 105:338–69
Farhi E, Werning I. 2010. Progressive estate taxation. *Q. J. Econ.* 125:635–73
Farhi E, Werning I. 2013. Insurance and taxation over the life cycle. *Rev. Econ. Stud.* 80:596–635
Findeisen S, Sachs D. 2016. Education and optimal dynamic taxation: the role of income-contingent student loans. *J. Public Econ.* 138:1–21
Findeisen S, Sachs D. 2017. Redistribution and insurance with simple tax instruments. *J. Public Econ.* 146:58–78
Golosov M, Kocherlakota N, Tsyvinski A. 2003. Optimal indirect and capital taxation. *Rev. Econ. Stud.* 70:569–87
Golosov M, Troshkin M, Tsyvinski A. 2016. Redistribution and social insurance. *Am. Econ. Rev.* 106:359–86
Golosov M, Tsyvinski A. 2007. Optimal taxation with endogenous insurance markets. *Q. J. Econ.* 122:487–534
Golosov M, Tsyvinski A, Werquin N. 2014. *A variational approach to the analysis of tax systems*. NBER Work. Pap. 20780
Grochulski B, Piskorski T. 2010. Risky human capital and deferred capital income taxation. *J. Econ. Theory* 145:908–43
Guvenen F, Kambourov G, Kuruscu B, Ocampo-Diaz S, Chen D. 2019. *Use it or lose it: efficiency gains from wealth taxation*. NBER Work. Pap. 26284
Heathcote J, Storesletten K, Violante GL. 2017. Optimal tax progressivity: an analytical framework. *Q. J. Econ.* 132:1693–754
Heathcote J, Storesletten K, Violante GL. 2019. *Optimal progressivity with age-dependent taxation*. NBER Work. Pap. 25617
Hicks J. 1970. Elasticity of substitution again: substitutes and complements. *Oxf. Econ. Pap.* 25:289–96
Kapicka M. 2013. Efficient allocations in dynamic private information economies with persistent shocks: a first-order approach. *Rev. Econ. Stud.* 80:1027–54
Kapicka M. 2015. Optimal Mirrleesean taxation in a Ben-Porath economy. *Am. Econ. J. Macroecon.* 7:219–48
Kapicka M, Neira J. 2014. *Optimal taxation in a life-cycle economy with endogenous human capital formation*. Work. Pap., Univ. Calif., Santa Barbara
Karabarbounis M. 2016. A road map for efficiently taxing heterogeneous agents. *Am. Econ. J. Macroecon.* 8:182–214
Kocherlakota NR. 2005. Zero expected wealth taxes: a Mirrlees approach to dynamic optimal taxation. *Econometrica* 73:1587–621
Koeniger W, Prat J. 2018. Human capital and optimal redistribution. *Rev. Econ. Dyn.* 27:1–26
Krueger D, Ludwig A. 2013. Optimal progressive labor income taxation and education subsidies when education decisions and intergenerational transfers are endogenous. *Am. Econ. Rev.* 103:496–501
Makris M, Pavan A. 2019. *Taxation under learning-by-doing*. Work. Pap., Northwest. Univ., Evanston, IL
Milgrom P, Segal I. 2002. Envelope theorems for arbitrary choice sets. *Econometrica* 70:583–601
Mirrlees J. 1971. An exploration in the theory of optimum income taxation. *Rev. Econ. Stud.* 38:175–208
Ndiaye A. 2017. *Flexible retirement and optimal taxation*. Work. Pap. WP-2018-18, Fed. Reserve Bank, Chicago
Pavan A, Segal I, Toikka J. 2014. Dynamic mechanism design: a Myersonian approach. *Econometrica* 82:601–53
Perrault L. 2015. *Optimal income taxation with unemployment induced loss of human capital*. Work. Pap., Ga. State Univ., Atlanta
Rogerson WP. 1985. The first-order approach to principal-agent problems. *Econometrica* 53:1357–67
Saez E. 2001. Using elasticities to derive optimal income tax rates. *Rev. Econ. Stud.* 68:205–29
Saez E, Stantcheva S. 2016. Generalized social marginal welfare weights for optimal tax theory. *Am. Econ. Rev.* 106:24–45
Saez E, Stantcheva S. 2018. A simpler theory of optimal capital taxation. *J. Public Econ.* 162:120–42
Samuelson PA. 1974. Complementarity: an essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *J. Econ. Lit.* 12:1255–89
Stantcheva S. 2014. *Optimal taxation and human capital policies over the life cycle*. Job Mark. Pap., Harvard Univ., Cambridge, MA

Stantcheva S. 2015a. *Learning and (or) doing: human capital investments and optimal taxation*. NBER Work. Pap. 21381

Stantcheva S. 2015b. *Optimal income, education and bequest taxes in an intergenerational model*. NBER Work. Pap. 21177

Stantcheva S. 2017. Optimal taxation and human capital policies over the life cycle. *J. Political Econ.* 125:1931–90

Weinzierl M. 2011. The surprising power of age-dependent taxes. *Rev. Econ. Stud.* 78:1490–518

Werning I. 2002. *Optimal unemployment insurance with unobservable savings*. Work. Pap., Mass. Inst. Technol., Cambridge, MA

Annual Review of Economics

Volume 12, 2020

Contents