

Optimal Income Taxation with Career Effects of Work Effort*

Michael Carlos Best & Henrik Jacobsen Kleven[†]

February 2013

Abstract

The literature on optimal income taxation assumes that wage rates are generated exogenously by innate ability and therefore do not respond to behavior and taxation. This is in stark contrast to a large empirical literature documenting a strong effect of current work effort on future wage rates. We extend the canonical Mirrleesian optimal tax framework to incorporate such career effects and provide analytical characterizations that depend on estimable entities. Besides the standard static earnings elasticity with respect to the marginal tax rate, the optimal tax schedule also depends on the elasticity of future wages with respect to current work effort. We explore the empirical magnitude of this “career elasticity” in a meta-analysis of the literature on the returns to work experience and tenure, concluding that a reasonable value for this elasticity lies between 0.2 and 0.4. Calibrating the model to US micro data (under reasonable values of the career elasticity), we present numerical simulations of optimal nonlinear tax schedules that depend on per-period earnings and potentially on age. In the case of age-independent taxation, the presence of career effects make the tax schedule substantially less progressive than in standard models with exogenous wage rates. In the case of age-dependent taxation, career effects create a strong argument for lower taxes on the old, opposite the recommendation in the recent literature on age-dependent taxation. This result reflects both a career incentive effect and an equity effect, where the latter effect arises because increasing earnings over the career path for each ability level imply that, *conditional on earnings*, age and ability are negatively correlated.

JEL Codes: H21, H24, J22, J24.

*We would like to thank Robin Boadway, Peter Diamond, Mike Golosov, Bas Jacobs, Vilen Lipatov, Alan Manning, Emmanuel Saez and Johannes Spinnewijn for discussions and comments. All errors remain our own.

[†]Best: STICERD & Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK (m.c.best@lse.ac.uk). Kleven: STICERD & Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK (h.j.kleven@lse.ac.uk)

Whatever muscles I have are the product of my own hard work and nothing else

-Evelyn Ashford: Olympic 100m champion

When I was young, I observed that nine out of ten things I did were failures. So I did ten times more work.

-George Bernard Shaw: Nobel laureate in literature

1 Introduction

The modern literature on optimal income taxation is cast in the Mirrleesian framework in which innate ability generates a wage rate that is exogenous and therefore unrelated to individual behavior and taxation. This holds both for static versions of the framework (e.g. [Mirrlees, 1971](#); [Diamond, 1998](#); [Saez, 2001](#)) and for recent dynamic versions (e.g. [Golosov et al., 2007, 2011](#); [Farhi & Werning, 2012](#)) in which the wage rate is allowed to change over time in a potentially non-deterministic fashion, but never depends on behavior. In this literature, earnings in any period of the life cycle respond to taxation only through contemporaneous changes in hours worked. This assumption stands in sharp contrast to a large body of work in labor economics studying the various ways in which current behavior—including work effort—affects future wages. Motivated by this research, we explore the optimal tax implications of breaking the simple mapping between abilities and wages by allowing current hours worked to affect future productivity and wages.

The link between work effort and future wages is widely documented in a vast literature in labor economics. This literature studies the relationship between the wage rate and various measures of work experience, including potential experience (age minus schooling), actual experience, tenure in an individual’s current job, and experience lost as a result of job losses (see [Blundell & MaCurdy, 1999](#) and [Farber, 1999](#) for surveys). Conceptually, a variety of mechanisms are likely to be in operation such as improvements in general and firm-specific human capital ([Ben-Porath, 1967](#)), improvements in employer-employee matches (e.g. [Manning, 2000](#)) and ability signaling effects ([Holmström, 1999](#)). In this paper, we capture all the channels through which current labor supply affects future wage rates in a simple reduced-form relationship, which keeps the otherwise very complicated dynamic optimal taxation problem tractable and allows us to obtain transparent analytical results that depend on empirical entities.

To explore the empirical magnitude of these effects, we conduct a meta-analysis of seventeen empirical studies that permit the derivation of an estimate of the elasticity of future wages with respect to current work effort—the parameter that we show is crucial for optimal income taxation. We find that 80% of the 108 estimates of this “career elasticity” lie between 0.19 and 0.38, implying that an additional 10% of work effort when young raises wages when old by between 2 and 4%. These effects are strong enough to have important qualitative and quantitative implications for optimal tax schedules.

Our paper also contributes to the recent debate about age-dependent taxation, as reviewed by [Banks & Diamond \(2011\)](#) in the recent *Mirrlees Review*. This work argues that age constitutes a useful tagging device ([Akerlof, 1978](#)), which can be used to relax the incentive compatibility constraints of the optimal income tax problem. For instance, applying the static Mirrlees model separately to different age groups, [Kremer \(2001\)](#) argues that earnings distributions and labor supply elasticities are so different across ages that the implied pattern of optimal tax rates would vary greatly by age. More recently, the dynamic optimal tax literature considers this question ([Weinzierl, 2011](#); [Golosov et al., 2011](#); [Farhi & Werning, 2012](#)) and finds that age-dependent tax schedules with higher tax rates on older workers are welfare-improving and able to realize most of the gains from a fully optimized history-dependent tax schedule. A key reason for the power of age-dependence in this literature is the fact that the observed wage distribution of older workers features both a higher mean and a higher variance than the wage distribution of younger workers ([Weinzierl, 2011](#)). Seen through the lens of the Mirrlees model, this translates to differences in the mean and variance of the ability distribution that creates an equity and insurance argument for higher taxes on the old. What this argument neglects is that the difference in the wage distributions of the young and the old reflects, not differences in exogenous ability, but the fact that the young and the old are observed at different stages of their (endogenous) careers. This is the issue that forms the basis of our paper, and we show that it can reverse previous conclusions in the literature.¹

As our framework of analysis, we consider a two-period Mirrlees model in which the wage rate as young equals innate ability while the wage rate as old is a general function of innate ability and hours worked as young. The young and the old have drawn their abilities from the same underlying ability distribution, but face different wage rates for two reasons. One reason is that effort as young serves as an investment in labor productivity as old (behavioral career effect). The other reason is that, independently of individual behavior, a given innate

¹[Weinzierl \(2011\)](#) also discusses the importance of modeling the endogeneity of wage paths in order to fully evaluate the case for age-dependent taxation.

ability may be associated with an age-varying wage profile rather than a constant wage over the career (what we call a mechanical career effect).²

We show that the presence of behavioral career effects provides a plausible micro-foundation for the well-documented empirical fact that labor supply elasticities are larger for older workers than for younger workers (e.g. [Blundell & MaCurdy, 1999](#)). Since the young are working to raise future wages as well as for consumption in the present while the old are working only to finance consumption in the present, the labor supply of the young is naturally less elastic than the labor supply of the old under the same preferences. Besides these implications for the *own-tax* elasticities of the young and the old, career effects have implications for the *cross-tax* elasticities as, for example, lower taxes on the old induce the young to work harder due to the effort investment effect, what we label the *aspiration effect* in the paper.

We consider a preference structure allowing us to bypass issues related to savings and capital taxation, and provide analytical characterizations of the optimal taxation of labor earnings that relate in intuitive and transparent ways to existing results without career effects.³ These characterizations show that the optimal tax schedule can be expressed as a function of long-run earnings elasticities for the young and the old that incorporate the implications of endogenous career paths. Since such long-run earnings responses are not what is captured by the empirical labor supply and taxable income literatures using short-run variation in micro data (as pointed out by, e.g., [Piketty & Saez, 2013](#)), we show that the relevant long-run elasticities depend on two underlying sufficient statistics: the standard static earnings elasticity with respect to the marginal net-of-tax rate (as estimated in the enormous taxable income elasticity literature) and the elasticity of future wages with respect to current work effort (the magnitude of which can be inferred from the large literature on the returns to work experience and tenure).

For the case of age-dependent taxation, this framework brings to the fore two important effects that have been ignored in previous optimal tax analyses. First, in the empirically relevant case of increasing wage profiles over the career, an old worker of a given ability level has a higher wage rate and higher earnings than a young worker of the same ability level. As a consequence, an old worker at a given *earnings* level must be of lower ability than young workers at the same earnings level. Therefore, conditional on earnings, age is negatively correlated with ability which creates a classical tagging argument for supplementing

²Here we will assume that this wage profile is deterministic, but there is no reason that our framework could not be extended to allow for it to be stochastic.

³Consistent with real-world tax policy, we focus on annual tax schedules that involve separate taxation of earnings in different time periods—but may be age-dependent—rather than fully history-dependent tax schedules.

an earnings-based income tax with a tax break to older workers. Second, the presence of behavioral career effects create an efficiency argument for lower taxes later in the career, an effect that operates through the own-tax and cross-tax elasticities of labor supply described above. In particular, lower taxes on the old are desirable both because older workers are relatively elastic with respect to their contemporaneous tax rate and because younger workers are elastic with respect to their future tax rate via the aspiration effect. In summary, both the age-ability correlation effect and the elasticity effects call for age-dependent taxation with lower income tax rates on older workers.⁴ This is directly opposite to the policy recommendation in the recent optimal tax literature, but is consistent with the policy debate outside economics in which age-dependence is typically discussed in the context of tax breaks for older workers.⁵

When taxes are constrained to be age-independent, we show that the optimal schedule of marginal tax rates can be written as a weighted average of the two optimal age-dependent marginal tax rate schedules. Since earnings increase over the career path, at higher income levels a greater fraction of the population is old and so the weight placed on the old relative to the young in the optimal marginal tax schedule is increasing in income. As the optimal age-dependent marginal tax rates are lower on old workers, the increasing weight on the old makes the optimal age-independent marginal tax rate schedules flatter—less progressive—than in the standard model with exogenous wage paths.

In order to ascertain the quantitative implications of the new effects we have identified, we carry out numerical simulations based on data for the United States, extending the simulation method set out by [Saez \(2001\)](#) to a setting with career effects. The simulations for age-dependent tax schedules reaffirm the theoretical arguments made above. In a setting with no behavioral career effects (but mechanical career effects generating an increasing wage profile over the life cycle at a given ability), the optimal tax system features a weak degree of age dependence with slightly higher taxes on older workers. However, even very modest behavioral career effects are sufficient to reverse this result and generate lower taxes on older workers. Under realistic assumptions about the strength of career effects (based on our empirical meta-study), it is possible to generate very strong age dependence with much lower

⁴Besides these two effects, a third offsetting effect is driven by the different hazard ratios of the earnings distributions of young and old workers. Optimal marginal tax rates on earnings depend positively on such hazard ratios (see e.g. [Saez, 2001](#), in the context of the standard Mirrlees model), and the empirical fact that earnings distributions of old workers feature higher hazard ratios than earnings distributions of young workers makes it more efficient to tax the old than the young, other things equal. This is precisely the effect that is central to the results in [Kremer \(2001\)](#) and [Weinzierl \(2011\)](#), as discussed above, but in our analysis it is not sufficiently strong to overturn the other arguments calling for lower taxes on the old.

⁵For example, the UK tax system involves limited age-dependence favoring old workers, and the *Mirrlees Review* proposes to go further in this direction.

taxes on older workers. This result is driven by the age-ability correlation and elasticity effects discussed above.⁶ The simulation results for age-independent tax schedules show that even modest career effects can have substantial impacts on optimal marginal tax rates, which are lower and flatter than in the absence of career effects.

The idea that work effort represents an investment in higher future wages (for example via learning by doing) is related to the large literature on human capital investments. Since the implications of standard human capital investments (formal education) for optimal taxation have been explored in earlier work (e.g. [Eaton & Rosen, 1980](#); [Bovenberg & Jacobs, 2005](#)), it is important to note that the tax implications of learning by doing are fundamentally different from the implications of education. First, education and work represent two substitutable uses of time, and the key cost of education is therefore the opportunity cost of foregone net-of-tax earnings during education. This implies that education costs are effectively tax deductible in which case income taxation need not distort human capital investments at all ([Eaton & Rosen, 1980](#)). By contrast, since learning by doing is a byproduct of work effort, income taxation will always distort this form of human capital investment. Second, formal education is an activity that can be observed and therefore directly subsidized or taxed by the government, whereas learning by doing cannot be separated from labor supply and so cannot receive a separate tax treatment. For both of these reasons, models of optimal taxation with endogenous education are conceptually very different from our framework and do not shed light on the issues that we highlight in this paper. As far as we are aware, the only previous paper that allows for learning-by-doing effects in the context of optimal income taxation is [Krause \(2009\)](#), who focuses on the implications of such effects for the no-distortion-at-the-top result in the context of a two-type [Stiglitz \(1982\)](#) model.

We will proceed as follows. Section 2 presents the setting and shows the implications of career effects for earnings elasticities. Section 3 characterizes optimal income tax schedules and discusses the implications of career effects for both age-dependent and age-independent taxation. Section 4 investigates empirically the career effect of work effort based on a meta-analysis of the literature on experience and tenure effects. Section 5 presents numerical simulations that demonstrate the quantitative importance of career effects for optimal tax policy, and finally section 6 concludes.

