Dynamic Job Market Signaling and Optimal Taxation*

André Sztutman

August 25, 2026

Abstract

How should workers' dynamic reputation-building efforts affect the design of optimal taxes? To answer this question, this paper presents a model in which job histories transmit information about workers' productivity. Workers' reputational benefits from overworking to signal higher productivity generate dynamic "rat race" externalities. The paper then provides general optimal taxation formulas, written as the composition of a redistributive and a corrective component. The latter depends on dynamic labor wedges: the ratio of the marginal contribution to output to the sum of current and future benefits the worker receives from supplying one extra unit of labor. Whether these wedges call for taxes on lifetime income alone or for corrections that depend on age and earnings history turns on how firms weigh the timing of past deliverables when reading resumes. Calibrating the taxation formulas to the existing evidence suggests a substantial and positive corrective tax component, especially among high-income earners.

*Carnegie Mellon University – Tepper School of Business. Email: asztutma@andrew.cmu.edu.
This paper is based on Chapter 1 of my Ph.D. thesis at MIT. I am grateful for the guidance and support of Jonathan Gruber, James Poterba, Robert Townsend, and Iván Werning. I also thank Daron Acemoglu, Laurence Ales, John Becko, V.V. Chari, Jonathan Cohen, Arnaud Costinot, Ashley Craig, Peter Diamond, Dave Donaldson, Lucas Finamor, Amy Finkelstein, Nicole Immorlica, Caio Lorecchio, Dylan Moore, Thi Mai Anh Nguyen, Florian Scheuer, Liyan Shi, Ali Shourideh, Lowell Taylor, Aleh Tsyvinski, Bryant Xia, Ariel Zetlin-Jones, as well as seminar participants at the 2024 AEA Meeting, 2023 NBER SI Macro Public Finance Workshop, 2023 IIPF Annual Meeting, 2023 SEA Annual Meeting, and 2022 EWMES, and the various other venues where this paper has been presented.

Keywords: job market signaling; career concerns; adverse selection; optimal taxation; labor wedges. JEL codes: D82, H21, H23, J20.

1 Introduction

Dynamic informational imperfections shape workers' incentives to exert effort, affect the distribution of income, and thus alter the tradeoff between efficiency and redistribution that the government faces when setting taxes. Early in their careers, foreseeing higher salaries in the future, workers exert great effort even though their current salaries are small. As they accumulate experience, employers become able to recognize who the more productive workers are and start paying them differentially [Farber and Gibbons, 1996, Altonji and Pierret, 2001, Lange, 2007]. Yet employers may never distinguish workers perfectly, and workers of different productivities may end up receiving the same wages. Standard benchmark taxation models ignore these dynamics, either adopting a frictionless view of labor markets in which salaries equal marginal products, or treating informational imperfections as static. This paper fills that gap: it builds a dynamic signaling model of labor markets, derives optimal taxation formulas in terms of sufficient statistics, and calibrates the key new statistic in the formulas from existing estimates. The analysis yields three main results.

First, when labor markets learn about workers from their job histories, the distortion that taxes should correct is not the gap between current wages and marginal products. A worker who completes one more deliverable is paid today's salary and, in addition, strengthens tomorrow's resume. What guides the worker's choices is the sum of the two. The relevant object is the dynamic labor wedge χ: the ratio of the marginal contribution to output to the sum of current and future benefits the worker receives from supplying one extra unit of labor, defined at each point in the career of the worker (Section 3). Comparing current wages with productivities reveals little about this wedge. Large distortions can coexist with no wage–product gap at all, and large gaps can arise when the timing of effort is undistorted.

Second, optimal taxes can be written as the composition of a redistributive component satisfying standard Mirrleesian formulas and a corrective, Pigouvian, component constructed from the labor wedges (Section 4). The structure of the corrective component is governed by how firms read resumes. When resumes are summarized by their "length," the cumulative sum of deliverables, there are no intertemporal distortions and the corrective factor is the end-of-career labor wedge expressed as a function of lifetime income, $\chi(y)$ (Proposition 4); when, in addition, households agree on the timing of consumption and labor supply (Assumption 1(a)), lifetime income is the optimal base for the entire tax system (Proposition 5). Proposition 4 also shows that, within its regularity domain, no other class of signals eliminates intertemporal distortions in every economy. When the weights firms place on past deliverables depend on the worker's current age, corrective taxes cannot in general be confined to lifetime income; when resumes are summarized by their "pace," the benefits of exerting effort are higher at early stages of the career, and, in a parametric example, the correction takes a closed form that depends only on age, high for young workers, declining with age, and vanishing at retirement (Corollary 2); it can be implemented with an age-dependent income tax of the kind several tax codes already contain. Because the length-case labor wedges satisfy $\chi \leq 1$ (Proposition 3), accounting for dynamic signaling pushes toward higher marginal taxes at every income level in that benchmark, holding measured elasticities and income distributions fixed; and there can be pure efficiency gains from age-dependent taxes that correct intertemporal distortions.

Third, the labor wedges can be calibrated from existing estimates. The model's free-entry condition ties the wedges to the signaling component of the returns to experience and to the shape of the lifetime income distribution; combining published lifecycle earnings-growth profiles [Guvenen, Karahan, Ozkan, and Song, 2021] with the signaling share estimated in the schooling context [Aryal, Bhuller, and Lange, 2022], the implied corrective component of taxes is of the order of 5% for the average worker and 11% at the top at the central signaling share (from 6% to 15% across shares); the top figures are high-gradient scenarios, since sorting on realized lifetime earnings steepens top profiles (Section 5). Estimating the wedges directly from tax variation raises identification questions that we leave for future work. The existing evidence thus suggests that, once the role of imperfect information in labor markets is taken into account, the current tax system is less redistributive than we previously thought.

The model is a *signaling-by-doing* model (Section 2). Firms can see and pay for the execution of clearly specified tasks—*deliverables*—but cannot assess any worker's individual contribution to output. Firms pay for the execution of these deliverables based on the worker's *resume*, which parsimoniously summarizes the history of deliverables the worker has produced so far in their career; in equilibrium, each worker is paid the average productivity of the workers with equivalent resumes. Forward-looking workers choose effort knowing that today's deliverables raise future salaries by moving employers' beliefs. Resumes convert the static cross-subsidy of adverse-selection models into a dynamic one, generating time-varying wedges that make the wage–marginal-product gap the wrong statistic and create a role for age- and history-dependent corrective taxes. Two polar cases illustrate the range of possibilities this setup can accommodate. When the resume is summarized by its "length," the cumulative sum of deliverables, salaries rise with experience and exceed marginal products at the peak of each career, yet the timing of effort is undistorted: the rat race is a lifetime rat race. When the resume is summarized by its "pace," the amount of deliverables generated per unit of time, there can be separation: workers of different types supply labor at different paces, with more productive workers supplying work at higher paces. Because of separation, firms offer salaries that, in equilibrium, are equal to marginal productivities. That does not imply there are no distortions. There is another source of benefits workers receive from exerting more effort, which comes from establishing their reputations and increasing the payments they will receive for future deliverables, and these reputational benefits are largest at early stages of the career. At an optimal tax system, the Pigouvian component of taxes is therefore higher at the beginning of workers' careers, correcting the incentives to overwork then, and declines with age. The pace case is developed in Section 4.3, jointly with the taxes that make its equilibrium tractable.

The model also delivers a welfare result with a technological interpretation. New monitoring and screening technologies arguably make deliverables a better measure of product. Building on Stantcheva [2014], we show that if society values redistribution, reducing informational asymmetries in labor markets *reduces* welfare: asymmetries force high-productivity workers to separate themselves by working harder, sustaining implicit cross-subsidies that do part of the planner's work (Proposition 12). The effect of reducing informational asymmetries on optimal marginal taxes is subtler, since the corrective factor pushes toward lower marginal taxes while the redistributive factor pushes toward higher ones; in a benchmark case with moderate welfare weights at the top the corrective effect dominates, and marginal taxes increase with the degree of asymmetry.

Related literature

The taxation results build on the nonlinear income taxation literature descending from Mirrlees [1971], Diamond [1998], and Saez [2001], and contribute to the branch that replaces the frictionless labor market with richer models: educational signaling with complete [Andersson, 1996] or limited instruments [Craig, 2023], competitive screening [Stantcheva, 2014, Bastani, Blumkin, and Micheletto, 2015], monopsonistic screening and market power [Hariton and Piaser, 2007, da Costa and Maestri, 2019, Hummel, 2024], performance pay and moral hazard [Doligalski, Ndiaye, and Werquin, 2023], multidimensional skills [Rothschild and Scheuer, 2013], task assignment [Ales, Kurnaz, and Sleet, 2015], and rent seeking [Piketty, Saez, and Stantcheva, 2014, Rothschild and Scheuer, 2016]. What this literature has not analyzed is *dynamic* job market screening—the case in which the friction operates through job histories—and that is the focus here. Relative to Stantcheva [2014], the model here is the dynamic counterpart of her static economy: there, every unit of a worker's labor is paid according to the worker's final resume, as if reputations were established instantaneously, whereas here workers separate from their less productive peers only gradually, generating cross-subsidies through the life cycle that the static model cannot express (see Section 2.4). Methodologically, the closest papers are Scheuer and Werning [2016] and Scheuer and Werning [2017], who show that standard optimal taxation formulas apply in a broad class of models with endogenous wages; this paper adds further generality, covering labor market frictions that generate wedges between pay and product, provided firms break even and earnings map one-to-one to labor supply choices. In the spirit of the principle of targeting [Sandmo, 1975, Dixit, 1985, Kopczuk, 2003], the formulas separate the correction of the labor market externality from the redistribution of income. The welfare comparison generalizes Proposition 13 of Stantcheva [2014] to intermediate degrees of informational friction and to any labor market model satisfying the same invertibility conditions. The result that lifetime income remains the optimal base extends Atkinson and Stiglitz [1976] to double adverse selection, echoing tax smoothing results [Werning, 2007b] and income-averaging ideas going back to Vickrey [1947]. The age-dependent corrective taxes of Corollary 2 arise from a mechanism new to the age-dependent taxation literature [Weinzierl, 2011, Farhi and Werning, 2013]: an informational externality whose strength declines with age, rather than age-varying elasticities or productivity profiles.

On the theory of labor markets, the signaling-by-doing model combines elements of static rat-race models [Akerlof, 1976, Miyazaki, 1977] with dynamic career concerns [Holmström, 1999, Gibbons and Murphy, 1992, Bonatti and Hörner, 2017, Cisternas, 2018, Hörner and Lambert, 2021]. As in Miyazaki [1977], firms pay for observable tasks whose quality they cannot assess; as in Holmström [1999], reputations evolve with histories, but they attach to the completion of deliverables, summarized by resumes, rather than to realized output, and firms can pay for the performance measures they observe. Unlike in Holmström [1999], where reputational incentives can leave effort too weak once reputations are consolidated, here the rat race keeps incentives high-powered over the career in the benchmark cases, above undistorted levels at every age in the length case and approaching undistorted levels only at retirement in the pace case.

Finally, an empirical literature asks how firms learn about worker productivity [Jovanovic, 1979, Farber and Gibbons, 1996, Acemoglu and Pischke, 1998, Altonji and Pierret, 2001, Lange, 2007, Kahn and Lange, 2014, Cella, Ellul, and Gupta, 2017, Aryal et al., 2022]. This paper draws on that literature for the inputs of its calibration, and connects it, through the sufficient-statistics formulas, to the design of taxes. A large literature on dynamic labor supply and human capital [Heckman, 1976, Imai and Keane, 2004, Altonji, Smith Jr., and Vidangos, 2013] emphasizes that workers anticipate the effect of today's labor supply on future wages; the same anticipation operates here, driven by signaling rather than skill accumulation, and Appendix B shows which conclusions survive when both operate at once.

Section 2 presents the model. Section 3 defines and characterizes the dynamic labor wedge. Section 4 derives optimal taxes, develops the pace case, and characterizes implementation with simple instruments. Section 5 calibrates the labor wedges from existing estimates, and Section 6 concludes; proofs are in Appendix A and extensions in Appendix B.

2 A Signaling-by-Doing Model

This section adapts a standard dynamic model of labor supply, demand, and taxation by adding a single constraint: firms have limited information about workers' productivities and can contract only on a subset of the activities workers perform. Employers see a worker's *resume*—a scalar that summarizes the history of completed deliverables at each point in the worker's career—and pay workers for the execution of these deliverables. Workers, aware of how their resumes will be read, choose their labor supply balancing costs and benefits, which include current wages and future wage increases. The model is a counterpart of the Arrow [1962] learning-by-doing model in which the completion of deliverables affects employers' perceptions of productivity rather than productivity itself; for this reason we call it a *signaling-by-doing* model.

2.1 Environment

Demographics and technology. The economy is populated by a continuum of overlapping generations of workers with finite lives. Workers in each cohort are indexed by a type θ, which determines both their productivity and their preferences. Each worker lives for a continuum of periods normalized to $[0,1]$, and at each date workers of all ages coexist. Workers supply *deliverables* at the flow rate $\tilde{h} : [0,1] \mapsto \mathbb{R}_+$ and consume at the flow rate $\tilde{c} : [0,1] \mapsto \mathbb{R}_+$; histories up to age a are denoted $\tilde{h}^a$ and $\tilde{c}^a$.