⁶In particular, these two effects dominate the effect coming from the difference in the earnings distributions of the young and the old (what we will call the hazard ratio effect), which is what drove the previous findings that age dependence should feature higher taxes on older workers.

2 The Setting

2.1 Individuals

We analyze the simplest possible setting that allows us to explore the implications of career effects for optimal tax schedules. Individuals live for 2 periods, $i \in \{y, o\}$, work in both of them and at any point in time there is a continuum of mass 1 of individuals of each age alive. They have time separable preferences with no discounting and their per-period utility is quasi-linear and given by $u(c_i, l_i) = c_i - \frac{1}{1+1/e} l_i^{1+1/e}$. This formulation has the virtue that individuals will not save and so we can focus on the analysis of wage effects without the additional complication of saving effects.

In the first period of life, individuals are paid according to their innate ability n , the distribution of which is given by the cdf $F(n)$. Therefore, earnings when young are $z_y = n l_y(n)$. Our key innovation is to allow the second-period wage to depend both on innate ability and on the first period's effort choice. We allow this effect to manifest itself in a very general way, merely positing that the wage rate when old ω is a general function of innate ability and first-period effort, i.e. $\omega = \omega(n, l_y)$. Earnings when old are then given by $z_o = \omega(n, l_y) l_o$. The responsiveness of the wage rate when old to innate ability may be captured by the elasticity $\eta = \frac{\partial \omega}{\partial n} \frac{n}{\omega}$ and reflects the *mechanical career effect* of higher ability on the life-cycle profile of wages. The responsiveness of wages when old to effort when young is captured by the elasticity $\delta = \frac{\partial \omega}{\partial l_y} \frac{l_y}{\omega} = \frac{\partial \omega}{\partial z_y} \frac{z_y}{\omega}$ and reflects the *behavioral career effect* due to the investment component of work effort as young. When we turn to simulations of the optimal tax schedules in section 5, we will assume that we are in the empirically plausible case where $\eta, \delta \geq 0$, however note that this restriction is not necessary for our derivations of the optimal tax schedules.

Since there are no savings, consumption at age i is simply equal to earnings net of income taxes at that age, i.e. $c_i = z_i - T_i(z_i)$. The income tax liability at age i , $T_i(z_i)$, depends on earnings at that age (but not on earnings at other ages) and possibly on age itself (as the $T_i(\cdot)$ function is allowed to vary with i). This is consistent with real-world tax schedules, which are always based on annual income and sometimes feature aspects of age-dependence (see, for example, the Mirrlees Review for a description of age-dependence in the UK tax system).

Lifetime utility is given by

$$U(z_y, z_o) = z_y - T_y(z_y) - \frac{1}{1+1/e} \left(\frac{z_y}{n}\right)^{1+1/e} + z_o - T_o(z_o) - \frac{1}{1+1/e} \left(\frac{z_o}{\omega(n, z_y/n)}\right)^{1+1/e} \quad (1)$$

which has first-order conditions for earnings chosen when young and when old given by

$$1 - \tau_y(z_y) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} = 0 \quad (2)$$

and

$$1 - \tau_o(z_o) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \frac{1}{\omega} = 0 \quad (3)$$

where $\tau_i(z) \equiv T'_i(z)$ is the marginal tax rate on earnings in period i .

2.2 Earnings Elasticities

To facilitate interpretation of our main results, this section starts by characterizing the relationship between the strength of career effects and earnings elasticities for the young and the old.⁷ At the extreme, when there are no behavioral career effects ($\delta = 0$), this model reduces to a simple two-period version of a standard optimal income tax model like that studied in [Diamond \(1998\)](#). In particular, the young are responsive only to the tax schedule they face when young even though they know the tax schedule they will face when old, and similarly for the old. This is because their behavior when young does not affect the decision-making problem as old, and vice versa. Moreover, it is easy to see from the first-order conditions (2) and (3) that the elasticity of earnings at age i with respect to the marginal net-of-tax rate at that age, $1 - \tau_i$, is given by the utility parameter e for both age groups. However, when we introduce career effects through $\delta > 0$, this changes.

We define the elasticity of earnings at age i with respect to the marginal net-of-tax rate at age j as $E_{ij} \equiv \frac{dz_i}{d(1-\tau_j)} \frac{1-\tau_j}{z_i}$. Applying the implicit function theorem to the pair of first-order conditions (2) and (3), [Appendix A](#) shows that the earnings elasticities can be expressed as

$$\begin{pmatrix} E_{yy} & E_{yo} \\ E_{oy} & E_{oo} \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} e & e(1+e) \delta \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)} \\ e(1+e) \delta & e \left[1 + \delta(1+\delta)(1+e) \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)} \right] \end{pmatrix} \quad (4)$$

where $\kappa \equiv 1 + \delta(1 - e\delta)(1 + e) \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)}$. The elasticities E_{yy} and E_{oo} are *contemporaneous* earnings elasticities of the young and the old with respect to the marginal net-of-tax rates faced at those respective ages, while E_{yo} and E_{oy} are *intertemporal* earnings elasticities of the young and the old that reflect the presence of career effects. The elasticity E_{yo} reflects what we refer to as the *aspiration effect*: since part of the return to current work effort is

⁷Throughout the paper, we focus on earnings elasticities (including hours-worked and wage-rate effects) rather than hours-worked elasticities, because it is the former elasticity concept that matters directly for optimal tax schedules. However, the main qualitative properties of earnings elasticities that we characterize in this section also applies to hours-worked elasticities.

higher future wages, and individuals anticipate the rate at which those future wages will be taxed, a higher tax rate later in life reduces the career investments made through work effort earlier in life. The elasticity E_{oy} reflects what we refer to as the *accumulation effect*: a higher tax rate on the young reduces work effort and therefore earnings by the young, which has a negative knock-on effect on the wage rate and labor supply of those individuals when they become old.

In the following, we present three lemmas that clarify the precise link between the career effect δ and the size of earnings elasticities. The proofs of these lemmas are provided in appendix A. The first lemma shows how the contemporaneous responsiveness of the two age groups is affected by the presence of career effects:

Lemma 1. *In the absence of behavioral career effects, $\delta = 0$, the contemporaneous earnings elasticities of the young and the old are given by $E_{yy} = E_{oo} = e$. In the presence of behavioral career effects, $\delta > 0$, the contemporaneous earnings elasticity of the young is lower while that of the old is larger than in the absence of such effects, i.e. $E_{yy} < e$ and $E_{oo} > e$.*

Intuitively, the young are working both for current wages (taxed at rate τ_y) and to raise their wages when old (taxed at rate τ_o), and so their earnings are naturally less elastic to their tax rate as young than is implied by the standard static elasticity e . Meanwhile, the earnings of the old respond to the tax rate when old both through a standard static hours-of-work response governed by the e -parameter and through a dynamic wage-rate response coming from the effect of the tax rate when old on the incentive while young to invest in higher wages as old. Notice that these earnings elasticities (and those discussed below) reflect full dynamic effects on earnings at different ages by taxpayers who plan their entire life cycle profile of earnings, perfectly anticipating the tax schedule faced in each period. These are, of course, the relevant elasticities to consider for the optimal tax analysis that follow, which focuses on the optimal tax policy by a government that can fully commit to future tax rates.

Next, we turn to the implications of career effects for the aspiration and accumulation elasticities:

Lemma 2. *In the absence of behavioral career effects, $\delta = 0$, the aspiration and accumulation elasticities are zero, i.e. $E_{yo} = E_{oy} = 0$. In the presence of behavioral career effects, $\delta > 0$, the aspiration and accumulation elasticities are positive and always increasing in the strength of the career effect, i.e. $\frac{\partial E_{yo}}{\partial \delta} > 0$ and $\frac{\partial E_{oy}}{\partial \delta} > 0$.*

The intuition behind these results follows naturally from the fact that, in this model, it is precisely the effect of current work effort on future wage rates that creates an intertemporal link between taxation and earnings across different periods. With positive career effects of

work effort, earnings in one period respond positively to the net-of-tax rate in another period, and the size of this response is increasing in the size of the career effect δ .

The elasticities considered so far measure earnings responses as young or old to the tax rate in one period of life taking as given the tax rate in the other period. It is useful to also consider *total* earnings responses by the young and the old to a change in the tax rate in both periods of life. Defining the total elasticity of earnings at age i as $E_i \equiv E_{iy} + E_{io}$, we can state the following:

Lemma 3. *In the presence of behavioral career effects, $\delta > 0$, the total elasticity of earnings in each period is larger than the standard static elasticity, i.e.*

$$E_y \equiv E_{yy} + E_{yo} > e$$

$$E_o \equiv E_{oy} + E_{oo} > e$$

Moreover, with $\delta > 0$, the total elasticity of earnings as old is larger than the total elasticity of earnings as young, i.e.

$$E_o > E_y$$

These results demonstrate that career effects increase the overall responsiveness of earnings to income taxation and therefore exacerbate the efficiency costs of taxation. Moreover, the degree to which career effects increase the responsiveness of earnings is stronger for the old than for the young. This last result not only provides an interesting and plausible micro-foundation for the often reported finding that labor supply and earnings elasticities are larger for the old than for the young (see, for example [Blundell & MaCurdy, 1999](#)),⁸ it also has potentially important implications for optimal tax structure and in particular the desirability and design of age-dependent taxes.

2.3 The Government

We consider a government imposing an “annual” income tax that may or may not depend on the age of the taxpayer. That is, an individual’s tax liability in a given period depends exclusively on within-period income and possibly on age. This is analytically and conceptually different from considering a government choosing fully history-dependent tax schedules in which an individual’s tax liability when old may depend directly on income earned when young. We focus on annual age-dependent tax schedules rather than fully history-dependent

⁸The relationships in Lemma 3, which are stated in terms of earnings elasticities, also apply to hours-of-work elasticities.

schedules, because the former is empirically more relevant: real-world income tax systems are based on annual time-separable tax liability and occasionally involves some age-dependence, but are in general not history dependent.⁹ While at present, age-dependence in the income tax system is used either in a very limited fashion in some countries (e.g. United Kingdom) or not at all in other countries (e.g. United States), it is interesting to analyze because of several recent proposals to introduce age as a tagging device in tax systems. We characterize optimal tax policy both when full age dependence is allowed (general schedules $T_y(z), T_o(z)$) and when no age dependence is allowed (schedules $T_y(z) = T_o(z) = T(z) \forall z$). We assume throughout that the government can fully commit to future tax rates.

In the case of age-dependence, the government chooses tax schedules for the young and the old $T_y(z), T_o(z)$ to maximize social welfare subject to incentive compatibility constraints and a revenue-raising constraint, i.e.

$$\begin{aligned} \max_{T_y(z), T_o(z)} \int_0^\infty \Psi [U(z_y(n), z_o(n))] dF(n) \\ \text{s.t. } \{z_y(n), z_o(n)\} \in \arg \max U(z_y, z_o) \forall n \\ \int_0^\infty T_y(z_y(n)) dF(n) + \int_0^\infty T_o(z_o(n)) dF(n) \geq R \end{aligned}$$

where $\Psi[\cdot]$ is an additively separable social welfare function defined over the lifetime utility of individuals, R is an exogenous revenue requirement, and the size of each generation has been normalized to 1. The government's redistributive tastes may be captured by social welfare weights equal to the social marginal utility of income for different individuals expressed in terms of the marginal value of public funds. For an individual of ability n , the social welfare weight is defined as $g(n) \equiv \Psi' [U(z_y(n), z_o(n))] / \lambda$ where λ is the Lagrange multiplier on the government budget constraint, the marginal value of public funds. It will be useful to translate this welfare weight from being a function of ability to being a function of income, so we also define $g_y(z) \equiv \Psi' [U(z, z_o(z))] / \lambda$ and $g_o(z) \equiv \Psi' [U(z_y(z), z)] / \lambda$, where $z_o(z)$ are the equilibrium earnings when old of an individual who earns z when young and $z_y(z)$ are the equilibrium earnings when young of an individual who earns z when old. $z_y(z), z_o(z)$ are increasing functions of z as long as $z_y(n), z_o(n)$ are increasing functions of n . As in the standard Mirrlees model, the condition that $z_y(n), z_o(n)$ are increasing in n is necessary and sufficient to ensure that a given path for $z_y(n), z_o(n)$ can be implemented by a truthful

⁹There is some history dependence in social security systems, which matters for retirement decisions. But here we focus on income taxation and do not model retirement.

mechanism or, equivalently, by a nonlinear tax system. The analytical characterization in section 3 assumes that this condition is satisfied while section 5 verifies this numerically.