A deliverable is a clearly specified, verifiable unit of work, the piece of information on which firms can condition payments. It should not be understood strictly as hours or effort. Two productivity parameters govern how deliverables relate to effort and to output. The parameter $b(\theta) > 0$ should be thought of as how many deliverables a worker can provide per unit of effort l, that is, $\tilde{h} = l \cdot b(\theta)$. The parameter $v(\theta) > 0$ should be interpreted as how much output the worker generates per unit of the deliverable, so that output is $v(\theta) \cdot \tilde{h}$. The total productivity as a function of effort is the product $v(\theta)b(\theta)$. Production is linear and firms are competitive.

Information. Neither parameter is directly observed by anyone other than the worker. Firms observe the execution of the deliverables they pay for, and the worker's resume (defined in Section 2.2), an imperfect record of the past history of deliverables; they do not observe effort l, b, v, the full history of past deliverables, or the worker's individual contribution to output.[1] The government observes *histories of earnings* $\tilde{y}^a$—the idea being that taxpayers file returns every year and the tax authority keeps

[1]The fact that firms observe deliverables but not individual output can be thought of through the lens of team production [Alchian and Demsetz, 1972]: production takes place inside a firm that aggregates the work of many workers, so the firm sees average output of a pool but cannot attribute it to individuals.

track of them—but not resumes, deliverables, effort, or types. Table 1 summarizes the information structure. Because neither firms nor the government observe types, the model features what Stantcheva [2014] calls "double adverse selection."

Table 1: Information structure

	Firms	Government
Type θ (hence b, v)	—	—
Effort l	—	—
Deliverables transacted with the firm	✓	—
Resume $I(\tilde{h}, a)$	✓	—
Earnings history $\tilde{y}^a$	—	✓
Individual output $v \cdot \tilde{h}$	—	—

Workers observe all rows of the table about themselves.

Preferences. All results in the paper are derived under the following assumption on preferences; each formal statement cites the case it uses.

Assumption 1 (Preferences). *Workers have preferences $U(\tilde{c}, \tilde{h}, \theta)$ over consumption and deliverable flows, increasing in $\tilde{c}$, decreasing in $\tilde{h}$, and ordered by the type so that willingness to supply deliverables sorts positively with output per deliverable: $v(\theta)$ is nondecreasing in θ, and the marginal rate of substitution of deliverables for consumption (in case (a) below, $MRS(C, H, \theta) \equiv -U_H/U_C$) is nonincreasing in θ, strictly so where a result invokes strict monotonicity. Two cases, not mutually exclusive, are used repeatedly:*

(a) Homogeneity over timing: *there is no heterogeneity in preferences over the timing of consumption and labor supply flows; that is, there are aggregators C and H of the flows, common across workers, such that $U = U(C(\tilde{c}), H(\tilde{h}), \theta)$.*

(b) Time-separable flows: $U = \int_0^1 e^{-\rho a}\, u\big(\tilde{c}(a), \tilde{h}(a); \theta\big)\, da$, *used in the pace-of-the-resume example of Section 4.3.*[2]

Throughout, the phrase "more productive types are more willing to supply deliverables" refers to the ordering of the assumption. It is an ordering on the output per

[2]Case (b) satisfies case (a) when the flow utility is additively separable between consumption and labor, the discount rate is common across workers, and the consumption felicity and the labor disutility are common functions scaled by the type, as in the example of Section 4.3 (Lemma 6 in Appendix A).

deliverable $v(\theta)$, not on total productivity $v(\theta)b(\theta)$: what the results require is that the types firms would like to identify are the ones that find deliverables cheaper to supply.

Under case (a), and provided that pay depends on the labor supply flows only through their discounted sum, the aggregators themselves can be dispensed with: worker behavior depends only on lifetime discounted consumption and lifetime discounted labor supply (Lemma 5 in Appendix A).

The worker's problem. Workers are forward-looking: they understand that today's labor supply affects the information firms will have about them in the future. They maximize utility subject to a lifetime budget constraint in which flows at age a are discounted at the exogenous rate $q(a)$:[3]

$$V = \max_{\tilde{c},\tilde{h}} U(\tilde{c},\tilde{h},\theta) \quad \text{s.t.} \quad \int_0^1 q(a)\Big[\tilde{c}(a) - w\big(I(\tilde{h},a,\theta)\big)\tilde{h}(a) + T(\tilde{y},a)\Big]da \leq 0, \quad (1)$$

where the salary per deliverable w depends on the information $I(\tilde{h},a,\theta)$ that firms have about the worker, and $T(\tilde{y},a)$ denotes tax payments, discussed below. A standard life-cycle labor supply problem is the special case in which wages equal productivity and are independent of labor supply decisions. Note also that human capital accumulation in the form of on-the-job learning would generate concerns for the worker analogous to those in (1), with labor supply affecting future wages through real productivity increases rather than through perceptions; we return to this observational-equivalence point in Section 2.3 and Appendix B.

2.2 Information, Contracts, and Equilibrium

Contracts. Firms are constrained to offer short-term contracts: they pay workers for the execution of a single deliverable and cannot commit not to renegotiate once more information becomes available. This makes the firm problem essentially static. There is free entry and exit.

[3]There is a technology for transferring resources across periods at rates q. All budget and resource constraints in the paper are lifetime constraints of a single cohort, so they are well defined for any positive rates, including the undiscounted benchmark $q \equiv 1$ maintained in the length case of Section 2.3.

Resumes. To assess the value of a worker's deliverables, firms observe a scalar signal of past experience: the worker's resume.

Assumption 2 (Resumes)**.** *The resume is a weighted sum of past completions of deliverables,*

$$I(\tilde{h}, a) = \int_0^a \varphi(\tilde{a}, a)\, \tilde{h}(\tilde{a})\, d\tilde{a}, \tag{2}$$

with weights φ *positive and continuous for* $0 \leq \tilde{a} \leq a$*,* $a > 0$*, and with uniformly bounded mass,* $\sup_a \int_0^a \varphi(\tilde{a}, a)\, d\tilde{a} < \infty$*.*

Positivity of the weights is what makes exerting effort improve the resume, and it is used throughout, beginning with Lemma 8 and Proposition 3. Continuity and the bounded mass guarantee that the resume and the conditional expectations of productivity are well defined. The lifetime marginal returns of Section 3 involve, in addition, the derivative of the equilibrium salary schedule; we focus throughout on equilibria whose salary schedules are nondecreasing and locally Lipschitz in the resume, as are those constructed in Proposition 7, so the derivative exists almost everywhere and the reputational term in (11) is well defined. Where more is needed, namely continuous differentiability of φ in its second argument in the necessity part of Proposition 4, it is stated there. The leading special case, developed in Section 2.3, is the *length* of the resume, $\varphi \equiv 1$ together with $q \equiv 1$, so that the resume is the cumulative amount of work the worker has supplied so far; its discounted counterpart, $\varphi(\tilde{a}, a) = q(\tilde{a})$, plays the same role when discount rates are general (Proposition 4). Signals that weight the *timing* of deliverables—such as the *pace* $\varphi(\tilde{a}, a) = 1/a$, which has unit mass at every age and so satisfies the assumption although it is unbounded near the start of the career—generate intertemporal distortions and are developed in Section 4.3, once the tax instruments needed to characterize their equilibria are on the table.

The resume is an imperfect measure of the past history of deliverables and their timing. Together with the overlapping-generations structure, formulation (2) is a parsimonious way of introducing noise in how employers assess what has been done and when: employers may pool together workers from different generations who had a different number of opportunities to work, as long as their past experiences are assessed as equivalent. In this construction, what is referred to as age a should be interpreted broadly as the *effective time the worker had to produce*—opportunities to work. Calendar age may be easy to extract from a resume, but it is much harder to verify how much time a worker effectively had available: health shocks, family responsibilities, and other obstacles are difficult for employers to observe. The weighting function $\varphi(\tilde{a}, a)$ depends on both the age at completion and the current age, so the formulation already accommodates employers who partially condition on calendar age.[4]

Firms can infer expected productivity given the signal—either because they have hired many employees with the same resume and observed their average output, or because they know the economy-wide distribution of productivities and labor supply. Figure 1 represents the flows of production, information, and payments: workers complete deliverables; the firm sees the deliverables and the resume, which adds up and weighs the completion history; payments are based on what the firm observes; output is realized, but the firm cannot attribute it to individual workers.

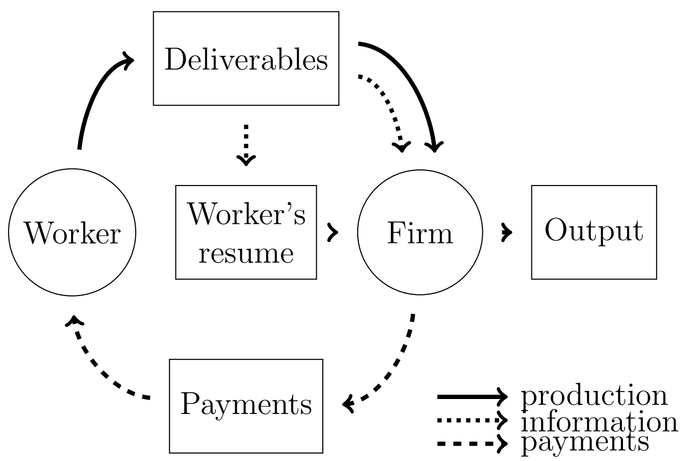

Figure 1: Flows of production, information, and payments

The government. The government faces a budget constraint and saves and borrows at the same rates q as workers. Because workers face a lifetime budget constraint, the timing of tax payments is not pinned down; we denote by $R(\tilde{y})$ the lifetime post-tax earnings of a worker with pre-tax earnings history $\tilde{y}$, and by $R(y)$ the same object when taxes can be written as a function of lifetime income y alone.

[4]Appendix B analyzes richer cases, including the one where the resume is the whole history of deliverables, so firms effectively track the (current and past) ages at which every deliverable was completed.

Definition 1 (Equilibrium). *An equilibrium consists of consumption and labor supply flows, salaries, and taxes such that: (i) workers choose $(\tilde{c}, \tilde{h})$ optimally, taking salary schedules and taxes as given and anticipating the effect of their labor supply on future salaries and taxes, as in (1); (ii) firms earn zero profits on each deliverable, paying each worker their expected productivity conditional on the resume,*

$$w\big(I(\tilde{h}, a, \theta)\big) = \mathbb{E}\big[v(\theta) \,\big|\, I(\tilde{h}, a, \theta)\big]; \tag{3}$$

[5] *and (iii) the government budget constraint holds in present value cohort by cohort,*

$$\mathbb{E}_\theta\left[\int_0^1 q(a)\, \tilde{y}_\theta(a)\, da \;-\; R(\tilde{y}_\theta)\right] \;\geq\; 0, \tag{4}$$

where $R(\tilde{y}_\theta)$ denotes the present value of post-tax lifetime earnings delivered to type θ, and the expectation is over the types of a single cohort; since firms break even on every pool, pre-tax lifetime earnings equal contributions to output on average across the cohort, though not type by type, the difference being precisely the cross-subsidies the model studies; hence (4) coincides with the resource constraint used in the planner problems of Section 4. Requiring the budget to balance within each cohort rules out transfers across cohorts, and it keeps the government problem well defined without taking a stance on the present value of resources across the infinite sequence of cohorts.

Equation (3) is the core of the model: at any date, workers with equivalent resumes are paid equally, and firms on average break even. It is an Akerlof [1970] lemons condition; under the maintained assumption that those who are more willing to work are more productive, selection is advantageous rather than adverse, so it might more precisely be called a "peach" condition.

Equilibrium concept. Contracts are nonexclusive and short-term: a firm cannot prevent a worker from completing deliverables for other firms, and cannot commit to long-term wage paths. In such settings, wages set competitively conditional on the public signal, as in condition (3), are the natural outcome, and Attar, Mari-

[5] Because profits are zero on each deliverable, the expectation in (3) is taken under the equilibrium distribution of deliverables: among workers whose resume equals I, each is weighted by their current deliverable flow; in the length case this is the stationary pool described after Lemma 1. Wages condition on the resume alone; the transacting firm observes the deliverable it buys, but the flow is not part of the public signal that the resume summarizes.

otti, and Salanié [2011] provide game-theoretic foundations in a static context.[6] As in other signaling models, condition (3) does not restrict out-of-equilibrium beliefs. Throughout, we focus on equilibria in which expected productivities conditional on the resume are determined in equilibrium by (3) and salaries are nondecreasing and locally Lipschitz in the resume (the natural class when the more productive supply more), and we ignore pooling equilibria sustained by off-path beliefs. Proposition 7 in Appendix A gives sufficient conditions for the existence of a separating equilibrium in this class in the length case. Equilibrium selection then enters the analysis as follows. The positive results of Section 3, the properties of careers and the decomposition of the wedges, hold at any equilibrium in the class. The optimal taxation results of Section 4 condition on the selected equilibrium: the tax formulas take the equilibrium salary schedule as the object that tax variation perturbs, and, when several equilibria exist, we make the usual mechanism design assumption that the planner can select among them. What does depend on the selection is the level of the wedges; what does not is the structure of the correction, which relies on firms breaking even, on salaries nondecreasing in the signal, and on the invertibility conditions of Lemma 2; for general signals it relies on the hypotheses of Proposition 6. Formally, the planner problem treats the equilibrium mapping as an object of choice: for every tax system, the planner selects among the equilibria of the class above, and the variational arguments of Section 4 perturb taxes within the selected family.

Why a scalar signal. If firms observed the detailed history of deliverables and types were single-dimensional, in most sensible cases the history would reveal the worker's type almost instantaneously, wages would equal marginal products, and the pooling at the heart of this paper would disappear, leaving a standard signaling model; the scalar resume is the tractable way to keep employer learning imperfect without postulating explicit noise processes. Lemma 3 in Appendix A formalizes this claim: under strict single crossing, in any separating equilibrium in which salaries vary smoothly with histories along equilibrium paths, salaries reveal the type of almost every worker at every age, and wages equal marginal products almost everywhere.

[6] The Miyazaki–Wilson–Spence concept used by Stantcheva [2014] presumes exclusive menus on which firms make offsetting losses and profits. With free entry and nonexclusive pay-per-deliverable contracts, the profitable side of any cross-subsidized menu can be poached by an entrant offering marginally better terms on those deliverables alone, so such menus cannot be sustained. Section 2.4 makes the comparison between the two concepts explicit.

Appendix B.1.3 analyzes the full-history case with a type space rich enough that all histories are on path, where no belief specification is needed.

2.3 The Length of the Resume

The benchmark case maintained through Sections 3–4.2 is the one in which there is no discounting within careers, $q \equiv 1$, and the resume is summarized by its *length*: the cumulative sum of deliverables, $I(\tilde{h}, a) = h(a) = \int_0^a \tilde{h}(\tilde{a})\, d\tilde{a}$.

This case is natural when most heterogeneity in labor supply is at the extensive margin and when firms find it hard to verify much more than the start and end dates of previous positions, or the hours worked in the current one. It is also natural when firms believe returns to experience reflect human capital: a firm observing that workers with longer resumes are on average more productive may attribute the difference to learning-by-doing rather than to the changing composition of the pool, posting the same salaries and breaking even either way. Workers themselves need not know whether working more makes them more productive or merely signals that they are; for their labor supply choices, only the mapping from experience to future salaries matters. The normative implications of signaling and of human capital accumulation, however, differ sharply, which is why Section 3 develops the statistics that distinguish them and Section 5 gauges the importance of the signaling component of the returns to experience.

Under this information structure, the Akerlof condition takes a particularly transparent form.

Lemma 1 (Salaries under the length case)**.** *If $q \equiv 1$ and $I(\tilde{h}, a) = \int_0^a \tilde{h}(\tilde{a}) d\tilde{a}$, then the zero-profit condition (3) is equivalent to*

$$w(h) = \mathbb{E}\big[v(\theta)\,\big|\,h(\theta) \geq h\big], \tag{5}$$

where $h(\theta) \equiv \int_0^1 \tilde{h}_\theta(\tilde{a})\, d\tilde{a}$ is the length of the type-θ resume at the end of the career.[7]

Salaries at resume length h equal the average productivity of all types who *eventually* accumulate a resume longer than h: for every type who eventually passes h,

[7]At an atom of the distribution of final lengths, the pool for the marginal deliverable at h is $\{h(\theta) \geq h\}$ and $y'(h)$ is the right derivative; statements involving w and y' hold almost everywhere. A mass of nonparticipants creates the same boundary at $h = 0$.

there is at any date someone—potentially from a different cohort—whose resume is exactly h today.[8]

It is worth clarifying over which population the expectation in (5) is taken. In the stationary cross-section, cohorts are of equal size, and with $q \equiv 1$ each type supplies exactly dh deliverables as its resume passes from h to $h + dh$, at whatever age it does so. The deliverables executed at length h are therefore distributed across the types who eventually pass h with the population weights, which is what (5) asserts.

Two consequences follow from (5) by a change of variables. Lifetime income can be written by integrating over increases in resume length rather than over time,

$$y(h) = \int_0^1 w\big(h(a)\big)\, \tilde{h}(a)\, da = \int_0^h w(z)\, dz, \tag{6}$$

so that (i) the lifetime gain from one extra unit of the deliverable is the same at every point in the career, and (ii) that gain equals the payment for the *last* unit supplied, $y'(h) = \mathbb{E}[v \mid h(\theta) \geq h]$. The economics behind (i) is a tradeoff between completing deliverables earlier—smaller direct payments now, larger future reputational benefits—and later—larger direct payments (the resume is already longer), smaller remaining reputational benefits. Under the length case these two forces exactly offset, so the timing of effort is undistorted; this is the observation that Section 3 generalizes into a characterization of *when* signaling distorts timing and when it does not. Property (ii) makes this dynamic career-concerns model as tractable as a static one.

In the economically sensible case where more productive workers are more willing to supply deliverables, salaries increase with experience. All workers start with an empty resume and the same salary—the average productivity of entrants. As hard-working individuals complete deliverables, they progressively separate from those who supply less; the more productive begin their careers subsidizing the less productive, and with every task some less productive workers are left behind with shorter resumes, so the remuneration of those who continue climbs. The size of the return to experience therefore depends on the joint distribution of preferences and productivities.[9]

[8] If labor supply heterogeneity is only at the extensive margin, or if deliverables are defined as working additional years, the pooled set consists only of workers of the same cohort.

[9] Section 3.1 establishes the properties that hold whenever the resume carries no information at the start of the career and signal paths are monotone: all workers start with identical salaries; there is a positive signaling return to experience; and at the peak of each career, salaries exceed marginal products. The pace case, informative from the first instant, shows the conditions have content.

A parametric example. Let preferences over lifetime consumption $C = \int_0^1 \tilde{c}(a)\, da$ and lifetime labor supply $H = \int_0^1 \tilde{h}(a)\, da = h(1)$ be quasilinear in consumption with a constant elasticity of labor supply,[10]

$$U(C, H, \theta) = C - \left(\frac{H}{b(\theta)}\right)^{1+\frac{1}{\epsilon}} \left(1 + \tfrac{1}{\epsilon}\right)^{-1}. \tag{7}$$

It is worth emphasizing that specifying preferences directly over these lifetime aggregates is without loss here. Lifetime earnings depend on the labor supply flows only through the lifetime labor supply (6), and the same is true of post-tax earnings once taxes condition on lifetime income (Proposition 5); by Lemma 5, worker behavior then depends on the flows only through C and H.

We parameterize $b(\theta) = \theta^{1-\delta}$ and $v(\theta) = \theta^{\delta}$, with θ Pareto-distributed with shape $\alpha > 1$ and $0 \leq \delta < 1$. Total productivity per unit of effort is $vb = \theta$, and δ governs the share of heterogeneity that is invisible to firms: a higher δ loads heterogeneity onto the value per deliverable v (which firms cannot see) rather than the deliverable rate b (which shows up in the volume of completed deliverables). Guessing and verifying that salaries are a power function of resume length delivers a log-linear wage-experience profile,

$$\log w = \gamma \log h + \kappa, \qquad \gamma = \frac{\delta}{1 - \delta + \epsilon}, \tag{8}$$

a constant proportional return to experience driven entirely by selection. The return is larger when δ is higher (more heterogeneity to be screened out by experience) and when the elasticity ϵ is lower (a longer resume is then a stronger indication of high productivity, because marginal increases in labor supply hurt the less productive relatively more).[11] Equation (8) illustrates that a wage-experience profile usually attributed to human capital accumulation can be generated by pure signaling; the two stories have very different normative implications, which is why the distinction—and the moments that identify it—are developed in Sections 3 and 5.

Because higher types are the workers willing to work longest, each worker faces, at the time of retirement, a salary above their own marginal product. And because

[10] Behind this specification there could be either heterogeneity only at the extensive margin, in which case workers of different cohorts would not be pooled together, or richer heterogeneity in preferences, in which case workers of different cohorts potentially would be pooled together.

[11] For h below the final length of the lowest type, the pool is the entire population and $w = \mathbb{E}[v]$; the power branch applies above it. The calibration of Section 5 evaluates the elasticity at ten years of experience, beyond this initial pooling region.

the lifetime benefit of an extra deliverable is timing-independent, these high-powered incentives operate at every moment of the career, not only near retirement: workers work too much over the lifetime, but not at the wrong times. In the parametric example the corrective implications are especially simple—marginal corrective taxes are constant, at rate δ/α, so they can be implemented with history-independent annual taxes—as shown in Section 4.2.

2.4 Relation to Static Double Adverse Selection

We now compare the resume economy with the static double adverse selection economy of Stantcheva [2014]. The primitives are close. Workers of unobserved type supply an observable quantity of labor, firms cannot see productivity, and the government observes only incomes, although her analysis works with a finite set of types while ours allows a continuum. The substantive differences are the horizon of information on which firms condition pay and the contract space.

Wages. The length case admits a static representation (6), which allows a term-by-term comparison. In the resume model, a worker who supplies lifetime labor h collects

$$y_{\text{resume}}(h) = \int_0^h \mathbb{E}\big[v(\theta) \mid h(\theta) \geq z\big]\, dz : \tag{9}$$

each successive unit is paid according to the pool of workers whose careers reach *that* unit. In the static MWS economy, pay depends on the final quantity alone. Stantcheva [2014] characterizes two regimes. When within-menu cross-subsidies are inactive, each contract breaks even and pay satisfies

$$y_{\text{MWS}}(h) = h \cdot \mathbb{E}\big[v(\theta) \mid h(\theta) = h\big]. \tag{10}$$

The difference is the treatment of inframarginal units. In MWS, a worker who increases lifetime labor supply enjoys higher payments on *all* inframarginal units, which are revalued as coming from a more productive worker. In the resume model, workers distinguish themselves from less productive peers only gradually, and increasing labor supply leaves the payments already collected on inframarginal units unchanged. Seen dynamically, MWS is the limiting case in which employers pay every unit according to the worker's *final* resume rather than the current one—as if reputations were established instantaneously. Both models generate high-powered incentives; the resume model additionally generates implicit cross-subsidies *through the life cycle*, from young high-productivity workers to the pools they pass through, that are absent from MWS.

The comparison comes with one caveat. The MWS equilibrium (her Definition 1) has firms post menus of pay–hours contracts that break even only on the overall portfolio; in the regime where within-menu cross-subsidies are active, high-productivity workers are paid less than their product and low-productivity workers more than theirs, so pay departs from (10). The direction of the implicit transfers is then the same as here, but the channel differs: there, within contemporaneous menus sustained by exclusivity; here, through the life cycle.

3 Dynamic Labor Wedges

The model of Section 2 implies that what workers are paid for a marginal deliverable and what that deliverable contributes to output are different objects, and—more importantly—that the gap between them is not measured by the difference between the current wage and the marginal product. A worker who completes a deliverable today collects the current salary *and* strengthens their resume in every future period. The object that governs distortions, and that taxes should correct, is therefore the *lifetime* marginal return to a deliverable, compared with its marginal product. This section establishes general properties of careers, defines these dynamic labor wedges, and characterizes when they are constant over the career of each worker and when dynamic signaling instead introduces intertemporal distortions.

3.1 General Properties of Careers

We begin with properties of salaries and the return to experience that hold for any signal of the form $I(\tilde{h}, a) = \int_0^a \tilde{h}(\tilde{a})\varphi(\tilde{a}, a)\, d\tilde{a}$ with $\varphi > 0$ bounded and continuous—the "strength of the resume."

First, for any preferences and distribution of productivities, every worker starts their career with nothing to show: the strength of the resume is zero, so everyone earns the same initial salary, equal to the average productivity of all entrants. Workers with above-average productivity start their careers earning less than their marginal product; all must climb the same ladder. This is illustrated in the left panel of Figure 2, where the pool of entrants is circled at the origin.

Second, workers at the peak of their careers—when they reach the highest value of the index—earn more than their marginal product whenever that value is exceeded with positive probability, because they are then pooled only with types willing to provide more deliverables across all periods, who are more productive; at an unexceeded top index the competitive wage equals the worker's own productivity. In the left panel of Figure 2, the dashed horizontal line collects the workers who at some point hold a resume of a given strength; wherever the line is tangent to the highest strength a worker achieves, that worker is pooled only with types reaching yet higher levels. The left panel also shows that an individual resume may deteriorate over time—for example, in the pace case, slowing down weakens the resume and is read as bad news about productivity—a case outside the monotone-path hypothesis of Proposition 1 but covered by the wedge definitions of Section 3.2.[12]

Third, provided more productive types are more willing to provide deliverables at all periods, exerting more effort raises future salaries: a stronger resume resembles the resumes of more productive types. In the right panel of Figure 2, a worker who exerts extra effort at some age moves from the dashed to the solid path, strengthening the resume at every subsequent age.

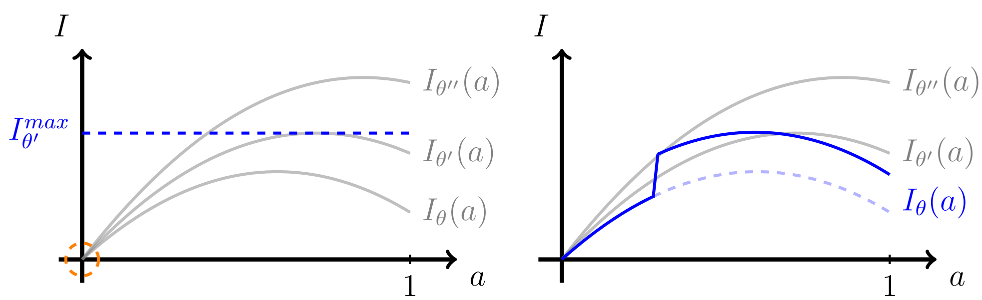

Figure 2: Careers of workers. Left panel: initial salaries and salaries at the peak. Right panel: signaling return to experience.