3 Optimal Tax Schedules

This section characterizes analytically the implications of career effects for the optimal nonlinear tax schedule in the age-dependent and age-independent cases. We derive optimal tax formulas using both Hamiltonian and tax perturbation approaches, where the latter is particularly useful for facilitating economic intuition about the role of different effects. The optimal marginal tax rates are expressed in terms of entities that are observable or estimable in the manner of Diamond (1998) and Saez (2001), which lends itself naturally to a calibration exercise as considered in section 5. As we describe in detail below, the implications of behavioral career effects for optimal income taxation can be split into *elasticity effects* coming from how careers affect the responsiveness of earnings by the young and the old to taxes, a *welfare weight effect* coming from how careers affect the social marginal utilities of income of the young and the old, and a *hazard ratio effect* coming from how careers generate different earnings distributions for the young and the old.

3.1 Optimal Age-Dependent Taxes

In this section we characterize the optimal age-dependent, nonlinear income tax schedule $\{T_y(z), T_o(z)\}$ with corresponding marginal tax rate schedules $\{\tau_y(z), \tau_o(z)\}$. We can show:

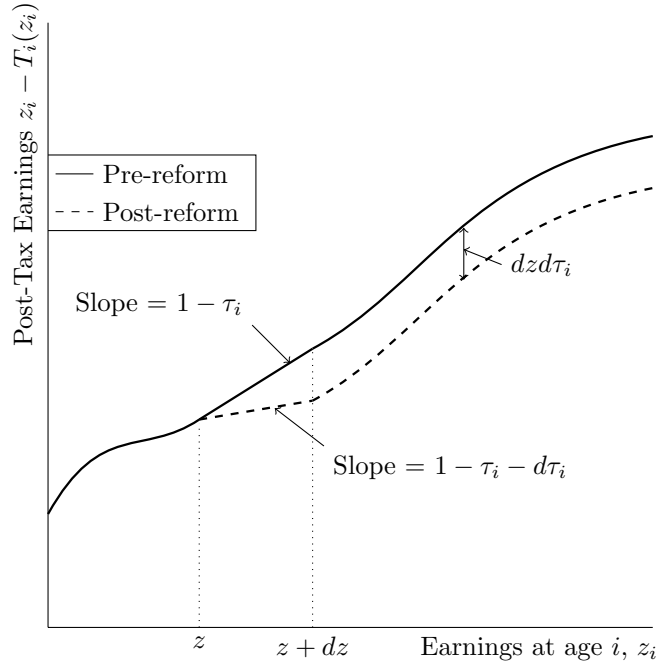
Proposition 1. *The optimal age-dependent tax schedule, $T_i(z)$ at age $i \in \{y, o\}$, is associated with marginal tax rates*

$$\frac{\tau_i(z)}{1 - \tau_i(z)} = A_i(z) B_i(z) C_i(z) \tag{5}$$

where (for $i \neq j$) we have

$$\begin{aligned} A_i(z) &= \left\{ E_{ii} + E_{ji} \frac{\tau_j(z_j(z)) z_j(z)}{\tau_i(z) z} \right\}^{-1} \\ B_i(z) &= \frac{\int_z^\infty [1 - g_i(z')] dH_i(z')}{1 - H_i(z)} \\ C_i(z) &= \frac{1 - H_i(z)}{zh_i(z)} \end{aligned}$$

Figure 1: Non Linear Tax Rate Perturbation



at any earnings level z . In these expressions, $H_i(\cdot)$ and $h_i(\cdot)$ denote the equilibrium cdf and pdf, respectively, of earnings at age i .

Proof. Here we prove the result directly using a tax perturbation method (as first developed by Piketty, 1997; Saez, 2001), first for the young and then for the old as this illustrates the intuition for the results better. A technically more rigorous proof based on the Hamiltonian approach is found in appendix B alongside a proof that the two methods produce equivalent results in the context of our model. For the young and the old separately, consider a small perturbation around the optimal tax schedule as depicted in Figure 1. The perturbation increases the marginal tax rate by a small amount $d\tau_i$ at age i on incomes falling in a small band $(z, z + dz)$ but is otherwise left unchanged.

The tax schedule of the young

We first consider the perturbation in the tax schedule of the young. The marginal tax rate increase $d\tau_y$ in the small band $(z, z + dz)$ has a mechanical effect on tax revenue and welfare for all young individuals above z as well as two behavioral effects on those with earnings between z and $z + dz$ as young. We proceed to analyze the three effects separately:

Mechanical Welfare Effect All young taxpayers with earnings above z pay $d\tau_y dz$ more in taxes (holding behavior constant), which creates a mechanical revenue gain for the government but reduces the utility of those individuals. The net social welfare effect of the mechanical tax increase of a young individual with income z' is given by $d\tau_y dz \cdot [1 - g_y(z')]$. Hence, the total mechanical effect on social welfare is given by

$$\Delta_y^M = d\tau_y dz \cdot \int_z^\infty [1 - g_y(z')] dH_y(z')$$

Contemporaneous Earnings Effect Using the definition of the contemporaneous earnings elasticity of the young E_{yy} in section 2.2, each young person in the band $(z, z + dz)$ reduces earnings by $-E_{yy} \cdot \frac{d\tau_y}{1 - \tau_y(z)} \cdot z$. Multiplying the earnings response by the marginal tax rate $\tau_y(z)$, we get the change in tax liability by each individual in this band. As there are $h_y(z) dz$ young individuals in the band, the total effect of contemporaneous earnings responses on tax revenue is given by

$$\Delta_y^E = -d\tau_y dz \cdot z h_y(z) \cdot E_{yy} \cdot \frac{\tau_y(z)}{1 - \tau_y(z)}$$

Accumulation Effect The labor supply response of young workers located in the band $(z, z + dz)$ affects human capital accumulation and therefore the wage rate and earnings of those young workers when they become old. As established earlier, a given tax system is associated with a mapping between earnings as young and earnings as old, so that a person with earnings z as young has earnings $z_o(z)$ as old. This implies that changing the tax rate on young workers at income level z has an accumulation effect on old workers at income level $z_o(z)$. Using the definition of the accumulation elasticity E_{oy} , an old person at $z_o(z)$ reduces earnings by $-E_{oy} \cdot \frac{d\tau_y}{1 - \tau_y(z)} \cdot z_o(z)$. The number of old workers whose earnings change as a result of this accumulation effect (those in the band $(z_o(z), z_o(z + dz))$ of the distribution $h_o(z_o)$) is equal to the number of young workers who changed their labor supply in response to the higher tax rate on the young (those in the band $(z, z + dz)$ of the distribution $h_y(z_y)$), i.e. we have $h_o(z_o(z)) \frac{dz_o}{dz} dz = h_y(z) dz$, and therefore the total effect on tax revenue due to the accumulation effect on all old workers affected is given by

$$\Delta^{AC} = -d\tau_y dz \cdot z_o(z) h_y(z) \cdot E_{oy} \cdot \frac{\tau_o(z_o(z))}{1 - \tau_y(z)}$$

Optimality At the optimal tax schedule, there should be no first-order welfare effect of this perturbation, and so we have

$$\Delta_y^M + \Delta_y^E + \Delta^{AC} = 0$$

Inserting the above expressions and rewriting gives the following optimality condition on the tax schedule for the young

$$\frac{\tau_y(z)}{1 - \tau_y(z)} = \left\{ E_{yy} + E_{oy} \cdot \frac{\tau_o(z_o(z)) z_o(z)}{\tau_y(z) z} \right\}^{-1} \cdot \frac{\int_z^\infty [1 - g_y(z')] dH_y(z')}{z h_y(z)}$$

which, after multiplying and dividing by $1 - H_y(z)$, is equivalent to the expression in Proposition 1 for $i = y$.

The tax schedule of the old

As in the tax perturbation for the young, the marginal tax rate increase on the old $d\tau_o$ in the band $(z, z + dz)$ gives rise to a mechanical welfare effect above z along with two behavioral effects on those between z and $z + dz$ as old. The mechanical welfare effect on the old is analogous to the expression for the young:

$$\Delta_o^M = d\tau_o dz \cdot \int_z^\infty [1 - g_o(z')] dH_o(z')$$

There is also a contemporaneous earnings effect on the old taking the same form as for the young:

$$\Delta_o^E = -d\tau_o dz \cdot z h_o(z) \cdot E_{oo} \cdot \frac{\tau_o(z)}{1 - \tau_o(z)}$$

Finally, instead of the accumulation effect of the tax perturbation for the young, we have an aspiration effect of the tax perturbation for the old.

Aspiration Effect The higher tax rate on old workers in the earnings band $(z, z + dz)$ discourages young workers who anticipate being in this band when old from investing in future productivity and earnings. Using the mapping between earnings as young and earnings as old, this behavioral effect on the young occurs in the earnings band $(z_y(z), z_y(z + dz))$. The change in earnings by each young worker who is affected equals $-E_{yo} \cdot \frac{d\tau_o}{1 - \tau_o(z)} \cdot z_y(z)$. The number of young workers affected (those in the band $(z_y(z), z_y(z + dz))$ of the distribution $h_y(z_y)$) is equal to the number of old workers facing a higher marginal tax rate (those in the band $(z, z + dz)$ of the distribution $h_o(z_o)$), so that $h_y(z_y(z)) \frac{dz_y}{dz} dz = h_o(z) dz$. This implies

that the total effect on tax revenue due to the aspiration effect can be written as

$$\Delta^{AS} = -d\tau_o dz \cdot z_y(z) h_o(z) \cdot E_{yo} \frac{\tau_y(z_y(z))}{1 - \tau_o(z)}$$

Optimality At the social optimal, we have

$$\Delta_o^M + \Delta_o^E + \Delta^{AS} = 0$$

which gives the expression in Proposition 1 for $i = o$. \square

We have thus characterized the optimal tax schedule in terms of two expressions that share several qualitative features with the standard formulas in Diamond (1998) and Saez (2001), but with some important differences that bear fleshing out. We will discuss these in the context of their implications for the optimal form and degree of age dependence in the tax system.

3.2 Age Dependence in the Optimal Tax System

The existence of career effects of work effort has implications for all three terms in the optimal income tax formula (5): the inverse elasticity term A_i , the welfare weight term B_i and the hazard ratio term C_i . Considering each of these terms separately, we now discuss the implications of career effects for optimal income tax structure. We emphasize how career effects change the three key terms in different ways for the young and the old, and therefore have important effects on the optimal form and degree of age dependence in the tax system.

The *elasticity effect* of careers operates through $A_y(z)$ and $A_o(z)$. For the taxation of old workers ($A_o(z)$ term), Lemmas 1 & 2 show that career effects $\delta > 0$ give rise to a contemporaneous earnings elasticity for the old that is larger than the standard static elasticity, $E_{oo} > e$, as well as a positive aspiration elasticity for the young, $E_{yo} > 0$. The combination of these effects imply $A_o(z) < 1/e$, so that the inverse elasticity term for old is always smaller than in standard models without career effects. This calls for lower taxes on the old, other things equal. For the taxation of young workers ($A_y(z)$ term), Lemmas 1 & 2 show that $\delta > 0$ implies a contemporaneous earnings elasticity for the young that is smaller than the standard elasticity, $E_{yy} < e$, along with a positive accumulation elasticity on the old, $E_{oy} > 0$. Hence, depending on the magnitudes of these elasticities, $A_y(z)$ may be either below or above $1/e$. Due to fact that the elasticities E_{yy}, E_{oy} (see equation 4) and the weighting term on E_{oy} in the optimal tax formula are endogenous to the tax system itself, it is not possible to analytically determine if $A_y(z)$ is smaller or greater than $1/e$. Nevertheless,

our numerical simulations (discussed in section 5) show that $A_y(z) \gtrsim 1/e$ under a wide range of reasonable parameter assumptions, so that the elasticity effect of careers calls for either unchanged or higher taxes on the young, other things equal. The combination of these insights imply that the elasticity effect on its own calls for age-dependent taxes with lower taxes on the old than on the young, conditional on earnings.

The *welfare weight effect* of careers operates through the terms $B_y(z)$ and $B_o(z)$. In the discussion, it is useful to denote by $G_i(z)$ the average social welfare weight on individuals of age i with earnings above z , so that we may write $B_i(z) = 1 - G_i(z)$. When considering the effect of age on the average social welfare weight $G_i(z)$, notice first that the social welfare weight on any given individual is a function of her lifetime utility which depends on her innate ability, but not on her age. However, the average social welfare weight over the earnings segment (z, ∞) is not independent of age, because this earnings segment is associated with different ability segments for the young and the old due to career effects. Since earnings profiles are increasing over the life cycle, the pool of old workers with earnings above z consists of all those whose earnings were above z as young *and also* some individuals whose earnings were below z as young. Given that earnings are increasing in ability n conditional on age (the condition for implementability of the direct mechanism), those below z as young must be of lower ability than those above z as young. Denoting the average welfare weight among workers who are below z as young but above z as old by $G_y(z^-)$, it follows that $G_y(z^-) > G_y(z)$ under concave social preferences. The average social welfare weight on old workers above z can then be written as $G_o(z) = s \cdot G_y(z) + (1 - s) \cdot G_y(z^-) > G_y(z)$ for $s \in (0, 1)$. Intuitively, with increasing earnings profiles over the career path at each ability n , older workers in a given earnings range are, on average, of lower ability than young workers in the same earnings range (age and ability are negatively correlated, conditional upon earnings), and therefore the average social welfare weight on the old is larger than on the young. This effect implies $B_o(z) < B_y(z)$, and so the welfare weight effect, like the elasticity effect discussed above, calls for age-dependent taxes with lower taxes on the old than on the young, conditional on earnings.