Proposition 1 (General positive properties)**.** *Let* $I = \int_0^a \varphi(\tilde{a}, a)\tilde{h}(\tilde{a})\, d\tilde{a}$ *with* $\varphi > 0$ *and* $\tilde{h} \geq 0$ *bounded and continuous, let more productive workers be more willing to supply deliverables (Assumption 1), and consider the equilibria of Section 2.2. Suppose*

[12] If firms believe there is more human capital depreciation than there actually is, they may also discount past experiences too heavily.

(i) the resume carries no information at the start of the career, $\int_0^a \varphi(\tilde{a}, a)\, d\tilde{a} \to 0$ as $a \to 0$, and (ii) equilibrium signal paths are continuous, nondecreasing in age, and increasing in the type, strictly at the terminal age, with types supplying deliverables at the ages at which their paths cross others' terminal values (automatic in the length case, where the resume grows only while deliverables are supplied). Then, for any distribution of preferences and productivities:

1. *all workers start with effectively empty resumes and earn the same initial salaries, the average productivity of entrants;*

2. *every worker whose final signal value is exceeded with positive probability is pooled, at the peak of the career, with all more productive types, whose paths cross that value on their way to higher terminal values; hence at the peak of the career of each such worker, salaries are higher than productivities, equal to a deliverable-weighted average of the productivities in the worker's upper tail;*

3. *there is a positive signaling return to effort: exerting more effort at any age strengthens the resume at every subsequent age, and with it weakly raises all subsequent salaries, strictly where the salary gradient is positive; a stronger resume resembles the resumes of more productive types.*

Proof. Under (i), signals at the start of the career converge to zero for every type, so entrants are pooled and paid the average productivity $\mathbb{E}[v]$. Under (ii), a less productive type's path never reaches the worker's terminal value, since its own terminal value lies strictly below it, while every more productive type's path starts below that value and ends above it, hence crosses it: the pool at the peak is the tail of more productive types, weighted by their deliverables at the crossing, which gives Parts 1 and 2. For Part 3, raising $\tilde{h}(a)$ raises the resume at every subsequent age, since $\varphi > 0$, and salaries are nondecreasing in the signal in the equilibria of Section 2.2, strictly increasing where the gradient is positive. □

Condition (i) fails for averaging signals such as the pace, which is informative from the first instant; there the equilibrium separates, salaries equal marginal products at every age, and all three properties fail, consistent with Section 4.3.

3.2 Definition and Decomposition

Fix a type θ following an optimal plan, and let $Y(\theta) = \int_0^1 q(a)\, w\big(I(\tilde{h}, a)\big)\tilde{h}(a)\, da$ denote lifetime earnings in present value. Consider a marginal increase in deliverables at age a, holding the rest of the plan fixed. The present-value earnings gain has two parts: the payment for the marginal deliverable itself, and the increase in the payments for all future deliverables induced by the stronger resume,

$$\frac{dY}{d\tilde{h}(a)} = q(a)\, w\big(I(a)\big) + \int_a^1 q(\tilde{a})\, w'\big(I(\tilde{a})\big)\, \frac{\partial I(\tilde{a})}{\partial \tilde{h}(a)}\, \tilde{h}(\tilde{a})\, d\tilde{a}. \tag{11}$$

Define the *marginal lifetime return per (present-value) unit of the deliverable* as

$$m(a;\theta) \equiv \frac{1}{q(a)} \frac{dY}{d\tilde{h}(a)}, \tag{12}$$

and the *dynamic labor wedge* at age a as the ratio of the marginal product to that return,

$$\chi(a;\theta) \equiv \frac{v(\theta)}{m(a;\theta)}. \tag{13}$$

When $\chi(a;\theta) < 1$, the worker collects more, over the lifetime, than the marginal deliverable contributes to output: incentives at age a are high-powered. When $\chi(a;\theta)$ *varies with* a, the worker additionally faces distorted incentives about *when* to work.

At the last unit of labor a worker supplies there are no future deliverables on which reputational gains could be collected, so the reputational term in (11) vanishes and the marginal return equals the salary on the final deliverable. We write $m(T(\theta);\theta)$ for this endpoint, where $T(\theta) \leq 1$ is the age at which the worker supplies their last unit of labor, with $T(\theta) = 1$ for a worker active throughout. The endpoint is well defined even when the signal keeps evolving with age, since the reputational term multiplies future labor supply, which is zero after retirement; in the length case the marginal return is age-invariant, so $m(T;\theta) = m(1;\theta)$. The wedge process then admits a decomposition.

Proposition 2 (Decomposition of dynamic labor wedges)**.** *Define the labor wedge at* *the time of retirement and the intertemporal factor*

$$\underbrace{\chi(\theta) \equiv \frac{v(\theta)}{m(T(\theta);\theta)}}_{\textit{wedge at retirement}} \qquad \textit{and} \qquad \underbrace{\iota(a;\theta) \equiv \frac{m(T(\theta);\theta)}{m(a;\theta)}}_{\textit{intertemporal distortions}}. \tag{14}$$

Then $\chi(a;\theta) = \chi(\theta) \cdot \iota(a;\theta)$ *at every age* $a \le T(\theta)$*. If, in addition, along the equilibrium path the discounted influence of an age-*a *deliverable on the resume at each later age at least matches the resume's own drift,*

$$\frac{q(\tilde{a})}{q(a)}\,\varphi(a,\tilde{a})\,\tilde{h}(\tilde{a}) \;\ge\; \dot{I}(\tilde{a}) \qquad \textit{for } a \le \tilde{a} \le T(\theta), \tag{15}$$

where $\dot{I}(\tilde{a})$ *denotes the drift of the resume along the equilibrium path, then* $\iota(a;\theta) \le 1$*, with equality at* $a = T(\theta)$*.*

Proof. The product identity is immediate from (13). For the inequality, by (11) and $m(T;\theta) = w(I(T))$,

$$m(a;\theta) - m(T;\theta) = \int_a^T w'(I(\tilde{a}))\left[\frac{q(\tilde{a})}{q(a)}\,\varphi(a,\tilde{a})\,\tilde{h}(\tilde{a}) - \dot{I}(\tilde{a})\right] d\tilde{a},$$

using $w(I(T)) - w(I(a)) = \int_a^T w'(I(\tilde{a}))\,\dot{I}(\tilde{a})\,d\tilde{a}$ and the fact that after T future labor supply is zero. The integrand is nonnegative under (15) and the monotonicity of salaries maintained in Section 2.2, so $m(a;\theta) \ge m(T;\theta)$; at $a = T(\theta)$ the integral vanishes. □

Condition (15) is the sense in which building a reputation earlier pays weakly more than later. It holds with equality at every age in the length case, where $\varphi \propto q$, and Proposition 4 shows that equality throughout characterizes the absence of intertemporal distortions. It can fail for signals with rapidly decaying memory: when early deliverables are soon forgotten, the return to effort is lower early in the career than at its end, and ι rises above one.

The wedge at retirement $\chi(\theta)$ compares the marginal product with the pay on the final deliverable. The retirement margin is where the wedge is measured, not where the distortion lives. In the length case the same wedge applies to every unit of labor the worker supplies over the career. The factor $\iota(a;\theta)$ compares the return to working at age a with the return at the end of the career; it falls below one whenever building a reputation earlier pays more than later. Two questions organize everything that follows. First, is $\chi(\theta)$ below one, so that over the lifetime workers are paid more than their marginal products for the last unit of labor they supply? Second, does $\iota(a;\theta)$ fall below one, so that dynamic signaling additionally distorts the timing of labor supply? There are no intertemporal distortions when $\iota \equiv 1$, in which case $\chi(a;\theta) = \chi(\theta)$ is a single number per worker.

The two polar cases of this paper give opposite answers to the two questions. When the resume is summarized by its length (Section 2.3), equation (6) implies $m(a;\theta) = w(h(\theta))$ for every a: $\iota \equiv 1$, and the whole distortion is the wedge at retirement, $\chi(\theta) = v(\theta)/\mathbb{E}[v \mid h(\tilde{\theta}) \geq h(\theta)] < 1$. When the resume is summarized by its pace, the amount of deliverables per unit of time available, salaries in the tax-supported equilibrium of Section 4.3 equal marginal products at every instant, so $\chi(\theta) = 1$, and the distortions are entirely intertemporal, with $\iota(a;\theta)$ below one early in the career and rising to one at retirement. In neither case does the comparison of current wages with productivities detect the distortion. In the pace case the comparison shows nothing while intertemporal distortions are large; in the length case the wage–product gap has the wrong sign early in the career, when salaries are still below marginal products, and it reaches the true wedge only at the final unit.

3.3 Minimal Ingredients

Which features of the environment generate the dynamic rat race, and which merely shift pay away from marginal products? The following proposition separates them.

Proposition 3 (The dynamic rat race)**.** *Let the signal take the form* (2) *and let equilibrium salaries satisfy* (3)*.*

(i) (Signaling return.) *Suppose more productive types are more willing to supply deliverables and the salary gradient is strictly positive, $w' > 0$ almost everywhere along the equilibrium path, as in the equilibria constructed in Proposition 7. Then for every worker with $\tilde{h} > 0$ on a positive-measure subset of $(a, 1]$, the reputational term in* (11) *is strictly positive: $m(a;\theta) > w(I(a))$ at every interior age.*

(ii) (Retirement-margin wedge.) *Under the conditions of Proposition 1 (no information at the start of the career, monotone signal paths), $\chi(\theta) < 1$ for every*

worker whose final signal value is produced, at some age, by a positive measure of other types. The pooling condition holds in the length case whenever the final length is exceeded with positive probability, and fails in the pace case, where signal values separate types and $\chi(\theta) = 1$.

(iii) (Vanishing, and what does not vanish.) *If* $v(\theta)$ *is independent of the preference component of* θ*, then* $w \equiv \mathbb{E}[v]$*, the signaling return is zero,* $\iota \equiv 1$*, and the remaining wedge* $\chi(a;\theta) = v(\theta)/\mathbb{E}[v]$ *reflects pure pooling cross-subsidies rather than a rat race. Full-history observation, by contrast, permits separation, and separation removes the pooling but not the rat race. In a separating equilibrium wages equal marginal products on path, yet incentive compatibility along the menu of equilibrium histories keeps the lifetime return to labor above the marginal product wherever* v *is strictly increasing in the type (Lemma 3).*

Proof. See Appendix A. □

Part (iii) draws a line worth emphasizing: the *dynamic rat race* requires effort to move beliefs. Pure pooling—being paid an average productivity unrelated to one's own—pushes pay away from marginal products but does not generate a signaling return to experience; it is the positive dependence between willingness to work and unobserved productivity that converts pooling into a race.

Under the timing condition (15) of Proposition 2, the wedge at the retirement margin is an upper bound on the wedge ratio at every age, so its complement $1-\chi(\theta)$ bounds the strength of career-long distortions from below.

3.4 Comparative Statics

How large are the wedges, and what moves them? In the parametric example of Section 2.3, the lifetime wedge takes the closed form

$$\chi(\theta) = \frac{v(\theta)}{\mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]} = \frac{\alpha - \delta}{\alpha}, \tag{16}$$

constant across workers: the Pareto tail makes the proportional cross-subsidy scale-free. The wedge deepens (moves away from one) as δ rises—more of the heterogeneity is invisible to firms, so the pool above any type is more productive relative to that type—and flattens as the tail parameter α rises, since thinner tails put less productivity mass above any given type. The signaling return to experience (8) moves with the same primitives, $\gamma = \delta/(1-\delta+\epsilon)$, which is what allows the return to experience and the wedge to be linked in the data (Section 5).

Beyond the parametric family, the labor wedges obey general monotone comparative statics, stated and proved in Appendix B.3. When information asymmetries decrease, in the sense that heterogeneity shifts from the unobserved productivities v toward the observable component b holding total productivities and the ranking of types fixed (Definition 4), the wedge at retirement rises pointwise toward one in the length case (Lemma 13). And holding the equilibrium path of salaries fixed, signal weights that reward effort later in the career deepen the intertemporal distortions (Lemma 14). The total effect, which includes the equilibrium adjustment of wages, is not signed in general; the pace case shows that the direct effect can dominate.

3.5 When Does Dynamic Signaling Introduce Intertemporal Distortions?

What makes the length case special is the shape of its signal weights. The following proposition shows that the absence of intertemporal distortions in every economy of a rich class characterizes exactly the signals whose weights are proportional to the discount factors of the execution dates, independent of the worker's current age. In other words, whether taxes on lifetime income can be free of intertemporal corrections across all such economies turns on a single feature of the environment, namely how firms weigh the timing of past deliverables when reading resumes.

Proposition 4 (Absence of intertemporal distortions)**.** *Within the class of signals* (2)*, write the weights as $\varphi(\tilde{a}, a)$, the weight that the age-a resume places on a deliverable completed at age $\tilde{a} \le a$. Then:*

- *(i)* (Sufficiency.) *If $\varphi(\tilde{a}, a) = c\, q(\tilde{a})$ for some $c > 0$—the resume is equivalent to the discounted cumulative sum of deliverables—then for every economy satisfying Assumption 1, $m(a;\theta) = w(h(\theta))$ for all a: $\iota \equiv 1$, and, absent bunching in lifetime income (so that the wedge is single-valued in y), the schedule of labor wedges $\chi(y)$ is a sufficient statistic for the corrective component of taxes.*
- *(ii)* (Necessity.) *Conversely, suppose φ is continuously differentiable in its second*

argument, and that its cumulative weight $\int_0^{\tilde{a}} \varphi(z, \tilde{a})\,dz$ has a continuous and strictly positive derivative in $\tilde{a}$, so that the resume drifts strictly upward along a career of constant flow. If $\iota \equiv 1$ in every economy of a rich class $\mathcal{E}$ (Definition 2 in Appendix A), then $\varphi(\tilde{a}, a) = c\, q(\tilde{a})$ for some $c > 0$, almost everywhere on the support of equilibrium allocations.