Finally, the *hazard ratio effect* of careers operates through the terms $C_y(z)$ and $C_o(z)$. These hazard ratios can be seen as measures of the thickness of the earnings distribution above a cutoff z for the young and the old, respectively. As an example, if earnings are distributed according to the Pareto distribution, these ratios are equal to the inverse of the Pareto parameter and measure the thickness of the upper tail. In our model, the presence of increasing earnings profiles over the career create an earnings distribution for older workers with a thicker upper tail than for younger workers, which implies $C_o(z) > C_y(z)$ at least

for a high enough z . This prediction is borne out by the data (hazard ratios under actual tax systems) and by our numerical simulations below (hazard ratios under the optimal tax system), in which the hazard ratio is larger for older than for younger workers, except at very low levels of earnings. On its own, this effect calls for higher taxes on the old than on the young, conditional on earnings, and therefore works to offset the elasticity and welfare weight effects described above. This hazard ratio effect is what drives the strong age-dependence results in [Weinzierl \(2011\)](#). In our framework, it is not possible to establish analytically whether the hazard ratio effect (calling for higher taxes on the old) is able to dominate the elasticity and welfare weight effects (calling for lower taxes on the old), and so we turn to numerical simulations based on U.S. micro data to explore this in [section 5](#).

3.3 Optimal Age-Dependent Top Tax Rates

Assuming that the upper tails of the earnings distributions for the young and the old are both Pareto distributed (with potentially different Pareto parameters), the optimal top marginal tax rates depend on career effects in a particularly simple way. We state the following proposition

Proposition 2. *Suppose that for very high incomes, the earnings of the young and the old are distributed according to Pareto distributions with Pareto parameters a_y and a_o respectively. Suppose further that the welfare weights on the young and the old converge to \bar{g}_y and \bar{g}_o and that the elasticities E_{ij} , $i, j \in \{y, o\}$ converge to constant values denoted by \bar{E}_{ij} . Then the optimal top marginal tax rates $\bar{\tau}_i$ on the young ($i = y$) and the old ($i = o$) are given by*

$$\frac{\bar{\tau}_i}{1 - \bar{\tau}_i} = \frac{1 - \bar{g}_i}{a_i \left[\bar{E}_{ii} + \bar{E}_{ji} \frac{a_j/(a_j-1)}{a_i/(a_i-1)} \frac{\bar{\tau}_j}{\bar{\tau}_i} \right]} \quad (6)$$

where $i, j \in \{y, o\}$, $i \neq j$.

Proof. To prove the proposition, we show that each of the components of [equation \(5\)](#) converges to a constant. $B_i(z)$ and $C_i(z)$ are straightforward. Clearly, if the welfare weights converge to \bar{g}_i then $B_i(z) \rightarrow 1 - \bar{g}_i$. It is a property of the Pareto distribution that $[1 - H_i(z)] / [zh_i(z)] = 1/a_i$ so $C_i(z) \rightarrow 1/a_i$. To establish the limiting value of $A_i(z)$ we use the property of the Pareto distribution with Pareto parameter a_i that $\mathbf{E}[z|z > x] = \frac{a_i}{a_i-1}x$. For individuals in age group i the limiting value of the ratio of their earnings when in the other age group j to their current earnings is $\lim_{z \rightarrow \infty} \frac{z_j(z)}{z} = \lim_{x \rightarrow \infty} \mathbf{E}[z_j|z_j > x] / \mathbf{E}[z_i|z_i > x] = \frac{a_j/(a_j-1)}{a_i/(a_i-1)}$. Combined with the assumption that the elasticities E_{ij} converge to constant values

this implies that $A_i(z) \rightarrow \bar{A}_i \equiv \left[\bar{E}_{ii} + \bar{E}_{ji} \frac{a_j/(a_j-1) \bar{\tau}_j}{a_i/(a_i-1) \bar{\tau}_i} \right]^{-1}$. Combining these pieces establishes the result in equation (6). \square

Equation (6) highlights the three conceptual effects discussed in section 3.2 in a very simple way. The *welfare weight effect* is captured by the term $1 - \bar{g}_i$ (where we have $\bar{g}_o > \bar{g}_y$ since increasing career-earnings profiles imply that, conditional on earnings, the old have lower abilities than the young), the *hazard ratio effect* is captured by the inverse of the Pareto parameter $1/a_i$ (where we have $a_y > a_o$ since increasing career-earnings profiles create a thicker upper tail in the earnings distribution of the old than in the earnings distribution of the young), and finally the *elasticity effect* is captured by the bracketed term in the denominator (where career effects imply $\bar{E}_{oo} > \bar{E}_{yy}$ and $\bar{E}_{yo} > \bar{E}_{oy}$, favoring lower taxes on the old). Note also that, in the limit where $z \rightarrow \infty$, the welfare weights on both age groups will asymptote to zero under standard concave social welfare functions, and so the welfare weight effect would not support any age-dependence at the limit. Therefore, at very high levels of earnings, optimal age dependence reflects a simple trade-off between the relative Pareto parameters—the key mechanism in previous work arguing for higher marginal rates on the old (Kremer, 2001; Weinzierl, 2011)—and career incentive effects which tend to call for lower marginal tax rates on the old as discussed above.

3.4 Age-Independent Taxes

As current tax systems in the world tend to make limited or no use of explicit age-dependence, it is of obvious interest to consider whether the career effects we introduce have any bite in influencing optimal age-independent tax schedules. This section therefore characterizes the optimal age-independent, nonlinear income tax schedule $T(z)$ with corresponding marginal tax rate schedule $\tau(z)$. We will see that it is still possible to express the optimal tax formula in terms of observable quantities and elasticities, and that the key effects discussed above are still present and affect the level and profile of marginal tax rates. In this setting, we have

Proposition 3. *The optimal age-independent tax schedule $T(z)$ is associated with marginal tax rates*

$$\frac{\tau(z)}{1 - \tau(z)} = \frac{\alpha(z) B_y(z) C_y(z) + [1 - \alpha(z)] B_o(z) C_o(z)}{\alpha(z) A_y(z)^{-1} + [1 - \alpha(z)] A_o(z)^{-1}} \quad (7)$$

where $\alpha(z) \equiv h_y(z) / [h_y(z) + h_o(z)]$ is the proportion of individuals with income z who are young and $A_y(z)$, $A_o(z)$, $B_y(z)$, $B_o(z)$, $C_y(z)$ and $C_o(z)$ are as defined in Proposition 1.

Proof: Again, we prove the result directly using the perturbation method, leaving the Hamiltonian method and the demonstration of their equivalence for appendix B. The perturbation that we consider is similar to the one depicted in Figure 1, except that it pertains to the unique tax schedule faced by both the young and the old. Hence, the marginal tax rate on both the young and the old is increased by a small amount $d\tau$ in a small earnings band $(z, z + dz)$. We now characterize the social welfare effects of this tax reform.

Mechanical Welfare Effect All taxpayers with earnings above z face a mechanical increase in tax liability of $d\tau dz$. For a young individual with earnings $z' > z$ the social value of this is given by $d\tau dz \cdot [1 - g_y(z')]$, while for an old individual at $z' > z$ the social value of this equals $d\tau dz \cdot [1 - g_o(z')]$. The total mechanical welfare effect is therefore given by

$$\Delta^M = d\tau dz \cdot \left\{ \int_z^\infty [1 - g_y(z')] dH_y(z') + \int_z^\infty [1 - g_o(z')] dH_o(z') \right\}$$

Contemporaneous Earnings Effects In the band $(z, z + dz)$, each young person reduces earnings by $-E_{yy} \cdot \frac{d\tau}{1-\tau(z)} \cdot z$ while each old person reduces earnings by $-E_{oo} \cdot \frac{d\tau}{1-\tau(z)} \cdot z$. The total tax revenue implications of these earnings response equal

$$\Delta^E = -d\tau dz \cdot z \cdot \{h_y(z) \cdot E_{yy} + h_o(z) \cdot E_{oo}\} \frac{\tau(z)}{1 - \tau(z)}$$

Aspiration Effect The higher tax rate in the earnings band $(z, z + dz)$ induces young workers who anticipate being in this band when old to invest less in future wage increases. In particular, each young person in the band $(z_y(z), z_y(z + dz))$ reduces earnings by $-E_{yo} \cdot \frac{d\tau}{1-\tau(z)} \cdot z_y(z)$. Since the total number of young workers responding through this channel is given by $h_y(z_y(z)) \frac{dz_y}{dz} dz = h_o(z) dz$, the total tax revenue implications of the aspiration effect can be written as

$$\Delta^{AS} = -d\tau dz \cdot z_y(z) h_o(z) \cdot E_{yo} \cdot \frac{\tau(z_y(z))}{1 - \tau(z)}$$

Accumulation Effect The labor supply response of young workers in the band $(z, z + dz)$ affects the wage rate and earnings of those workers when they become old. This effect implies that each old person in the band $(z_o(z), z_o(z + dz))$ reduces earnings by $-E_{oy} \cdot \frac{d\tau}{1-\tau(z)} \cdot z_o(z)$. The number of old workers affected $h_o(z_o(z)) \frac{dz_o}{dz} dz = h_y(z) dz$, and so the total accumulation

effect on tax revenue is given by

$$\Delta^{AC} = -d\tau dz \cdot z_o(z) h_y(z) \cdot E_{oy} \cdot \frac{\tau(z_o(z))}{1 - \tau(z)}$$

Optimality At the optimal tax schedule, the sum of the different social welfare effects derived above must be zero:

$$\Delta^M + \Delta^E + \Delta^{AS} + \Delta^{AC} = 0$$

By inserting the above effects in this optimality and noting that by the definition of $\alpha(z)$, $h_y(z) = \frac{\alpha(z)}{1-\alpha(z)} h_o(z)$, we obtain the result in Proposition 3. \square

The optimal age-independent tax schedule in Proposition 3 depends on weighted averages of the terms that were also present in the age-dependent tax schedules for the young and the old. Both the numerator and the denominator of equation (7) are averages of their counterparts for the age-dependent case in Proposition 1, where the weight on the young is given by the proportion of individuals at that earnings level who are young, $\alpha(z)$. Hence, the same basic effects that we discussed earlier in section 3.2 are still at play in the determination of age-independent taxes.

Because individuals have increasing earnings profiles over the life cycle, higher income levels will be populated to a larger degree by older workers than by younger workers, and vice versa at lower income levels, implying that $\alpha(z)$ is decreasing in z . This implies that at the bottom the optimal age-independent tax rate $\tau(z)$ puts a relatively high weight on the young and is therefore closer to the age-dependent tax rate on the young $\tau_y(z)$, whereas at the top the optimal age-independent tax rate $\tau(z)$ puts a relatively high weight on the old and is thus closer to the age-dependent tax rate on the old $\tau_o(z)$. This in turn implies that the earlier conclusions regarding optimal age-dependence (i.e., the difference between $\tau_y(z)$ and $\tau_o(z)$ at each earnings level) in the age-independent case manifest themselves as an effect on the profile of the marginal tax rate with respect to earnings (progressivity of $\tau(z)$ with respect to earnings). If there is a welfare argument for age-dependence favoring the old ($\tau_o(z) < \tau_y(z)$), this would in itself lower marginal tax rate progressivity in the age-independent schedule as higher earnings levels put more weight on $\tau_o(z)$.

4 How Big Are Career Effects?

Having established how the optimal way for governments to tax income depends on the career effects of work effort, the natural next question is how large these career effects actually are

in practice. In sections 2 and 3 above we have shown that the key sufficient statistics for optimal income taxation are the long-run earnings elasticities (including career effects) of the two age groups to the tax rate at each age. As argued earlier, this is not what is identified by the micro literature on labor supply and taxable income responses, which mostly studies *short-run* earnings responses to *contemporaneous* tax rates. In practice, this literature comes closer to estimating the static elasticity e in our framework than the dynamic career-inclusive elasticities E_{ij} elasticities (see [Piketty & Saez, 2013](#) for a similar argument).

Nevertheless, as equation (4) shows, the E_{ij} elasticities are functions of the underlying static elasticity e and the elasticity of future wage rates with respect to current earnings δ . While the voluminous literature on labor supply and taxable income responses can serve as a guide to what a reasonable value for the static elasticity e is, there is no such ready guidance when it comes to a reasonable value of the career elasticity δ . A careful estimation of this parameter is beyond the scope of this paper, but there is a very large literature on experience-earnings profiles in labor economics from which we can learn something about the likely size of δ . We therefore conduct a meta-analysis of this literature, focusing on 17 empirical papers studying the effects of experience, tenure and seniority on wages whose estimates permit the derivation of an estimate of δ .

In order to derive this estimate, we must perform a simple transformation of the reported estimates as most of these papers model log wages as polynomials in experience along the lines of

$$\ln(w) = \alpha + \beta_1 EXP + \beta_2 EXP^2 + \varepsilon$$

whereas we want to estimate an elasticity $\delta = \partial \ln(w) / \partial \ln(EXP)$. To derive an estimate of δ we note that by the chain rule $\frac{\partial \ln(w)}{\partial \ln(x)} = \frac{\partial \ln(w)}{\partial x} \frac{\partial x}{\partial \ln(x)}$ and by the inverse rule of calculus $\frac{\partial x}{\partial \ln(x)} = \left[\frac{\partial \ln(x)}{\partial x} \right]^{-1} = x$ and so we can derive an estimate of δ as

$$\hat{\delta} = \left[\hat{\beta}_1 + 2\hat{\beta}_2 E\bar{X}P \right] E\bar{X}P \tag{8}$$

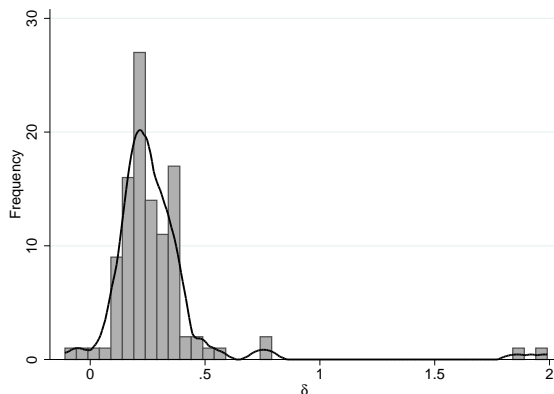
where $E\bar{X}P$ is the sample mean of EXP , and we can obtain standard errors by the delta method wherever the papers provide the necessary variances. We can also extend this to higher-order polynomials where the appropriate sample means are available.

Many papers use multiple measures of experience, for example total labor market experience EXP and tenure in the individual's current job TEN as in equation (9).

$$\ln(w) = \alpha + \beta_1 EXP + \beta_2 EXP^2 + \gamma_1 TEN + \gamma_2 TEN^2 + \varepsilon \tag{9}$$

In this case, a similar derivation to that above suggests that we should use variants of

Figure 2: Derived estimates of δ



Percentiles					mean	s.d.
10	25	50	75	90		
0.12	0.19	0.24	0.34	0.38	0.29	0.26

Notes: The 108 estimates of δ whose distribution is shown are derived using variants of equation (8) where appropriate (the vast majority of cases) and the regression of predicted wage levels on log experience levels as outlined in the text in the remaining cases. The black line is a kernel density estimate, and summary statistics of the distribution are displayed in the table below the figure.

$\hat{\delta} = [\hat{\beta}_1 + 2\hat{\beta}_2 E\bar{X}P] E\bar{X}P + [\hat{\gamma}_1 + 2\hat{\gamma}_2 T\bar{E}N] T\bar{E}N$ as our estimate. Finally, some papers, particularly those using more structural methods, present tables of predicted wages at various levels of experience rather than polynomials in experience. For these, we combine the estimated wage levels by simply regressing the predicted log wage on the log of experience, and again obtaining standard errors by the delta method where possible.

Applying these methods we are able to derive 108 estimates of δ . A full table of the estimates along with references to the exact locations in the papers and the methods used by the authors is in the online appendix, but Table 1 summarizes our findings. For each of the 17 papers, Table 1 presents the dataset(s) used, the population(s) studied, and the method(s) employed, as well as the average derived δ and its standard error, where the average δ is weighted by the number of observations used to estimate each δ in the paper.

Table 1 shows that while the estimates vary slightly from paper to paper, they mostly agree that δ lies roughly between 0.15 and 0.4 implying that a 10% increase in experience is associated with an increase in wages of between 1.5% and 4%. To reinforce this point, Figure 2 shows the distribution of all 108 estimates of δ with an overlaid kernel density alongside some summary statistics of the distribution which again show that 80% of the estimates lie between 0.19 and 0.38 with a mean of 0.29.

Table 1: Implied $\hat{\delta}$ From Existing Estimates

Paper	Dataset(s) and period	Population	Method(s)	$\hat{\delta}$	s.e.
Borjas (1981)	1966 National Longitudinal Survey of Mature Men	Men aged 45-59	OLS	0.15	(N/A)
Abraham & Farber (1987)	PSID (SRC Subsample) 1968–1981	Non-union, male household heads aged 18–60	OLS / IV	0.24	(0.035)
Altonji & Shakotko (1987)	PSID (SRC Subsample) 1968–1981	White, male household heads aged 18–60	IV - GLS	0.35	(0.073)
Topel (1991)	PSID 1968–1983	White, male household heads aged 18–60	2-step bounding exercise	1.92	(N/A)
Filer (1993)	National Longitudinal Sample 1966-1984 & 1980 Census	Women aged 14–62 in NLS; Random Sample of Women from Census	OLS with predicted experience by sector	0.23	(0.009)
Light & Ureta (1995)	National Longitudinal Survey 1966–1984	Men born 1942–1952 during 1966–1981; Women born 1944–1954 during 1968–1984	IV-GLS potential and actual experience	0.23	(0.016)
Neal (1995)	1984–1990 Displaced Worker Surveys	Full-time, nonagricultural workers whose job was lost due to establishment closing	OLS with selection correction for job loss	0.29	(0.049)
Blau & Kahn (1997)	PSID 1980 & 1989	Full-time, nonagricultural employees aged 18-65	OLS	0.32	(0.039)
Bratsberg & Terrell (1998)	NLSY 1979–1991	Male, high school graduates not employed in agriculture or military/government	OLS; Altonji and Shakotko (1987); Topel (1991)	0.20	(0.019)

Implied $\hat{\delta}$ From Existing Estimates (cont.)

Paper	Dataset(s) and period	Population	Method(s)	$\hat{\delta}$	s.e.
Flabbi & Ichino (2001)	HR data from Italian bank 1992–1995	Male workers	OLS	0.16	(0.007)
Dohmen (2004)	HR data from Dutch aircraft manufacturer 1987–1996	Permanent workers over 23	OLS	0.23	(0.004)
Schönberg (2007)	German Social Security Data 1975–1994; NLSY 1979–1994	Men in non-military employment	OLS	0.17	(0.004)
Altonji <i>et al.</i> (2009)	PSID (SRC and SEO Subsamples) 1978–1996	Male household heads aged 18–62	Structural	0.16	(N/A)
Grogger (2009)	Florida Family Transition Program 1994–1995	Work-contingent financial aid recipients	Selection corrected OLS	0.26	(0.101)
Regan & Oaxaca (2009)	NLSY 1979–1990; PSID 1968–1990; Integrated Public Use Microdata Sample 1990	Individuals aged 25–33 in 1990 (NLSY/IPUMS); Household heads aged 18–55 in 1990 (PSID/IPUMS)	OLS with predicted Experience	0.36	(0.051)
Yamaguchi (2009)	NLSY 1979–2004	White, male, high school and college graduates	Structural	0.26	(N/A)
Buchinsky <i>et al.</i> (2010)	PSID 1975–1992	Household heads aged 18–65 appearing ≥ 3 times	Structural	0.21	(0.017)

Notes: The $\hat{\delta}$ column shows the average of the estimates of δ derived from the estimates in the paper according to variants of equation (8) where appropriate (the vast majority of cases) and the regression of predicted wage levels on log experience levels as described in the text in the remaining cases. The estimates are weighted by the number of observations used to estimate them and where possible, the standard error of the estimate is computed using the delta method.

5 Numerical Simulations

5.1 Methodology

Our simulation method extends the procedure developed by Saez (2001) to a setting with dynamic wage rate effects. To perform numerical simulations, we first have to calibrate the three primitives of the model: the distribution of innate ability $F(n)$, the function relating the wage rate when old to innate ability and effort when young $\omega(n, z_y/n)$, and the static earnings elasticity parameter e . In the existing literature where the wage rate is exogenously given by innate ability, it is sufficient to use the first-order condition for earnings and an assumption about the earnings elasticity e to infer the ability level of an individual from the observed earnings and marginal tax rate of the individual. In our setting where the wage rate when old is endogenous to effort when young, the ability distribution cannot be determined quite so straightforwardly. Below we describe how $F(n)$ and $\omega(n, z_y/n)$ are calibrated in a manner that maintains the spirit of the method in previous work.

The calibration starts from micro data containing information about earnings, marginal tax rates and age in the United States. We obtain data on earnings and age from the 2007 round of the Panel Study of Income Dynamics (PSID), which we combine with the NBER TAXSIM model to get data on marginal tax rates.¹⁰ To operationalize the simplification to two age groups in our model, we split the sample into the young and the old using the median age in the sample (equal to 41 years) as a cutoff. We estimate smooth earnings distributions of the young and the old from the PSID data using an adaptive kernel density estimator. Since the data are sparse for high earners and affected by top-coding of income, we follow the standard approach in the literature and fit a Pareto distribution to the upper tail of the earnings distribution. In particular, we assume that earnings are Pareto distributed above an annual income level of \$150,000 for both the young and the old. To estimate the Pareto distribution's shape parameter a , we note that a Pareto distribution implies $z_m/z = a/(a-1)$ where $z_m \equiv \mathbf{E}[z_i | z_i > z]$ is defined as average earnings among those with earnings above z . Hence, the Pareto parameter a is estimated by regressing z^m/z for z between \$100,000 and \$150,000 on a constant, and take the estimated intercept as our estimate of $a/(a-1)$.

Having obtained the empirical distributions of earnings and marginal tax rates of the young and the old, the calibration method proceeds in the following steps. First, using the first-order condition for earnings as old (3) and an assumption about the value of e , we can back out the wage rate ω for each old person based on information about z_o and $\tau(z_o)$.

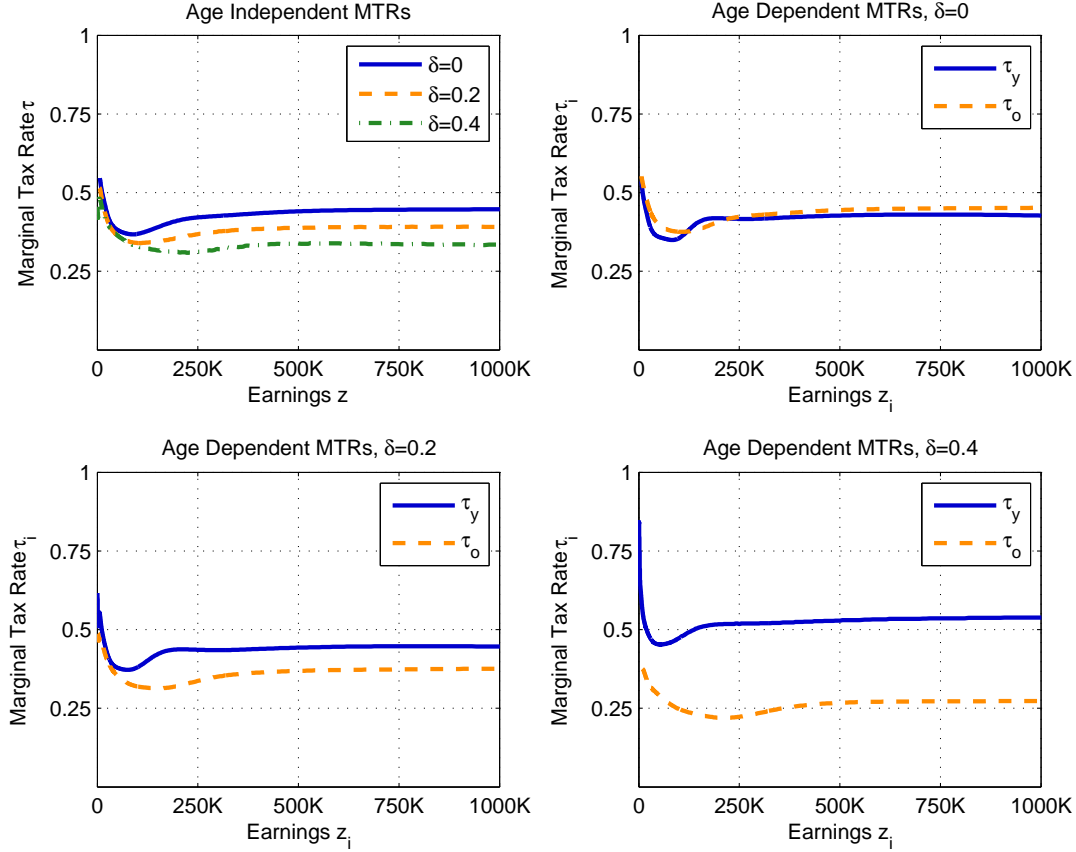
¹⁰Specifically, we use the `taxsim9` module for stata available at <http://www.nber.org/~taxsim/taxsim-calc9/>

This gives us a wage rate distribution for the old, $J(\omega)$, associated with the current tax system and earnings choices. Second, we create life-cycle earnings profiles by pairing earnings observations for the old z_o with earnings observations for the young z_y in the cross-sectional data that we use. This pairing is done using a no rank-reversal assumption that corresponds to the theoretical model: it is assumed that each individual's rank in the earnings distribution when old is the same as in the earnings distribution when young, i.e. $H_o(z_o(n)) = H_y(z_y(n))$. Hence, each earnings observation for the old is linked to an earnings observation for the young according to $z_y = H_y^{-1}[H_o(z_o)]$. Notice that the optimal tax problem considered above already makes such a no rank-reversal assumption by requiring that $z_y(n), z_o(n)$ are monotonically increasing in n to guarantee implementability. What we do here is to extend the assumption to the *current* (potentially non-optimal) tax system.