Proof. See Appendix A. Sufficiency is the change of variables behind (6). For necessity, constancy of $m(\cdot;\theta)$ across ages forces, economy by economy, the present-value reputational kick of an age-a deliverable to integrate to the same total regardless of a; richness of $\mathcal{E}$ makes the equilibrium wage gradients vary enough that this can hold only if the weight on an age-a deliverable is proportional to $q(a)$ and independent of the evaluation age. □

Richness asks only that the class contain the extensive-margin family used in the proof, so the necessity conclusion applies to any larger class of economies.

The proposition answers the question of which features of the environment map into the structure of corrective taxes. If the weights firms place on past deliverables are aligned with the discount rates and do not depend on the worker's current age, as in the length case, where $q \equiv 1$ and the weights are constant, then dynamic signaling distorts how much people work over their lifetimes but not when, in every economy. The corrective instrument is a schedule of labor wedges over lifetime incomes, and, when households also agree on the timing of consumption and labor supply (Assumption 1(a)), lifetime income is the optimal tax base (Section 4). Outside this class, and within the regularity domain of the necessity part, the proposition guarantees only that *some* economy features intertemporal distortions; whether a given economy does, and in which direction, turns on the timing condition (15) of Proposition 2. In the "pace-of-the-resume" case, where $\varphi = 1/a$, the benefits of exerting effort are higher at early stages of the career, and the corrective instrument must be age-dependent; under the parametric assumptions of Corollary 2, age and current earnings suffice. Section 4.3 develops the pace case in full once the tax instruments needed to characterize its equilibrium are available, and derives the closed-form age-dependent correction (Corollary 2).

Section 5 puts numbers on the schedule $\chi(y)$, reading it off the free-entry condition and existing estimates of the signaling component of the returns to experience.

4 Optimal Taxation

This section characterizes optimal taxes. The government maximizes a welfarist functional of worker utilities, observing histories of earnings but not types, resumes, or deliverables (Table 1). The results are organized by the structure of the labor wedges of Section 3. When there are no intertemporal distortions (Proposition 4), *lifetime income* suffices for the corrective component and, under Assumption 1(a), is an optimal base for the entire system; optimal taxes take a generalized Mirrleesian form whose corrective factor is the lifetime wedge (Sections 4.1–4.2); when it is not, the corrective layer must additionally undo timing distortions, and becomes age- and history-dependent (Section 4.3). Section 4.4 characterizes implementation with simple instruments, and Appendix B.4 studies the welfare effects of changes in information technologies.

4.1 The Lifetime Income Base

We first place special focus on the case where the government conditions taxes only on lifetime income y. The focus is motivated by the following extension of the Atkinson and Stiglitz [1976] uniform-taxation result, and, on the practical side, by the observation that many actual taxes and transfers condition on lifetime income—most notably, the key determinant of US Social Security benefits is a measure of average lifetime earnings.

Proposition 5 (Lifetime income base). *Suppose preferences are homogeneous over the timing of consumption and labor supply flows (Assumption 1(a)), with the aggregators C and H strictly increasing, and let resumes be summarized by their length, $q \equiv 1$ and $I(\tilde{h}, a) = \int_0^a \tilde{h}(\tilde{a})\, d\tilde{a}$. Then there exists an optimal tax system that conditions only on lifetime income, even with double adverse selection.*

Proof. See Appendix A. □

The content of the proposition is that when dynamic signaling introduces no intertemporal distortions and households agree on the timing of consumption and labor, there is no reason for the tax system to distort intertemporal decisions. Its logic is more general than its statement: optimal policy should, whenever possible, make *post-tax* lifetime earnings a function of lifetime labor supply. In the length case, *pre-tax* lifetime earnings are already such a function (6), so lifetime income taxes suffice. Away from that benchmark, when pre-tax earnings depend on the timing of labor supply (Section 4.3), optimal policy entails additional taxes that undo the intertemporal distortions—effectively restoring post-tax earnings as a function of discounted lifetime labor supply. Conversely, if the homogeneity assumption is relaxed, the intertemporal margin becomes a potentially useful screening instrument, and lifetime income is no longer guaranteed to suffice as the base of the *redistributive* component, even though it still suffices for the *corrective* component; relaxing both assumptions at once removes the guarantee from both components (Appendix B). Although the tax base is lifetime income, taxes need not be collected at the end of life: they can be raised annually, provided each year's taxes may depend on the history of earnings to date.

4.2 Redistribution and Correction: A Decomposition

We now derive necessary conditions for optimal taxes in terms of sufficient statistics, as in Saez [2001]. The government chooses the lifetime retention function $R(y)$ mapping pre-tax to post-tax lifetime income, to solve

$$\max_{R(y)} \mathbb{E}\big[W\big(V(R;\theta),\theta\big)\big] \quad \text{s.t.} \quad \mathbb{E}\big[y\big(h(R;\theta)\big) - R\big(y(h(R;\theta))\big)\big] \geq 0, \tag{17}$$

and subject to pre-tax wages being determined by the differential equation $y'(h) = \mathbb{E}[v(\theta) \mid h(R;\theta) \geq h]$ with $y(0) = 0$.[13]

Solving directly for $R(y)$ is complicated. Changing taxes around an income level y changes the incentives of workers there; by selection, this changes the joint distribution of resumes and productivities; firms, recognizing the new pool, update salaries; salary updates change incentives again, and so on—a cascade. With income effects the cascade also propagates upward through the income distribution. The following lemma cuts through it.

Lemma 2 (Solving in post-tax space)**.** *Without loss, the planner can solve directly for $\tilde{R}(h) = R(y(h))$—choosing post-tax income as a function of lifetime labor supply,*

[13] The welfarist formulation is general: choosing linear functionals $W = \lambda(\theta)V(R;\theta)$ converts the problem into Pareto-efficiency tests as in Werning [2007a], Hosseini and Shourideh [2019], and Bierbrauer, Boyer, and Hansen [2023].

with labor valued at its marginal product—and then recover $y(h)$ and $R(y)$:

$$\max_{\tilde{R}} \ \mathbb{E}\big[W\big(V(\tilde{R};\theta),\theta\big)\big] \quad s.t. \quad \mathbb{E}\big[v(\theta)\,h(\tilde{R};\theta) - \tilde{R}\big(h(\tilde{R};\theta)\big)\big] \geq 0. \tag{18}$$

Proof. See Appendix A. □

The intuition is a version of the targeting principle. The cascading salary adjustments are transfers *among workers*: firms break even throughout, so every dollar one worker gains through the revaluation of labor, another loses. A planner equipped with a flexible retention schedule can undo any such transfer, so none of the cascade constrains welfare; what constrains welfare is the resource cost of labor supply, which depends only on true productivities. The lemma requires only that firms make zero profits[14] and that $y(h)$ is well defined and invertible—so it applies to other labor market frictions whose effects run through wages and firm revenues: production externalities of that form, imperfect competition compressing wages, competitive or monopsonistic screening. Externalities outside firm revenues, and untaxed profits or rents, would enter the resource constraint separately.

Theorem 1 (Optimal tax formula)**.** *Let θ be single-dimensional, let the distribution of lifetime labor be atomless with density f, and suppose the optimal allocation has no bunching. If a tax schedule is optimal, it satisfies, at every income level in the interior of the support,*

$$\frac{\chi(y)-r(y)}{r(y)}\,\epsilon_r^c(y)\,g(y)\,y = \int_y^\infty \big(1-\lambda(\tilde{y})\big)g(\tilde{y})\,d\tilde{y} + \int_y^\infty \frac{\chi(\tilde{y})-r(\tilde{y})}{r(\tilde{y})}\,\eta_I(\tilde{y})\,g(\tilde{y})\,d\tilde{y}, \tag{19}$$

where $r(y)$ is marginal retention, ϵ_r^c and η_I are the compensated and income elasticities of lifetime income with respect to retention (holding pre-tax salary schedules fixed), g is the density of lifetime income, $\lambda(y)$ is the normalized marginal welfare weight (the social value of a unit of lifetime income to the worker earning y, in units of public funds), and $\chi(y) \equiv v(y)/y'(h(y))$ is the lifetime wedge of Section 3 expressed as a function of lifetime income.

Proof. See Appendix A. □

[14] Or that profits are fully taxed, or uniformly shared, as in Scheuer and Werning [2017]. Zero profits are what make every discrepancy between pay and product a transfer among workers rather than a resource.

Equation (19) is the standard Saez [2001] condition with one new ingredient, $\chi(y)$. Reading it term by term: when the planner raises marginal retention on a small income interval (holding post-tax salaries fixed elsewhere), workers there supply more labor—compensated effects, left-hand side—and each extra dollar earned eases the resource constraint not by $1 - r$ as in the standard model but by $\chi - r$: a worker paid one extra dollar contributes $\chi(y)$ dollars of output. The mechanical and welfare effects (first term on the right) are unchanged. Income effects (second term) likewise ease or damage the resource constraint by $\chi - r$ per dollar.

The same condition can be stated as a two-block decomposition:

$$r(y) = r_m(y) \cdot \chi(y), \tag{20}$$

with r_m satisfying the *standard* Mirrleesian formula

$$\frac{1 - r_m(y)}{r_m(y)}\, \epsilon_r^c(y)\, g(y)\, y = \int_y^\infty \big(1 - \lambda\big) g\, d\tilde{y} + \int_y^\infty \frac{1 - r_m}{r_m}\, \eta_I\, g\, d\tilde{y}. \tag{21}$$

Optimal retention is the product of a redistributive (Mirrleesian) factor and a corrective (Pigouvian) factor equal to the lifetime wedge. The decomposition captures the externality that career concerns make workers impose on one another: holding remuneration schedules fixed, a worker who increases lifetime labor supply to earn one more dollar contributes χ dollars of output—as if producing one dollar while emitting a negative production externality of $1 - \chi$.

Three observations sharpen the interpretation. First, the decomposition has content beyond bookkeeping. Consider three economies: (a) no information asymmetries; (b) asymmetries only between government and workers (the standard Mirrlees economy); and (c) asymmetries among firms, workers, and government. The first- and second-best utility possibility frontiers share a point (zero taxes in (b) implements the first best). More surprisingly, the first- and *third*-best frontiers also share a point, attained by setting the Mirrleesian factor to one, $r = \chi$: purely corrective taxes restore first-best efficiency at one point of the frontier (Proposition 11 in Appendix B). Second, since workers supplying their last unit of labor are the least productive in their resume pool (Proposition 3(ii)), $\chi(y) \leq 1$: holding estimated elasticities and densities fixed, accounting for career concerns unambiguously pushes toward *higher* marginal taxes at every income level. Third, the elasticities in (19) hold pre-tax salary schedules fixed, while estimated elasticities of taxable income do not. When increases in retention draw marginal, less productive types into work, pre-tax salaries fall and the observed response understates the fixed-schedule elasticity, though other wage-setting responses can offset this. The formula's inputs are also long-run elasticities of *lifetime* income, which may exceed short-run annual elasticities [Imai and Keane, 2004, Saez, Slemrod, and Giertz, 2012, Kurnaz, Michelini, Özdenören, and Sleet, 2026].

In the parametric example of Section 2.3, the Pigouvian factor is constant, $\chi = (\alpha - \delta)/\alpha$ (16), so the corrective layer is a *linear* tax at rate δ/α—implementable with history-independent annual income taxes. Corrective taxation of dynamic signaling does not, in this benchmark, require any sophistication in the tax code.

4.3 Intertemporal Corrections: The Pace of the Resume

When the signal weights depend on the worker's current age, outside the condition of Proposition 4, corrective taxes must undo intertemporal distortions as well as the lifetime rat race; in the pace case, the benefits of exerting effort are higher at early stages of the career. The general result stacks the corrective and redistributive layers as follows.