Third, having obtained the variables $(z_y, z_o, \tau_y, \tau_o, \omega)$ for each individual in the sample, it is now possible to use the first-order condition for earnings as young (2) along with assumptions about the values of e and δ to infer innate ability n for every individual. For simplicity, we assume that the career elasticity δ is constant across individuals of different abilities, i.e. we assume that the wage rate when old $\omega(n, z_y/n)$ is iso-elastic with respect to effort when young z_y/n , and show simulation results for three different scenarios: a benchmark scenario with $\delta = 0$ and scenarios with career elasticities $\delta = 0.2$ and $\delta = 0.4$ in order to span the realistic range established in the meta-analysis above. Fourth, we specify that the wage rate when old is $\omega = \omega(n, z_y/n) = \omega_0(n) \cdot \left(\frac{z_y}{n}\right)^\delta$ where $\omega_0(n)$ is the baseline wage for an old person with innate ability n in the absence of any career investment effects. As the preceding steps have established information on n, z_y, ω for each individual and we make an assumption about δ , we can back out a baseline wage $\omega_0(n)$ ensuring that the function $\omega(\cdot)$ is satisfied for every individual. This concludes the calibration as we now have information about all the primitives of the model.

Finally, in order to simulate optimal tax rates, we must specify the social welfare criterion and the aggregate tax revenue requirement R . We follow the literature and adopt a CRRA social welfare function $\Psi[U] = U^{1-\gamma}/(1-\gamma)$, where $\gamma \geq 0$ measures preferences for equity. We consider a case with “moderate” equity preferences ($\gamma = 1$) and a case with “strong” equity preferences ($\gamma = 10$). The revenue requirement R is set equal to 10,000. The aggregate income varies from one simulation to the other as income levels are endogenous to the tax schedule, but this revenue requirement corresponds to between 8% and 11% of aggregate income. We always check that the optimal tax schedule leads to $z_y(n), z_o(n)$ that are everywhere increasing in n as this is a necessary and sufficient condition for the path of $z_y(n), z_o(n)$ to be implementable via a truthful mechanism (as described earlier).

Figure 3: Optimal Marginal Tax Rates, $e = 0.5$, $\gamma = 1$

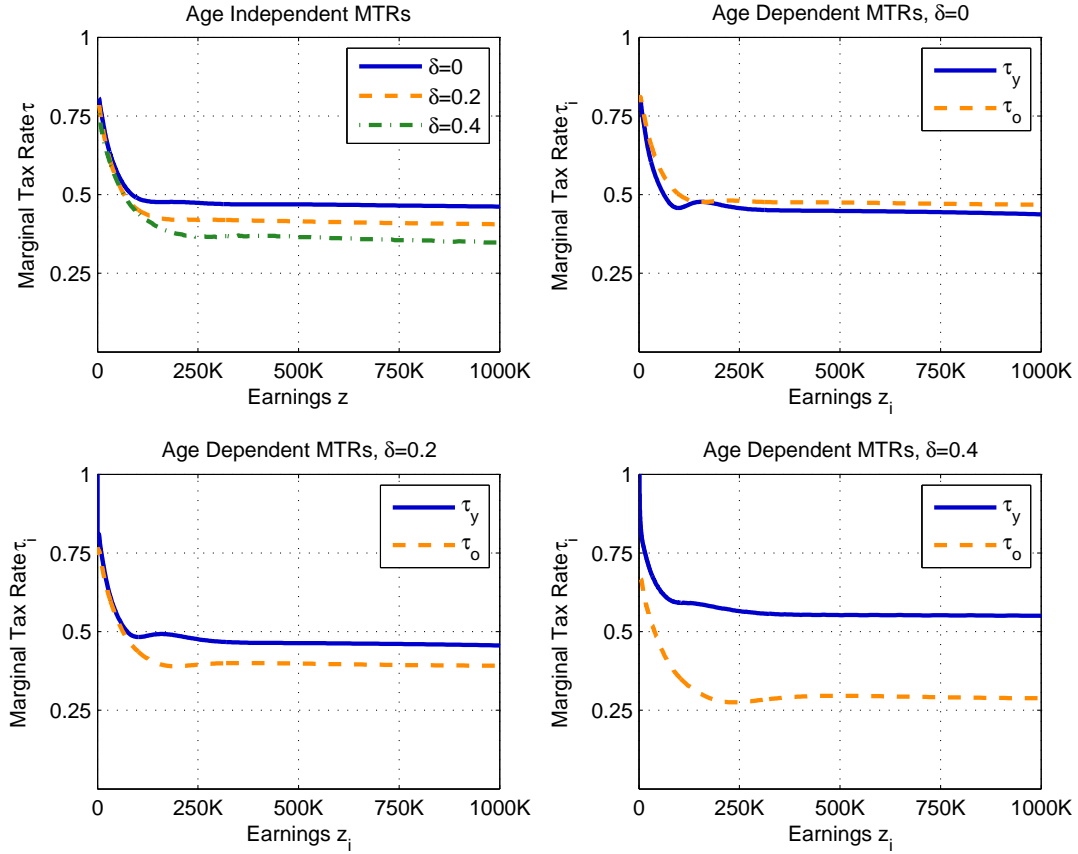


5.2 Results

Figures 2 and 3 show simulation results for age-independent and age-dependent tax schedules under various plausible levels of the parameters of the model. We assume that the elasticity parameter in the utility function is given by $e = 0.5$ throughout (corresponding to the static earnings elasticity without career effects), and consider three different values for the behavioral career elasticity $\delta \in \{0, 0.2, 0.4\}$ as well as two values of inequality aversion $\gamma \in \{1, 10\}$.

In Figure 2 where $\gamma = 1$, simulations of the age-independent tax schedule in the top-left panel show that when behavioral career effects of work effort are stronger, marginal tax rates are reduced everywhere. The age-independent marginal tax rate asymptotes to about 44% when $\delta = 0$, 39% when $\delta = 0.2$, and 34% when $\delta = 0.4$. We can also see that the U-shape of the optimal tax schedules becomes less pronounced as career effects become stronger, demonstrating our earlier conclusion that age-independent tax schedules exhibit less progressivity when accounting for endogenous career effects than in standard models

Figure 4: Optimal Marginal Tax Rates, $e = 0.5$, $\gamma = 10$



without such effects.¹¹

For the age-dependent case, when there are no behavioral career effects of work effort (so that the wage as old is determined mechanically from innate ability according to $\omega = \omega_0(n)$), the optimal tax system is associated with weak age dependence favoring the young—the marginal tax rate asymptotes to about 45% for the old, but only 43% for the young. However, once we start introducing behavioral career effects through a positive δ , optimal age dependence quickly shifts in favor of the old. At a modest behavioral career effect of $\delta = 0.2$, marginal tax rates asymptote to around 45% for the young and 37% for the old. For a stronger behavioral career effect of $\delta = 0.4$, age dependence in favor of the old becomes extremely strong with the young asymptoting to a tax rate around 53% and the old asymptoting to a tax rate around 27%.

In Figure 3 where inequality aversion is stronger at $\gamma = 10$, marginal tax rates are everywhere higher with stronger effects at the bottom than at the top of the distribution due

¹¹Moreover, notice that due to the same effect, age-independent tax schedules also exhibit less progressivity than age-dependent tax schedules for a given strength of the career effect δ .

to the fact that the social welfare weight $g(n)$ converges to zero as n becomes very large under any γ . Nevertheless, the qualitative conclusions regarding the implications of career effects for the level of marginal tax rates and age dependence remain the same. These simulations thus demonstrate that for plausible parameter values, the effects our analytical results highlight are quantitatively important with significant implications for both age-independent and age-dependent tax schedules. The stronger are behavioral career effects, the lower should be age-independent marginal tax rates and the greater should be the degree of age dependence in favor of the old.

6 Conclusion

The fact that wage paths are endogenous to hours worked and therefore to tax rates has been neglected in the optimal income tax literature. In this paper, we have presented analytical results on optimal income taxation when future wages depend on current hours worked and explored the quantitative importance of such career effects through simulations using US data. In the case of age-independent annual tax schedules, career effects of work effort reduce the level of marginal tax rates at all earnings levels and make marginal tax rate schedules flatter. In the case of age-dependent annual tax schedules, career effects of work effort reduce the level of marginal tax rates on the old and increase the level of marginal tax rates on the young, conditional on earnings, thereby providing an argument for age-dependence favoring the old. Our simulations for the US suggest that reasonably modest career effects of work effort may call for substantially lower taxes on the old. We interpret these results as being driven by two effects: an elasticity effect coming from how career effects change earnings elasticities with respect to taxes at different points in the life cycle, and an equity effect coming from how career effects create a negative correlation between age and innate ability conditional on earnings. These two effects dominate an offsetting effect coming from the fact that the hazard ratio of the earnings distribution is larger among the old than among the young. These findings are opposite to the recent literature on age-dependent taxation based on the standard framework with exogenous wage paths ([Kremer, 2001](#); [Weinzierl, 2011](#); [Golosov *et al.*, 2011](#); [Farhi & Werning, 2012](#)).

Our framework is highly stylized in order to highlight the implications of career effects as starkly as possible. In future work it would be interesting to relax the assumption of quasilinear utility so as to introduce savings into the model, which may interact with the optimal income tax schedule and age dependence in important ways. Also, while our analytical framework did not assume that the strength of behavioral career effects was constant

throughout the ability distribution, our numerical simulations for the US were based on this simplifying assumption (i.e. we assumed that δ was constant). However, some empirical evidence suggests that experience effects on future wages are larger at the top than at the bottom of the distribution (e.g. [Card & Hyslop, 2005](#)), which has potentially important implications for the optimal progressivity of marginal tax rates. Perhaps the most important call for future research emerging from this paper is the need to explore ways to credibly estimate earnings elasticities that incorporate dynamic wage rate effects to allow for proper implementation of the expressions for optimal income taxes derived here.