Proposition 6 (Corrective layers under general signals)**.** *Suppose preferences satisfy Assumption 1(a), the signal takes the form (2), and the optimal allocation is separating, in the sense that no two types choose the same labor supply history. (Separation holds, absent bunching, when higher types choose strictly higher lifetime aggregates and paths are strictly increasing in the aggregate, as under strict versions of the conditions of Lemma 8.) Assume also the conditions of Lemma 9: the weights are pointwise bounded, and salaries, as a function of the signal, are bounded below away from zero and Lipschitz.*[15] *If the retention functional $R(\tilde{y})$ is optimal, there exist R_m, R_p with $R(\tilde{y}) = R_m(R_p(\tilde{y}))$ satisfying the three conditions below, with all derivatives evaluated at the equilibrium earnings histories; the corrective layer is constructed as a globally differentiable function of discounted lifetime labor, so its derivatives are well defined in every direction, and the redistributive layer is completed confiscatorily off the equilibrium menu. Denote by $m_p(a;\theta)$ the post-tax counterpart of the marginal*

[15] Assumption 2 bounds the mass of the weights, not their pointwise values, and the pace kernel $1/a$ violates the pointwise bound near the start of the career. The pace case is nonetheless covered by the explicit construction of Corollary 2, which does not use Lemma 9.

lifetime return $m(a;\theta)$ of (12): the increase in post-tax lifetime earnings under the corrective layer, per present-value unit of the deliverable,

$$m_p(a;\theta) \equiv \frac{1}{q(a)} \left[\frac{dR_p(\tilde{y})}{d\tilde{y}(a)}\, w(\tilde{h}^a) + \int_a^1 \frac{dR_p(\tilde{y})}{d\tilde{y}(\tilde{a})}\, \frac{dw(\tilde{h}^{\tilde{a}})}{d\tilde{h}(a)}\, \tilde{h}(\tilde{a})\, d\tilde{a} \right], \tag{22}$$

read as a function of earnings flows through the inverse operator $\tilde{h}(\tilde{y})$ of Lemma 9. Then:

1. (Intertemporal, Pigouvian) *the corrective layer neutralizes the timing of labor supply: for any ages $a, \bar{a}$,*

$$m_p(\bar{a};\theta) = m_p(a;\theta); \tag{23}$$

2. (Lifetime, Pigouvian) *that common return equals the marginal product: for every age a,*

$$m_p(a;\theta) = v(\tilde{h}); \tag{24}$$

3. (Lifetime, redistributive) *on top of the corrective layer, R_m satisfies the standard Mirrleesian formula (21).*

In the notation of Proposition 2, conditions 1–2 say that the corrective layer sets the post-tax counterparts of the wedges to one at every age: $\chi_p(a;\theta) \equiv v(\theta)/m_p(a;\theta) = 1$ and $\iota_p \equiv 1$. The construction rests on three supporting lemmas, stated and proved in Appendix A: optimal allocations feature efficient timing and lifetime optimality (Lemma 7); salaries are nondecreasing in the signal, as in the equilibria of Section 2.2, with age-conditional foundations in Lemma 8; and no two labor supply histories generate the same earnings history (Lemma 9), so the government can infer labor supply from tax records and decentralize the allocation. Separation is used in this last step. Recovering the labor supply history then identifies the worker's type, and with it the productivity $v(\theta)$ that the corrective layer targets.

Proof. See Appendix A. □

In words: the corrective layer R_p makes the worker's lifetime reward per marginal deliverable equal to the marginal product *at every age*—it cancels the reputational gains that make $m(a;\theta)$ exceed the current wage, undoing the intertemporal distortions and the lifetime rat race at once—and the redistributive layer then faces a conventional Mirrleesian problem. If a tax system already corrects intertemporal distortions, the economy is back to the environment of Theorem 1; the corollary below records the converse.

Corollary 1. *If the information structure generates no intertemporal distortions (Proposition 4), optimal taxes satisfy* (19).

The pace of the resume. The leading example of a timing-sensitive signal is the *pace*: $I(\tilde{h}, a) = \int_0^a \tilde{h}(\tilde{a})/a \; d\tilde{a}$, the amount of deliverables per unit of time available. This case is natural when heterogeneity is at the intensive margin—every worker starts and ends the career at the same ages, but different types are willing to work at different rates—and employers screen on the rate itself.

Now that we have introduced taxes, we are ready to present the pace case. Its characterization is a property of the equilibrium *jointly with* the corrective taxes that implement it. Without taxes, workers front-load effort to establish high paces early, paces are not constant, and the equilibrium is far less tractable (Remark 1).

Let preferences be time-separable (Assumption 1(b)) with isoelastic disutility,

$$U = \int_0^1 e^{-\rho\tilde{a}} \left(\tilde{c}(\tilde{a}) - \frac{\tilde{h}(\tilde{a})^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \, b(\theta)^{-(1+\frac{1}{\epsilon})} \right) d\tilde{a},$$

discounting $q(a) = e^{-\rho a}$, and the parameterization $v(\theta) = \theta^{\delta}$, $b(\theta) = \theta^{1-\delta}$, $0 < \delta < 1$.

Any constrained-optimal allocation has each worker supplying labor at a constant rate. Focusing for concreteness on the allocation whose redistributive component is zero (each worker's marginal rate of substitution equals their marginal product), more productive types supply higher constant paces; each type reveals a constant pace $\bar{h}$; and, whenever there is separation, pre-tax salaries satisfy a log-linear relationship in the pace,

$$\log w = \gamma \log \bar{h}, \qquad \gamma = \frac{\delta}{1+\epsilon-\delta}, \tag{25}$$

with salaries equal to marginal products at (almost) every instant: $\bar{\chi}(\theta) = 1$. Yet distortions are large. A worker who raises labor supply at age a raises their measured pace at every future age $a' > a$ by $1/a'$, and hence their future salaries: the elasticity of the age-a' salary to age-a effort is γ/a'. The whole distortion is intertemporal, with

$\iota(a;\theta) < 1$ at every interior age, invisible to any comparison of wages and marginal products. Corrective taxes must exactly cancel these reputational gains—and in this example they take a strikingly simple form.

Corollary 2 (Purely age-dependent corrective taxes)**.** *Under the parametric assumptions above, the corrective layer depends only on age:*

$$\frac{dR_p(\tilde{y})}{d\tilde{y}(a)} = a^{\gamma}\left[e^{-\rho} + \rho\int_a^1 s^{-\gamma}e^{-\rho s}\,ds\right], \qquad \gamma = \frac{\delta}{1+\epsilon-\delta}, \tag{26}$$

which reduces to a^{γ} without discounting.

Proof. See Appendix A. □

The corrective retention rate in age-a units, $r_y(a) = e^{\rho a}\,dR_p/d\tilde{y}(a)$, rises from zero to one over the career and stays below one at every interior age (see the proof). Implementing it requires the government to observe age alongside current earnings; tax records contain birthdates, so this is innocuous, and firms' pay, by (3), conditions on the resume alone. Young workers, whose effort carries the largest reputational payoff, face the largest corrective taxes, which fade to zero as the reputational motive disappears. Two features are worth emphasizing. First, implementation requires no information beyond age and current earnings—no histories—so the "history-dependent" corrective layer is, in this case, an age-dependent income tax of the kind that already exists in several tax codes. Second, the *mechanism* generating age dependence is new relative to the age-dependent taxation literature: in Weinzierl [2011] and Farhi and Werning [2013], age dependence tracks the age profiles of elasticities and productivities; here it tracks the age profile of an informational externality—the declining reputational return to effort—and would call for age-dependent taxes even with age-invariant elasticities and productivities.

4.4 Implementation with Simple Instruments

The optimal system of Proposition 6 uses age- and history-dependent instruments, which may seem far removed from actual tax codes. It is useful to collect what the previous results already imply about how much of the correction survives when only simple instruments are available.

In the "length-of-the-resume" case, the answer is that nothing is lost. Under the parametric assumptions of Section 2.3, the Pigouvian component is constant, $1 - \chi = \delta/\alpha$, so the corrective layer is a linear tax on earnings. Linearity is one property that guarantees that lifetime income taxes can be implemented with history-independent annual taxes. Besides linearity, whenever there is heterogeneity only at the extensive margin, lifetime income taxes can be implemented with history-independent non-linear annual taxes. In this benchmark, correcting for dynamic signaling requires no sophistication in the tax code at all.

In the "pace-of-the-resume" case, history dependence is also dispensable, but age dependence is not. Corollary 2 shows that the corrective layer depends only on age and current earnings; expressed in age-a units, its retention rate on age-a earnings is $r_y(a) = e^{\rho a}\, dR_p/d\tilde{y}(a)$, which rises from zero to one over the career. The correction can be implemented exactly with an age-dependent income tax, an instrument of a kind that several tax codes already contain,[16] A purely age-independent tax cannot implement it, since it would either leave young workers overworking or overcorrect older workers, whose reputational motive has largely expired.

To summarize, in both polar cases the corrective component of taxes can be implemented with annual income taxes, history-independent in the length case and age-dependent in the pace case. What we do not quantify is the welfare cost of running the *redistributive* component through the same simple instruments when the assumptions of Proposition 5 fail, or the loss from an age-independent code in the pace case; both are well-posed exercises that we leave for future work.

New monitoring and screening technologies change how workers are paid and how resumes are read. Appendix B.4 shows that, if society values redistribution, decreasing information asymmetries in labor markets decreases welfare (Proposition 12): asymmetries sustain implicit cross-subsidies from high- to low-productivity workers that do part of the planner's work. The impact on taxes is more subtle and is characterized in Appendix B.5.

[16] Age dependence appears in a variety of forms: statutory provisions, such as the additional standard deduction for taxpayers over 65 in the United States; social security and pension rules keyed to age and to earnings histories, including the earnings test; and implicit age variation in effective marginal rates through age-targeted transfers. See Weinzierl [2011] for a survey of such provisions and of the case for explicit age dependence.

5 Magnitudes for the Labor Wedges

The previous sections offered an understanding of the tradeoffs a government faces between incentives and redistribution through the lens of simple sufficient statistics. In the benchmark case where workers' resumes are summarized by the cumulative sum of deliverables they have supplied so far in their careers, the schedule of labor wedges $\chi(y)$ is the single new statistic that the corrective component of taxes requires. In this section we put numbers on that schedule. We take published estimates of the returns to experience, of the share of those returns that can be attributed to signaling, and of the shape of the lifetime income distribution, and we use the free-entry condition (5) to translate them into wedges.

There is documented evidence that workers experience large growth rates of earnings as a function of experience. In fact, Guvenen et al. [2021] have documented that the top 1% lifetime earners have a very steep growth rate of annual earnings, of around 2700% over a 30-year period, or 11.3% per year (and approximately 3% per year on average across workers); the life-cycle wage profiles assembled by Lagakos, Moll, Porzio, Qian, and Schoellman [2018] give similar magnitudes for the average worker. While part of this pattern may be thought of as the result of human capital accumulation, and another part of it may be thought of as the result of pure luck, it is reasonable to expect that at least another part of it is due to selection and the career concerns logic we have uncovered. Moreover, there is evidence that signaling and learning are important to explain the dynamics of salaries and tenure in some occupations: Pan, Wang, and Weisbach [2015] show that the volatility of stock returns declines with the tenure of CEOs of large US firms in the way implied by the market gradually learning about CEO ability, and Kahn and Lange [2014] estimate that employers' uncertainty about worker productivity persists throughout the career, which is what allows experience to keep transmitting information late in working life.

For the purpose of this paper, the key question is determining the share of the growth rate of salaries that is due to signaling and how it can be translated into an estimate of the corrective component of taxes. There is no direct estimate in the literature that can readily be used to answer these questions. To get a sense of what would be reasonable magnitudes for the signaling component of the growth rate of salaries, we can look at the evidence on the return to schooling and the role of signaling in that context. Recent work has concluded that on average 30% of the returns to schooling can be attributed to signaling and 70% to human capital accumulation [Aryal et al., 2022]; the speed of employer learning estimated by Lange [2007] places the signaling share of the returns to schooling below roughly a quarter to a third, and estimates elsewhere in the schooling literature range from the negligible diploma effects of Clark and Martorell [2014] to substantially larger shares. We therefore treat 30% as a central value and vary it widely. Using the free-entry condition (5) under the parametric assumptions of Section 2.3, the return to experience is related to the Pigouvian component of taxes $(1-\chi)$ and the shape of the lifetime income distribution by the formula

$$\gamma = \frac{\alpha_y(1-\chi)}{1-\alpha_y(1-\chi)}, \tag{27}$$

where γ is the elasticity of salaries with respect to experience and α_y is the tail parameter of the lifetime income distribution.[17]

Two inputs map these annual earnings growth rates into (27): the elasticity γ of salaries with respect to experience, and the tail parameter α_y of lifetime income. First, in the length benchmark the resume grows linearly with experience along the common career track, so the elasticity of salaries with respect to the resume at ten years of experience equals ten times the annual growth rate of salaries, giving $\gamma = 0.09$ for the average worker (30% of 3% per year times ten) and $\gamma = 0.34$ at the top (30% of 11.3% times ten).[18] Second, the tail parameter of lifetime income. Both values come from the male lifetime earnings distribution of Guvenen, Kaplan, Song, and Weidner [2022]. For the average worker, the P90/P50 ratio of 2.5 gives $\alpha_y = 1.7$. For top earners, the share of lifetime earnings accruing to the top 1% of men, together with the mean and the median, gives $\alpha_y = 2.4$.[19] Formula (27) is exact in the constant-elasticity economy. We apply it locally, so the two columns of Table 2 are two local calibrations. Table 2 reports the implied Pigouvian component. At the

[17] In the parametric example, $\gamma = \delta/(1-\delta+\epsilon)$, $1-\chi = \delta/\alpha$, and lifetime income is proportional to $\theta^{1+\epsilon}$, so $\alpha_y = \alpha/(1+\epsilon)$; substituting $\delta = (1-\chi)\,\alpha_y(1+\epsilon)$ into γ delivers (27).

[18] Ten years is the horizon within which most of lifecycle wage growth occurs [Topel and Ward, 1992]. The growth rates are of annual earnings, while γ concerns the salary per deliverable; in the common-track benchmark each worker supplies a constant flow, so the two coincide within a career, and we use the empirical earnings profiles as proxies for salary profiles. The mapping is exact when hours and participation are constant over the career.

[19] The top 1% of men earn 8% of cohort lifetime earnings, the mean is \$52,200, and the median is about \$38,000. With a Pareto tail above P90 $\approx$ \$95,000, the value $\alpha_y = 2.4$ implies a 99th percentile of about \$250,000 and average top-1% earnings of about \$420,000, matching the 8% share. For comparison, the annual labor-earnings tail coefficient is about 2 [Lee, Sasaki, Toda, and Wang, 2022, de Vries and Toda, 2022].

central signaling share, the corrective component is of the order of 5% for the average worker and 11% at the top. As the signaling share varies from 15% to 50%, the corrective component for the average worker stays in single digits, and for top earners it goes from 6% to 15%. Since (27) is bounded above by $1/\alpha_y$, even large signaling shares cannot push the externality much further.

Table 2: Calibrated Pigouvian component $1-\chi$

Signaling share of returns to experience	Average worker (3%/yr)	Top earners (11.3%/yr)
15%	0.03	0.06
30%	0.05	0.11
50%	0.08	0.15

Implied by (27), evaluating the elasticity of salaries with respect to experience at ten years, with lifecycle earnings growth from Guvenen et al. [2021], the signaling share centered on Aryal et al. [2022], and $\alpha_y = 1.7$ for the average worker and $\alpha_y = 2.4$ for top earners, both computed from the male lifetime earnings distribution of Guvenen et al. [2022].

A second route to the wedges does not require transporting the signaling share from the schooling context. The wedge at the retirement margin of Section 3 satisfies $1-\chi(\theta) = \big(\mathbb{E}[v \mid h(\tilde{\theta}) \geq h(\theta)] - v(\theta)\big)/\mathbb{E}[v \mid h(\tilde{\theta}) \geq h(\theta)]$: it measures the productivity dispersion remaining among the equally paid workers in the marginal retiree's pool. This is the object the employer-learning literature estimates. Using repeated performance measures for managers of a single large US firm, in the upper part of the earnings distribution, Kahn and Lange [2014] find that the standard deviation of employers' expectation error is about 0.15 in logs for most of the life cycle. With log productivity dispersed $\sigma = 0.15$ within the pool and the marginal retiree z standard deviations below the pool mean, the wedge is $1-\chi = 1-e^{-z\sigma-\sigma^2/2}$, about 1% for a marginal worker at the median of the pool, 11% at the 25th percentile, and 18% at the 10th. Because the marginal retiree's location in the pool is not observed, these values should be read as a sensitivity range rather than as estimates.[20] Under the reported location assumptions, the range overlaps with the returns-based calibration of Table 2, without using the returns to experience. Consistent with this, randomly granting workers detailed public evaluations substantially raised their sub-

[20]Under pure productivity sorting the marginal retiree would be the pool's least productive worker; with preference heterogeneity, retirement selects partly on preferences, so the marginal retiree sits in the interior of the pool's productivity distribution, which is what z parametrizes.

sequent earnings [Pallais, 2014].[21]

An alternative would be to estimate the wedges directly. Adapting to labor markets the logic used to quantify adverse selection in insurance markets [Einav, Finkelstein, and Cullen, 2010], exogenous variation in after-tax wages identifies the externality at the participation margin: when a tax change moves marginal workers out of a pool of workers with equivalent resumes, the salaries of those who remain move with the productivity composition of the pool, and the ratio of the log-salary response to the log-participation response recovers $\chi - 1$. Two identification problems remain. Participation responses are small relative to their sampling error, so the ratio inherits the instability of a weak denominator. And posted wages respond to tax changes through firms' wage-setting power as well as through composition, a force that is plausibly stronger for older workers, whose mobility is lower [Manning, 2003, Topel and Ward, 1992, Card et al., 2018], with direct evidence that taxes move wages through firm behavior [Saez, Schoefer, and Seim, 2019]; the two channels are not separable with the same moments. Developing a design that separates them, and that decomposes participation changes into the gross inflows and outflows that both move the composition of pools, is left for future work.

6 Conclusion

While dynamic signaling is a key feature of labor markets, standard benchmark taxation models often ignore it, or adopt a static perspective on information transmission. In this paper, we developed a simple model that allows job histories and resumes to play this informational role, with firms using them to predict productivities and forward-looking individuals making labor supply decisions that anticipate their impact on future wages. In the model, firms' interest in learning workers' productivity arises even when they can pay for performance and all parties are risk neutral.

We derived generalized Mirrleesian and Pigouvian formulas that apply not only to this model, but more generally to labor market frictions, as long as firms break even and, given how firms set wages, earnings map one-to-one to labor supply choices; the general-signal decomposition requires, in addition, the timing-homogeneity and sep-

[21] Table 2 isolates signaling; monopsony operates in the opposite direction, putting pay below marginal products [Manning, 2003, Card, Cardoso, Heining, and Kline, 2018], and since job mobility declines with age [Topel and Ward, 1992] that friction is plausibly stronger for older workers. The numbers therefore measure the correction that the signaling friction alone would call for.

aration conditions of Proposition 6. The formulas decompose optimal taxes into two components: a redistributive component, and a corrective component that depends on dynamic labor wedges, the ratio of the marginal contribution to output to the sum of current and future benefits the worker receives from supplying one extra unit of labor. The instruments the corrective component requires depend on how firms weigh the timing of past deliverables when reading resumes (Proposition 4), and the two benchmark cases indicate the range. To the extent that firms read resumes as in the length case, taxes on lifetime income suffice; to the extent that the timing of deliverables matters, as in the pace case, the correction is concentrated on the young and fades to zero at retirement, and using the tax system to correct the intertemporal distortions from the dynamic "rat race" could generate pure efficiency gains. Actual labor markets presumably combine elements of both cases.

While a large non-linear income taxation literature has explored and estimated the statistics that appear in the redistributive component of taxes, there is limited work estimating the corrective component, especially under dynamic imperfect information in labor markets. Calibrating the model with published estimates of the returns to experience and of the signaling share of those returns, and reading the results through the length-of-the-resume benchmark, in which lifetime income is the optimal base, the calibration implies that for an average worker the corrective component of taxes is of the order of 5%, while for high earners it ranges from 6% to as high as 15% across signaling shares, with 11% at the central share. An alternative calibration, based on the productivity dispersion, produces comparable magnitudes. It implies that, once we take into account the impact of imperfect information in labor markets, we may conclude that the current tax system is less redistributive than we previously thought.

The calibration transports the signaling share of the returns to schooling to the returns to experience. Direct estimates of the signaling component of experience profiles do not exist yet, and the range of the calibrated wedges is correspondingly wide. Reading the calibrated wedges as a corrective schedule over lifetime income relies on the length benchmark; under signals that front-load reputational returns, as in the pace case, the same retirement-margin numbers would understate the correction and shift it toward earlier ages, and signals with decaying memory could shift it the other way. Finally, the wedges are calibrated rather than estimated; a design based on tax variation must separate the informational composition of pools from firms' wage-setting responses, a task we leave for future work.

References

Daron Acemoglu and Jörn-Steffen Pischke. Why do firms train? theory and evidence. *The Quarterly Journal of Economics*, 113(1):79–119, 1998.

George Akerlof. The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, 90(4):599–617, 1976.

George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

Armen A. Alchian and Harold Demsetz. Production, information costs, and economic organization. *The American Economic Review*, 62(5):777–795, 1972.

Laurence Ales, Musab Kurnaz, and Christopher Sleet. Technical change, wage inequality, and taxes. *American Economic Review*, 105(10):3061–3101, 2015.

Joseph G. Altonji and Charles R. Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.

Joseph G. Altonji, Anthony A. Smith Jr., and Ivan Vidangos. Modeling earnings dynamics. *Econometrica*, 81(4):1395–1454, 2013.

Fredrik Andersson. Income taxation and job-market signaling. *Journal of Public Economics*, 59(2):277–298, 1996.

Kenneth J. Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173, 1962.

Gaurab Aryal, Manudeep Bhuller, and Fabian Lange. Signaling and employer learning with instruments. *American Economic Review*, 112(5):1669–1702, 2022.

Anthony B. Atkinson and Joseph E. Stiglitz. The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 6(1-2):55–75, 1976.

Andrea Attar, Thomas Mariotti, and François Salanié. Nonexclusive competition in the market for lemons. *Econometrica*, 79(6):1869–1918, 2011.

Spencer Bastani, Tomer Blumkin, and Luca Micheletto. Optimal wage redistribution in the presence of adverse selection in the labor market. *Journal of Public Economics*, 131:41–57, 2015.

Felix J. Bierbrauer, Pierre C. Boyer, and Emanuel Hansen. Pareto-improving tax reforms and the earned income tax credit. *Econometrica*, 91(3):1077–1103, 2023.

Alessandro Bonatti and Johannes Hörner. Career concerns with exponential learning. *Theoretical Economics*, 12(1):425–475, 2017.

David Card. Intertemporal labor supply: An assessment. In Christopher A. Sims, editor, *Advances in Econometrics: Sixth World Congress*, volume II, pages 49–78. Cambridge University Press, Cambridge, 1994.

David Card, Ana Rute Cardoso, Jörg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1): S13–S70, 2018.

Cristina Cella, Andrew Ellul, and Nandini Gupta. Learning through a smokescreen: Earnings management and ceo compensation over tenure. Research paper, Riksbank, 2017.

Gonzalo Cisternas. Career concerns and the nature of skills. *American Economic Journal: Microeconomics*, 10(2):152–189, 2018.

Damon Clark and Paco Martorell. The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318, 2014.

Earl A. Coddington and Norman Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

Ashley C. Craig. Optimal income taxation with spillovers from employer learning. *American Economic Journal: Economic Policy*, 15(2):82–125, 2023.

Carlos E. da Costa and Lucas J. Maestri. Optimal Mirrleesian taxation in non-competitive labor markets. *Economic Theory*, 68(4):845–886, 2019.

Tjeerd de Vries and Alexis Akira Toda. Capital and labor income Pareto exponents across time and space. *Review of Income and Wealth*, 68(4):1058–1078, 2022.

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

Peter A. Diamond and James A. Mirrlees. Optimal taxation and public production i: Production efficiency. *The American Economic Review*, 61(1):8–27, 1971.

Avinash Dixit. Tax policy in open economies. In *Handbook of Public Economics*, volume 1, pages 313–374. Elsevier, 1985.

Paweł Doligalski, Abdoulaye Ndiaye, and Nicolas Werquin. Redistribution with performance pay. *Journal of Political Economy Macroeconomics*, 1(2):371–402, 2023.

Liran Einav, Amy Finkelstein, and Mark R. Cullen. Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics*, 125(3): 877–921, 2010.

Henry S. Farber and Robert Gibbons. Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047, 1996.

Emmanuel Farhi and Iván Werning. Insurance and taxation over the life cycle. *The*

Review of Economic Studies, 80(2):596–635, 2013.

Robert Gibbons and Kevin J. Murphy. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, 100(3): 468–505, 1992.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What do data on millions of US workers reveal about lifecycle earnings risk? *Econometrica*, 89(5): 2303–2339, 2021.

Fatih Guvenen, Greg Kaplan, Jae Song, and Justin Weidner. Lifetime earnings in the United States over six decades. *American Economic Journal: Applied Economics*, 14(4):446–479, 2022.

Cyril Hariton and Gwenaël Piaser. When redistribution leads to regressive taxation. *Journal of Public Economic Theory*, 9(4):589–606, 2007.

James J. Heckman. A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy*, 84(4):S9–S44, 1976.

Bengt Holmström. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169–182, 1999.

Johannes Hörner and Nicolas S. Lambert. Motivational ratings. *The Review of Economic Studies*, 88(4):1892–1935, 2021.

Roozbeh Hosseini and Ali Shourideh. Retirement financing: An optimal reform approach. *Econometrica*, 87(4):1205–1265, 2019.

Albert Jan Hummel. Monopsony power, income taxation and welfare. Working paper, University of Amsterdam, 2024.

Susumu Imai and Michael P. Keane. Intertemporal labor supply and human capital accumulation. *International Economic Review*, 45(2):601–641, 2004.

Boyan Jovanovic. Job matching and the theory of turnover. *Journal of Political Economy*, 87(5):972–990, 1979.

Lisa B. Kahn and Fabian Lange. Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *The Review of Economic Studies*, 81(4):1575–1613, 2014.

Greg Kaplan. Inequality and the life cycle. *Quantitative Economics*, 3(3):471–525, 2012.

Wojciech Kopczuk. A note on optimal taxation in the presence of externalities. *Economics Letters*, 80(1):81–86, 2003.

Musab Kurnaz, Martin Michelini, Hakkı Özdenören, and Christopher Sleet. Tax design for the long run. *Journal of Political Economy Macroeconomics*, 4(1), 2026.

David Lagakos, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849, 2018.

Renée M. Landers, James B. Rebitzer, and Lowell J. Taylor. Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, 86(3):329–348, 1996.

Fabian Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1): 1–35, 2007.

Ji Hyung Lee, Yuya Sasaki, Alexis Akira Toda, and Yulong Wang. Capital and labor income Pareto exponents in the United States, 1916–2019. Working paper, arXiv:2206.04257, 2022.

Alan Manning. *Monopsony in Motion: Imperfect Competition in Labor Markets.* Princeton University Press, Princeton, NJ, 2003.

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

Hajime Miyazaki. The rat race and internal labor markets. *The Bell Journal of Economics*, 8(2):394–418, 1977.

Amanda Pallais. Inefficient hiring in entry-level labor markets. *American Economic Review*, 104(11):3565–3599, 2014.

Yihui Pan, Tracy Yue Wang, and Michael S. Weisbach. Learning about CEO ability and stock return volatility. *Review of Financial Studies*, 28(6):1623–1666, 2015.

Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva. Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–271, 2014.

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

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

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

Emmanuel Saez, Joel Slemrod, and Seth H. Giertz. The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50, 2012.

Emmanuel Saez, Benjamin Schoefer, and David Seim. Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in sweden. *American Economic Review*, 109(5):1717–1763, 2019.

Agnar Sandmo. Optimal taxation in the presence of externalities. *The Swedish Journal of Economics*, 77(1):86–98, 1975.

Florian Scheuer and Iván Werning. Mirrlees meets Diamond-Mirrlees. Working Paper 22076, NBER, 2016.