A Proofs of Lemmas 1 - 3

The derivation of the earnings elasticities is a straightforward application of the implicit function theorem to the system of two first-order conditions.

$$f(z, 1 - \tau) = \begin{pmatrix} f_y \\ f_o \end{pmatrix} = \begin{pmatrix} 1 - \tau_y (z_y) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} \\ 1 - \tau_o (z_o) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \frac{1}{\omega} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (10)$$

The implicit function theorem states that

$$\begin{aligned} D_{1-\tau} z &= -[D_z f(z, 1 - \tau)]^{-1} D_{1-\tau} f(z, 1 - \tau) \\ &= \begin{pmatrix} E_{yy} \frac{z_y}{1-\tau_y} & E_{yo} \frac{z_y}{1-\tau_o} \\ E_{oy} \frac{z_o}{1-\tau_y} & E_{oo} \frac{z_o}{1-\tau_o} \end{pmatrix} \end{aligned}$$

where $E_{ty} \equiv \frac{\partial z_t}{\partial 1-\tau_y} \frac{1-\tau_y}{z_t}$ and $E_{to} \equiv \frac{\partial z_t}{\partial 1-\tau_o} \frac{1-\tau_o}{z_t}$ are the earnings elasticities. Tedious algebra shows that

$$D_{1-\tau} z = \frac{e^2 z_y z_o}{(1-\tau_y)(1-\tau_o)\kappa} \begin{pmatrix} \frac{1}{e z_o} (1-\tau_o) & \frac{1+e}{e} \frac{\delta}{z_y} (1-\tau_o) \\ \frac{1+e}{e} \frac{\delta}{z_y} (1-\tau_o) & \frac{1-\tau_y}{e z_y} \left[1 + \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)} \delta (1+e)(1+\delta)\right] \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where $\kappa = 1 + \delta(1 - e\delta)(1 + e) \frac{z_o[1-T'_o]}{z_y[1-T'_y]}$, and hence that

$$\begin{pmatrix} E_{yy} & E_{yo} \\ E_{oy} & E_{oo} \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} e & e(1+e) \delta \frac{z_o[1-T'_o]}{z_y[1-T'_y]} \\ e(1+e) \delta & e \left[1 + \delta(1+\delta)(1+e) \frac{z_o[1-T'_o]}{z_y[1-T'_y]}\right] \end{pmatrix}$$

From here, proving Lemmas 1 and 3 is straightforward. Proving Lemma 2 also requires differentiation of $\frac{z_o[1-T'_o]}{z_y[1-T'_y]}$. For this, we repeat the procedure above applying the implicit function again to the first order conditions to get that

$$\begin{aligned} D_\delta z &= \frac{e^2 z_y z_o}{(1-\tau_y)(1-\tau_o)\kappa} \begin{pmatrix} \frac{1}{e z_o} (1-\tau_o) & \frac{1+e}{e} \frac{\delta}{z_y} (1-\tau_o) \\ \frac{1+e}{e} \frac{\delta}{z_y} (1-\tau_o) & \frac{1-\tau_y}{e z_y} \left[1 + \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)} \delta (1+e)(1+\delta)\right] \end{pmatrix} \begin{pmatrix} \frac{z_o}{z_y} (1-\tau_o) \\ 0 \end{pmatrix} \\ &= \frac{1}{\kappa} \begin{bmatrix} e z_o \frac{1-T'_o}{1-T'_y} \\ e(1+e) z_o \frac{z_o[1-T'_o]}{z_y[1-T'_y]} \delta \end{bmatrix} \end{aligned}$$

and from here proving Lemma 2 is just further tedious algebra.

B The Full Hamiltonian Method and Its Equivalence With the Direct Method

Analogously to the method of [Mirrlees \(1971\)](#) and [Diamond \(1998\)](#), we will treat lifetime utility as the state variable and the earnings levels as the control variables. Recall that lifetime utility is given by

$$U(z_y, z_o) = z_y - T_y(z_y) + z_o - T_o(z_o) - \frac{\left(\frac{z_y}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} - \frac{\left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}}$$

Which has first order conditions

$$\begin{aligned} 1 - \tau_y(z_y) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} &= 0 \\ 1 - \tau_o(z_o) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \frac{1}{\omega} &= 0 \end{aligned}$$

Differentiating utility and letting dots denote derivatives with respect to ability

$$\dot{U}(n) = [1 - T'_y] \dot{z}_y + [1 - T'_o] \dot{z}_o - \left(\frac{z_y}{n}\right)^{1+\frac{1}{e}} \left[\frac{\dot{z}_y}{z_y} - \frac{1}{n} \right] - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \left[\frac{\dot{z}_o}{\omega} - \frac{\dot{\omega}}{\omega} \frac{z_o}{\omega} \right]$$

where $\frac{\dot{\omega}}{\omega} = \eta \frac{1}{n} + \delta \left(\frac{\dot{z}_y}{z_y} - \frac{1}{n} \right)$ and $\eta \equiv \frac{\partial \omega}{\partial n} \frac{n}{\omega}$. Using the first order conditions, this reduces to

$$\dot{U}(n) = \frac{1}{n} \left\{ \left(\frac{z_y}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} (\eta - \delta) \right\}$$

B.1 Age-dependent Taxes

Turning first to the case of age-dependent taxes, the government's problem is to

$$\max \int_0^\infty \Psi [U(n)] f(n) dn$$

subject to

$$\begin{aligned} \int_0^\infty \{z_y(n) - T_y[z_y(n)] + z_o(n) - T_o[z_o(n)]\} f(n) dn &\leq \int_0^\infty [z_y(n) + z_o(n)] f(n) dn - R \\ 1 - \tau_y(z_y) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} &= 0 \\ 1 - \tau_o(z_o) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \frac{1}{\omega} &= 0 \end{aligned}$$

where we use the first-order approach to substitute the first-order conditions for the complete incentive compatibility constraints.¹² Substituting out the tax system, this becomes

¹²In general, in dynamic moral hazard problems, the first-order approach is not always justified as sophisticated possible deviations may not be captured by the first-order conditions (the so-called "double deviation" problem). However, since in our setting the individual's problem is globally concave, we do not face these

$$\max \int_0^\infty \Psi [U(n)] f(n) dn$$

subject to

$$\int_0^\infty \left\{ U(n) + \frac{\left(\frac{z_y(n)}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} + \frac{\left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} \right\} f(n) dn \leq \int_0^\infty [z_y(n) + z_o(n)] f(n) dn - R$$

$$\dot{U}(n) = \frac{1}{n} \left\{ \left(\frac{z_y}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} (\eta - \delta) \right\}$$

Forming the Hamiltonian,

$$H(n) = \left\{ \Psi [U(n)] - p \left[U(n) + \frac{\left(\frac{z_y(n)}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} + \frac{\left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} - z_y(n) - z_o(n) \right] \right\} f(n)$$

$$+ \mu(n) \frac{1}{n} \left\{ \left(\frac{z_y(n)}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}} (\eta - \delta) \right\}$$

When taxes are dependent on age, the planner has two control variables available, z_y and z_o which she can manipulate independently and so the optimality conditions are that

$$0 = \frac{\partial H}{\partial z_y} = -pf(n) \left[\frac{1}{n} \left(\frac{z_y}{n}\right)^{\frac{1}{e}} - 1 - \left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} \right] + \mu(n) \frac{1}{n} \left(1 + \frac{1}{e}\right) \left\{ \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} - \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} (\eta - \delta) \right\}$$

$$0 = \frac{\partial H}{\partial z_o} = -pf(n) \left[\left(\frac{z_o(n)}{\omega(n)}\right)^{\frac{1}{e}} \frac{1}{\omega(n)} - 1 \right] + \mu(n) \frac{1}{n} \left(1 + \frac{1}{e}\right) (\eta - \delta) \left(\frac{z_o(n)}{\omega(n)}\right)^{\frac{1}{e}} \frac{1}{\omega(n)} \quad (12)$$

$$-\dot{\mu} = \frac{\partial H}{\partial U} = [\Psi' [U(n)] - p] f(n) \quad (13)$$

First note that integrating equation (13), and using the transversality condition,

$$\mu(n) = \int_n^\infty [\Psi' [U(n)] - p] dF(n) = p \int_n^\infty [g(n) - 1] dF(n) \quad (14)$$

For the young, substituting the first order conditions into (11) we get that

$$-pf(n) T'_y + \mu \left(1 + \frac{1}{e}\right) \frac{1}{n} \left\{ [1 - T'_y] + [1 - T'_o] \frac{z_o}{z_y} \delta (1 + \delta - \eta) \right\} = 0$$

which combined with equation (14) yields

$$\frac{T'_y}{1 - T'_y} = \frac{\int_n^\infty [1 - g(n)] f(n) dn}{nf(n)} \left(1 + \frac{1}{e}\right) \left(1 + \delta \frac{z_o [1 - T'_o]}{z_y [1 - T'_y]} [1 + \delta - \eta]\right) \quad (15)$$

issues. For technical details on double deviations and the requirements for double deviations not to be an issue, see [Kocherlakota \(2004\)](#) in the context of unemployment insurance and [Ábrahám & Pavoni \(2009\)](#) in the context of income taxation with hidden saving.

Turning to the old, equation (12) together with the first order conditions imply that

$$\begin{aligned} pf(n)T'_o &= -\mu \left(1 + \frac{1}{e}\right) (\eta - \delta) \frac{[1 - T'_o]}{n} \\ \frac{T'_o}{1 - T'_o} &= \frac{\int_n^\infty [1 - g(n)] f(n) dn}{nf(n)} \left(1 + \frac{1}{e}\right) (\eta - \delta) \end{aligned} \quad (16)$$

B.2 Age-Independent Taxes

In the case of age-independent taxes, the setup is the same as above for age-dependent taxes. However, the planner faces an additional constraint, namely that taxes be age-independent, or that $T_y(z) = T_o(z)$. To operationalize this constraint, note that it implies that $T_y(z_y) = T_o(z_o)$ whenever $z_y(n) = z_o(n')$. In particular, defining $n_o(n)$ as the ability level of the individual whose earnings when old are equal to the earnings when young of an individual of ability n , it is the case that $z_y(n) = z_o[n_o(n)]$. In effect, this constraint limits the number of control variables available to the planner to one (either $z_y(n)$ or $z_o(n)$). Without loss of generality we will work with $z_y(n)$. It further implies that changes in the control variable $z_y(n)$ are also changes in the earnings when old of individuals with ability $n_o(n)$.

This means that the optimality conditions on the Hamiltonian are instead that

$$\begin{aligned} 0 &= \frac{\partial H(n)}{\partial z_y(n)} + \frac{\partial H[n_o(n)]}{\partial z_o(n)} \\ &= -pf(n) \left[\frac{1}{n} \left(\frac{z_y(n)}{n}\right)^{\frac{1}{e}} - 1 - \left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y(n)} \right] \\ &\quad - pf[n_o(n)] \left[\left(\frac{z_o[n_o(n)]}{\omega[n_o(n)]}\right)^{\frac{1}{e}} \frac{1}{\omega[n_o(n)]} - 1 \right] \\ &\quad + \mu(n) \frac{1}{n} \left(1 + \frac{1}{e}\right) \left\{ \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} - \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y(n)} (\eta - \delta) \right\} \\ &\quad + \mu[n_o(n)] \frac{1}{n_o(n)} \left(1 + \frac{1}{e}\right) (\eta - \delta) \left(\frac{z_o[n_o(n)]}{\omega[n_o(n)]}\right)^{\frac{1}{e}} \frac{1}{\omega[n_o(n)]} \end{aligned} \quad (17)$$

$$-\dot{\mu} = \frac{\partial H}{\partial U} = [\Psi' [U(n)] - p] f(n)$$

Using the first order conditions to substitute back in the tax terms, we can rewrite equation (17) as

$$\begin{aligned} 0 &= -pf(n)T'_y[z_y(n)] - pf[n_o(n)]T'_o[z_o[n_o(n)]] \\ &\quad + \mu(n) \frac{1}{n} \left(1 + \frac{1}{e}\right) \left\{ 1 - T'_y[z_y(n)] + (1 - T'_o[z_o(n)]) \frac{z_o(n)}{z_y(n)} \delta (1 + \delta - \eta) \right\} \\ &\quad + \mu[n_o(n)] \frac{1}{n_o(n)} \left(1 + \frac{1}{e}\right) (\eta - \delta) (1 - T'_o[z_o[n_o(n)]]) \end{aligned}$$

or that

$$\frac{T'[z_y(n)]}{1 - T'[z_y(n)]} = \left\{ \frac{\int_n^\infty [1 - g(n)] f(n) dn}{n} \left(1 + \delta \frac{z_o(n) (1 - T'[z_o(n)])}{z_y(n) (1 - T'[z_y(n)])} [1 + \delta - \eta] \right) + \frac{\int_{n_o(n)}^\infty [1 - g(n)] f(n) dn}{n_o(n)} (\eta - \delta) \right\} \times \left(1 + \frac{1}{e} \right) \frac{1}{f(n) + f[n_o(n)]} \quad (18)$$

B.3 Equivalence of the Hamiltonian and Direct Methods

Here we demonstrate the equivalence of the two methods for the age-dependent tax schedule for the young. The demonstration for the age-dependent tax schedule for the old and the age-independent tax schedule follow the same steps, and are left to the interested reader. The following lemma akin to Lemma 1 in [Saez \(2001\)](#) but for our setting will be useful in demonstrating this equivalence.

Lemma 4. *For any tax schedule T not necessarily optimal and not necessarily age-dependent, the earnings functions z_{yn} and z_{on} are non-decreasing and satisfy the following system of differential equations.*

$$\frac{\dot{z}_y}{z_y} = \frac{1}{n} [1 + E_{yy} + E_{yo}\eta] - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{yy} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{yo} \quad (19)$$

$$\frac{\dot{z}_o}{z_o} = \frac{1}{n} [E_{oy} + (1 + E_{oo})\eta] - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{oy} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{oo} \quad (20)$$

If equations (19) and (20) lead to $\dot{z}_y < 0$ or $\dot{z}_o < 0$ then z_{yn} (z_{on}) is discontinuous and (19) or (20) does not hold.

Proof. Starting with the young, first note that $\dot{z}_{yn}/z_{yn} = (\dot{l}_{yn}/l_{yn}) + (1/n)$. Since there are no income effects, we can write the labor supply of a young individual of ability n as a function of the wages in the two periods $l_{yn} = l_y(w_{yn}, w_{on})$ where $w_{yn} = n(1 - T_y')$ and $w_{on} = \omega_n(1 - T_o')$. This means that

$$\dot{l}_{yn} = \frac{\partial l_y}{\partial w_y} [1 - T_y' - nT_y'' \dot{z}_{yn}] + \frac{\partial l_y}{\partial w_o} [\dot{\omega}_n (1 - T_o') - \omega_n T_o'' \dot{z}_{on}]$$

Now $\dot{\omega}_n = \frac{\partial \omega_n}{\partial n} = \eta \frac{\omega_n}{n}$ where $\eta \equiv \frac{\partial \omega}{\partial n} \frac{n}{\omega}$ is the elasticity of the wage when old with respect to ability. Then, using the labor supply/earnings elasticities $E_{ij} \equiv \frac{\partial z_i}{\partial 1 - T_j'} \frac{1 - T_j'}{z_j} = \frac{\partial l_i}{\partial w_j} \frac{w_j}{l_i}$ we get that

$$\frac{\dot{l}_y}{l_y} = E_{yy} \frac{1}{n} - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{yy} + E_{yo} \frac{\eta}{n} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{yo}$$

and plugging everything in and rearranging we get that

$$\frac{\dot{z}_y}{z_y} = \frac{1}{n} [1 + E_{yy} + E_{yo}\eta] - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{yy} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{yo}$$

By exactly the same reasoning, $\dot{z}_{on}/z_{on} = (\dot{l}_{on}/l_{on}) + (\dot{\omega}_n/\omega_n)$ where $l_{on} = l_o(w_{yn}, w_{on})$ so that

$$\dot{l}_{on} = \frac{\partial l_o}{\partial w_y} [1 - T_y' - nT_y'' \dot{z}_{yn}] + \frac{\partial l_o}{\partial w_o} [\dot{\omega}_n (1 - T_o') - \omega_n T_o'' \dot{z}_{on}]$$

and

$$\frac{\dot{l}_o}{l_o} = E_{oy} \frac{1}{n} - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{oy} + E_{oo} \frac{\dot{\omega}}{\omega} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{oo}$$

so plugging in and rearranging

$$\frac{\dot{z}_o}{z_o} = \frac{1}{n} [E_{oy} + (1 + E_{oo}) \eta] - \dot{z}_y \frac{T_y''}{1 - T_y'} E_{oy} - \dot{z}_o \frac{T_o''}{1 - T_o'} E_{oo}$$

which finishes the proof. \square

In particular, lemma 4 shows that for a tax schedule linearized around the optimum, we will have

$$\dot{z}_y = \frac{z_y}{n} [1 + E_{yy} + E_{yo} \eta] \quad (21)$$

$$\dot{z}_o = \frac{z_o}{n} [E_{oy} + (1 + E_{oo}) \eta] \quad (22)$$

It is also useful to note that combining (15) and (16),

$$\frac{T_o' z_o}{T_y' z_y} = \frac{(\eta - \delta) \tilde{z}}{1 + \delta \tilde{z} (1 + \delta - \eta)} \quad (23)$$

where we define $\tilde{z} = \frac{[1 - T_o'] z_o}{[1 - T_y'] z_y}$ for convenience.

The direct approach for the young gave us

$$\frac{\tau_y(z^*)}{1 - \tau_y(z^*)} = \frac{\int_{z^*}^{\infty} (1 - g_y(z)) dH_y(z)}{z^* h_y(z^*) \left[E_{yy} + E_{oy} \frac{\tau_o[z_o(z^*)] z_o(z^*)}{\tau_y(z^*) z^*} \right]} \quad (24)$$

Using (23) in the definition of $A_y(z)$ in equation (5),

$$A_y(z)^{-1} = \frac{1 + \delta \tilde{z} (1 + \delta - \eta)}{E_{yy} [1 + \delta \tilde{z} (1 + \delta - \eta)] + E_{oy} (\eta - \delta) \tilde{z}}$$

and using the definitions of the elasticities in (4)

$$\begin{aligned} A_y(z) [1 + \delta \tilde{z} (1 + \delta - \eta)] &= E_{yy} [1 + \delta \tilde{z} (1 + \delta)] - E_{oy} \delta \tilde{z} + [E_{oy} \tilde{z} - E_{yy} \delta \tilde{z}] \eta \\ &= \frac{e}{\kappa} \left\{ \frac{e}{1 + e} + \frac{\kappa}{1 + e} + e \delta \tilde{z} \eta \right\} \\ &= \frac{e}{1 + e} [1 + E_{yy} + E_{yo} \eta] \end{aligned} \quad (25)$$

Then, combining (25) with (21) and noting that by definition $h_y(z_y) \dot{z}_y = f[n_y(z_y)]$ and that $-\mu(n)/p = \int_n^{\infty} [1 - g(n)] dF(n) = \int_{z_y(n)}^{\infty} [1 - g_y(z)] dH_y(z)$, we demonstrate the equivalence of the Hamiltonian solution (15) and the direct solution (24) for the young. Exactly analogous steps and noting that $h_o(z_o) \dot{z}_o = f[n_o(z_o)]$ and $-\mu(n)/p = \int_n^{\infty} [1 - g(n)] dF(n) = \int_{z_o(n)}^{\infty} [1 - g_o(z)] dH_o(z)$ demonstrate the equivalence of the Hamiltonian solution (16) and the direct solution for the old. Combining these two sets of results demonstrates the equivalence for the age-dependent case.

References

- ÁBRAHÁM, ÁRPÁD, & PAVONI, NICOLA. 2009. *Optimal Income Taxation and Hidden Borrowing and Lending: The First-Order Approach in Two Periods*. mimeo: University College London.
- ABRAHAM, KATHARINE G., & FARBER, HENRY S. 1987. Job Duration, Seniority, and Earnings. *American Economic Review*, **77**, 278–297.
- AKERLOF, GEORGE. 1978. The Economics of “Tagging” as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review*, **68**, 8–19.
- ALTONJI, JOSEPH G., & SHAKOTKO, ROBERT A. 1987. Do Wages Rise with Job Seniority? *Review of Economic Studies*, **54**, 437–459.
- ALTONJI, JOSEPH G., SMITH, ANTHONY, & VIDANGOS, IVAN. 2009. *Modeling Earnings Dynamics*. NBER Working Paper 14743.
- BANKS, JAMES, & DIAMOND, PETER. 2011. The base for Direct Taxation. *Chap. 6 of: MIRRLEES, JAMES, ADAM, STUART, BESLEY, TIM, BLUNDELL, RICHARD, BOND, STEPHEN, CHOTE, ROBERT, GAMMIE, MALCOLM, JOHNSON, PAUL, MYLES, GARETH, & POTERBA, JAMES (eds), Dimensions of Tax Design: The Mirrlees Review*. Oxford: Oxford University Press.
- BEN-PORATH, YORAM. 1967. The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, **75**, 352–365.
- BLAU, FRANCINE D., & KAHN, LAWRENCE M. 1997. Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics*, **15**, 1–42.
- BLUNDELL, RICHARD, & MACURDY, THOMAS. 1999. Labor Supply: A Review of Alternative Approaches. *Chap. 27, pages 1559–1695 of: ASHENFELTER, ORLEY, & CARD, DAVID (eds), Handbook of Labor Economics, Volume 3*. North Holland: Elsevier.
- BORJAS, GEORGE J. 1981. Job Mobility and Earnings over the Life Cycle. *Industrial and Labor Relations Review*, **34**, 365–376.
- BOVENBERG, ARY LANS, & JACOBS, BAS. 2005. Redistribution and Education Subsidies are Siamese Twins. *Journal of Public Economics*, **89**, 2005–2035.
- BRATSBERG, BERNT, & TERRELL, DEK. 1998. Experience, Tenure, and Wage Growth of Young Black and White Men. *Journal of Human Resources*, **33**, 658–682.
- BUCHINSKY, MOSHE, FOUGÈRE, DENIS, KRAMARZ, FRANCIS, & TCHERNIS, RUSTY. 2010. Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States. *Review of Economic Studies*, **77**, 972–1001.
- CARD, DAVID, & HYSLOP, DEAN R. 2005. Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers. *Econometrica*, **73**, 1723–1770.

- DIAMOND, PETER A. 1998. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, **88**, 83–95.
- DOHMEN, THOMAS J. 2004. Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles. *Labour Economics*, **11**, 741–763.
- EATON, JONATHAN, & ROSEN, HARVEY S. 1980. Taxation, Human Capital, and Uncertainty. *American Economic Review*, **70**, 705–715.
- FARBER, HENRY S. 1999. Mobility and Stability: The Dynamics of Job Change in Labor Markets. *Chap. 37 of: ASHENFELTER, ORLEY, & CARD, DAVID (eds), Handbook of Labor Economics, vol.3.*
- FARHI, EMMANUEL, & WERNING, IVÁN. 2012. *Insurance and Taxation over the Life Cycle*. forthcoming, Review of Economic Studies.
- FILER, RANDALL K. 1993. The Usefulness of Predicted Values for Prior Work Experience in Analyzing Labor Market Outcomes for Women. *Journal of Human Resources*, **28**, 519–537.
- FLABBI, LUCA, & ICHINO, ANDREA. 2001. Productivity, Seniority and Wages: New Evidence from Personnel Data. *Labour Economics*, **8**, 359–387.
- GOLOSOV, MIKHAIL, TSYVINSKI, ALEH, & WERNING, IVAN. 2007. New Dynamic Public Finance: A User’s Guide. *Chap. 5, pages 317–387 of: ACEMOGLU, DARON, ROGOFF, KENNETH, & WOODFORD, MICHAEL (eds), NBER Macroeconomic Annual 2006*. Cambridge: MIT Press.
- GOLOSOV, MIKHAIL, TROSHKIN, MAXIM, & TSYVINSKI, ALEH. 2011. *Optimal Dynamic Taxes*. Mimeo: Yale University.
- GROGGER, JEFFREY. 2009. Welfare Reform, Returns to Experience, and Wages: Using Reservation Wages to Account for Sample Selection Bias. *Review of Economics and Statistics*, **91**, 490–502.
- HOLMSTRÖM, BENGT. 1999. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, **66**, 169–182.
- KOCHERLAKOTA, NARAYANA. 2004. Figuring out the Impact of Hidden Savings on Optimal Unemployment Insurance. *Review of Economic Dynamics*, **7**, 541–554.
- KRAUSE, ALAN. 2009. Optimal Nonlinear Income Taxation With Learning-by-Doing. *Journal of Public Economics*, **93**, 1098–1110.
- KREMER, MICHAEL. 2001. *Should Taxes be Independent of Age?* Mimeo: Harvard University.
- LIGHT, AUDREY, & URETA, MANUELITA. 1995. Early-Career Work Experience and Gender Wage Differentials. *Journal of Labor Economics*, **13**, 121–154.

- MANNING, ALAN. 2000. Movin' On Up: Interpreting the Earnings-Experience Profile. *Bulletin of Economic Research*, **52**, 261–295.
- MIRRLEES, JAMES A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, **38**, 175–208.
- NEAL, DEREK. 1995. Industry-Specific Human Capital: Evidence from Displaced Workers. *Journal of Labor Economics*, **13**, 653–677.
- PIKETTY, THOMAS. 1997. La Redistribution Fiscale face au Chômage. *Revue Française d'Economie*, **12**, 157–201.
- PIKETTY, THOMAS, & SAEZ, EMMANUEL. 2013. *Optimal Labor Income Taxation*. Forthcoming in Handbook of Public Economics, Vol.5.
- REGAN, TRACY L., & OAXACA, RONALD L. 2009. Work Experience as a Source of Specification Error in Earnings Models: Implications for Gender Wage Decompositions. *Journal of Population Economics*, **22**, 463–499.
- SAEZ, EMMANUEL. 2001. Using Elasticities to Derive Optimal Tax Rates. *Review of Economic Studies*, **68**, 205–229.
- SCHÖNBERG, UTA. 2007. Wage Growth Due to Human Capital Accumulation and Job Search: A Comparison between the United States and Germany. *Industrial & Labor Relations Review*, **60**, 562–586.
- STIGLITZ, JOSEPH. 1982. Self-selection and pareto efficient taxation. *Journal of Public Economics*, **17**, 213–240.
- TOPEL, ROBERT. 1991. Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority. *Journal of Political Economy*, **99**, 145–176.
- WEINZIERL, MATTHEW C. 2011. The Surprising Power of Age-Dependent Taxes. *Review of Economic Studies*, **78**, 1490–1518.
- YAMAGUCHI, SHINTARO. 2009. The Effect of Match Quality and Specific Experience on Career Decisions and Wage Growth. *Labour Economics*, **17**, 407–423.