Florian Scheuer and Iván Werning. The taxation of superstars. *The Quarterly Journal of Economics*, 132(1):211–270, 2017.

Jesús Seade. On the sign of the optimum marginal income tax. *The Review of Economic Studies*, 49(4):637–643, 1982.

Stefanie Stantcheva. Optimal income taxation with adverse selection in the labour market. *Review of Economic Studies*, 81(3):1296–1329, 2014.

Robert H. Topel and Michael P. Ward. Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479, 1992.

Donald M. Topkis. *Supermodularity and Complementarity.* Princeton University Press, Princeton, NJ, 1998.

William Vickrey. *Agenda for Progressive Taxation.* Ronald Press, New York, 1947.

Matthew Weinzierl. The surprising power of age-dependent taxes. *The Review of Economic Studies*, 78(4):1490–1518, 2011.

Iván Werning. An elementary proof of positive optimal marginal taxes. mimeo, MIT, 2000.

Iván Werning. Pareto efficient income taxation. mimeo, MIT, 2007a.

Iván Werning. Optimal fiscal policy with redistribution. *The Quarterly Journal of Economics*, 122(3):925–967, 2007b.

A Proofs

A.1 Salaries Under the Length Case (Proof of Lemma 1)

Proof. First, notice that for every type who supplies $h(\theta) > h$ over their lifetime, there is at any date someone (potentially from a different cohort) who has experience exactly h today. Moreover, with $q \equiv 1$, every such type supplies exactly dh deliverables while its resume passes from h to $h + dh$, regardless of the age at which it does so, so in the stationary cross-section the deliverables executed at length h are distributed across the tail types with the population weights. Thus $\mathbb{E}[v(\theta) \mid h(a, \theta) = h] = \mathbb{E}[v(\theta) \mid h(\theta) \geq h]$. Now suppose a firm sets salaries not equal to this conditional expectation, so that (5) is violated for some h. Then a firm offering a contract conditional on experience h makes either losses or strictly positive profits. In the first case the firm is better off not offering the contract; in the second, an entrant can attract the same deliverables by offering a slightly higher salary, still below the conditional expected productivity, and make strictly positive profits, so the posited salary cannot survive entry. □

A.2 Unraveling Under Full-History Observation

Lemma 3 (Unraveling under full-history observation)**.** *Let $I(\tilde{h}, a, \theta) = \tilde{h}^a$, let θ be single-dimensional, and suppose equilibrium policies are strictly monotone in the type: for every age $a > 0$, if $\theta' > \theta$ then $\tilde{h}_{\theta'}$ exceeds $\tilde{h}_\theta$ on a positive-measure subset of $[0, a]$ (which holds under Assumption 1 with strict single crossing and $\partial^2 H / \partial\tilde{h}(a)\partial\tilde{h}(a') < 0$). Then in any separating equilibrium in which salaries vary smoothly with histories along equilibrium paths, salaries reveal the type of almost every worker at every age $a > 0$, and wages equal marginal products almost everywhere: $w(\tilde{h}^a_\theta) = v(\theta)$ for a.e. θ and every $a > 0$.*

Proof. Fix $a > 0$. By strict monotonicity of equilibrium policies, if $\theta' > \theta$ then $\tilde{h}_{\theta'}$ and $\tilde{h}_\theta$ differ on a positive-measure subset of $[0, a]$, so the map $\theta \mapsto \tilde{h}^a_\theta$ is injective on the support of types. In a separating equilibrium—one in which expected productivities conditional on observed histories are determined on path by (3)—firms therefore invert the history: $\mathbb{E}[v(\tilde{\theta}) \mid \tilde{h}^a_\theta] = v(\theta)$, and $w(\tilde{h}^a_\theta) = v(\theta)$ for a.e. θ and every $a > 0$: the comparison of wages with productivities detects no distortion. The allocation is nonetheless distorted wherever v is strictly increasing. Producing the equilibrium history of another type is an on-path deviation with known pay: type θ producing $\tilde{h}_{\theta'}$ receives $v(\theta')$ on every deliverable, for lifetime pay $Y(\theta') = v(\theta')\, h(\theta')$, with $h(\theta') = \int_0^1 q\, \tilde{h}_{\theta'}$, at disutility aggregate $H(\tilde{h}_{\theta'})$. In the absence of taxes, or reading Y as pre-tax lifetime pay so that the calculation isolates the signaling component, incentive compatibility along this menu gives, at $\theta' = \theta$, $MRS = Y'/H' = (v'h + vh')/H'$; when the disutility aggregate is the discounted sum, $H = h$, this is $v + v'h/h' > v$. Hence the pre-tax lifetime return to labor exceeds the marginal product, and the lifetime wedge along the menu is $v/(v + v'h/h') < 1$. In this separating equilibrium there is no pooling by hypothesis, and yet the rat race remains. Note that a perturbation at a single age takes the worker off the menu, and its pay depends on how salaries extend to off-path histories.

Two remarks delimit what this argument establishes. First, strict monotonicity of policies follows, by the argument of Lemma 8, from Assumption 1 with strict single crossing and $\partial^2 H/\partial\tilde{h}(a)\partial\tilde{h}(a') < 0$. Smoothness of salaries along equilibrium paths holds in the equilibria constructed in Proposition 7, since there the allocations solve a system of ordinary differential equations with continuous coefficients, and w is then differentiable in the type and hence along the path. Second, the statement concerns revelation on the equilibrium path. To assign wages to off-path histories one needs to specify beliefs, and we do not rule out pooling equilibria sustained by off-path beliefs. Appendix B.1.3 analyzes the full-history case with a type space rich enough that all histories are on path, where no belief specification is needed. □

A.3 Minimal Ingredients (Proof of Proposition 3)

Proof. (i) By hypothesis, the salary gradient is strictly positive, $w' > 0$ almost everywhere along the equilibrium path. The weights satisfy $\varphi > 0$ and $\partial I(\tilde{a})/\partial\tilde{h}(a) = \varphi(a, \tilde{a}) > 0$ for $\tilde{a} > a$, so every term of the reputational integral in (11) is nonnegative, and strictly positive whenever $\tilde{h}(\tilde{a}) > 0$ on a positive-measure subset of $(a, 1]$. Hence $m(a; \theta) > w(I(a))$.

(ii) We argue for a worker active throughout, $T = 1$; with early retirement, replace 1 by $T(\theta)$ throughout, using that the reputational term vanishes after $T(\theta)$. At $a = 1$ the reputational term vanishes and $m(1; \theta) = w(I_\theta(1))$. A less productive type never produces the value $I_\theta(1)$, since its signal path lies below I_θ at every age and, being nondecreasing, never exceeds its own terminal value, which lies strictly below $I_\theta(1)$.

The pool at $I_\theta(1)$ therefore consists of type θ and of strictly more productive types, the latter present with positive probability by the pooling hypothesis, so $w(I_\theta(1)) = \mathbb{E}[v \mid I_\theta(1)] > v(\theta)$ and $\bar{\chi}(\theta) = v(\theta)/m(1;\theta) < 1$. In the length case resumes are nondecreasing and every longer resume passes through the worker's final length, so the pooling hypothesis holds for every type whose final length is exceeded with positive probability; in the pace case each index value identifies the type, the hypothesis fails, and $\bar{\chi}(\theta) = 1$, consistent with Section 3.2.

(iii) Independence: if $v(\theta)$ is independent of the preference component of the type, then for any signal realization $\mathbb{E}[v \mid I] = \mathbb{E}[v]$, so w is constant, $w' = 0$, the reputational term vanishes at every age, $m(a;\theta) = \mathbb{E}[v]$ for all a, and $\iota \equiv 1$. The wedge $\chi(a;\theta) = v(\theta)/\mathbb{E}[v]$ differs from one type by type but carries no signaling return: it is a pure pooling cross-subsidy, and in particular it averages to one across workers. Full history: by Lemma 3, wages equal marginal products on path, while the menu argument in its proof shows the lifetime return to labor exceeds the marginal product wherever v is strictly increasing in the type, so the pooling vanishes while the rat race remains. $\square$

A.4 Absence of Intertemporal Distortions (Proof of Proposition 4)

Definition 2 (Rich class)**.** *A class of economies $\mathcal{E}$ is* rich *if it contains the extensive-margin family: economies in which types are indexed by a retirement age T, workers supply $\tilde{h} = 1$ until T and zero afterward, and productivity is assigned by a function $v = \nu(T)$ with $\nu : [a_0, 1] \to [\underline{v}, \bar{v}]$ strictly increasing and continuously differentiable (so that more productive types are more willing to supply deliverables, as Assumption 1 requires); the family ranges over all such assignments ν and all continuous distributions of T with full support on $[a_0, 1]$, for every $a_0 \in (0, 1)$.*

Within the extensive-margin family, every worker still active supplies the same flow $\tilde{h} \equiv 1$, so the resume path $I(\tilde{a})$ and its drift are common across these economies and across active workers; only the equilibrium wage function varies. The following lemma makes precise which wage profiles the family generates; it is the only property of $\mathcal{E}$ that the necessity argument uses. The construction requires the common resume path to be strictly increasing, so that the resume identifies age along the path; this is the cumulative-weight condition added in the necessity part of Proposition 4. The pace kernel violates it, but the pace also violates the hypothesis $\iota \equiv 1$, so nothing is lost.

Lemma 4 (Exact generation of wage profiles)**.** *Assume the common resume path $I(\tilde{a}) = \int_0^{\tilde{a}} \varphi(z, \tilde{a})\, dz$ has a continuous and strictly positive derivative in $\tilde{a}$. Fix $a_0 \in (0, 1)$ and let $g : [a_0, 1] \to (\underline{v}, \bar{v}]$ be continuously differentiable with $g' > 0$. Then there is an economy in the extensive-margin family whose equilibrium wage along the common resume path is exactly $w(I(\tilde{a})) = g(\tilde{a})$ for every $\tilde{a} \in [a_0, 1)$.*

Proof. Let $\varepsilon(s) = \varepsilon_0(1 - s)$ with $\varepsilon_0 > 0$ small enough that $g - \varepsilon > \underline{v}$ on $[a_0, 1]$, and set $\nu = g - \varepsilon$. Then $\nu' = g' + \varepsilon_0 > 0$ and ν takes values in $(\underline{v}, \bar{v}]$, so ν is an admissible assignment. Give the retirement age T the hazard rate $\lambda(s) = g'(s)/\varepsilon(s)$ on $[a_0, 1)$. The hazard is continuous and strictly positive, and $\int_{a_0}^1 \lambda(s)\, ds = \infty$, because g' attains a positive minimum on the compact interval while ε vanishes linearly at 1. Hence the survival function $S(s) = \exp(-\int_{a_0}^s \lambda)$ defines a continuous distribution for T with full support on $[a_0, 1]$ and $S(1) = 0$. The equilibrium wage at age $\tilde{a}$ is the tail mean $M(\tilde{a}) = \mathbb{E}[\nu(T) \mid T \geq \tilde{a}]$, which satisfies $M' = \lambda(M - \nu)$; the target satisfies the same equation, since $g' = \lambda\varepsilon = \lambda(g - \nu)$. Therefore

$$\frac{d}{d\tilde{a}}\Big[S\,(g - M)\Big] = -\lambda S(g - M) + S\lambda(g - \nu) - S\lambda(M - \nu) = 0,$$

so $S \cdot (g - M)$ is constant. Since $g - M$ is bounded and $S(\tilde{a}) \to 0$ as $\tilde{a} \to 1$, the constant is zero, and $M = g$ wherever $S > 0$, that is, on $[a_0, 1)$. The profile is generated exactly, not approximately. $\square$

Positive perturbations of admissible targets remain admissible, so the wage gradients realized along the common path include all continuous strictly positive functions on $[a_0, 1]$ of sufficiently small amplitude, and differences of two such gradients span the continuous functions; this is what the variational step below requires. Since the drift $\dot{I}$ is continuous and strictly positive, the induced salary schedule has gradient $w'(I(t)) = g'(t)/\dot{I}(t)$, continuous and strictly positive, so it is locally Lipschitz in the signal and admissible.

Proof. (i) Sufficiency. Let $\varphi(\tilde{a}, a) = c\, q(\tilde{a})$, so $I(a) = c\, h(a)$ and equilibrium wages can be written as a function of $h(a)$. For the reputational term in (11), using

$\partial h(\tilde a)/\partial \tilde h(a) = q(a)$ for $\tilde a > a$ and $dh(\tilde a) = q(\tilde a)\tilde h(\tilde a)\, d\tilde a$,

$$\int_a^1 q(\tilde a)\, w'(h(\tilde a))\, \frac{\partial h(\tilde a)}{\partial \tilde h(a)}\, \tilde h(\tilde a)\, d\tilde a = q(a) \int_{h(a)}^{h(1)} w'(z)\, dz = q(a)\, \big[w(h(1)) - w(h(a))\big].$$

Hence $m(a;\theta) = w(h(a)) + w(h(1)) - w(h(a)) = w(h(\theta))$, independent of a, so $\iota \equiv 1$ in every economy.

(ii) Necessity. Suppose $\iota \equiv 1$ in every economy of a rich class $\mathcal{E}$. Fix an economy and a type with $\tilde h > 0$; constancy of $m(\cdot;\theta)$ and $m(1;\theta) = w(I(1))$ give, for every a,

$$\frac{1}{q(a)} \int_a^1 q(\tilde a)\, w'(I(\tilde a))\, \varphi(a,\tilde a)\, \tilde h(\tilde a)\, d\tilde a = w(I(1)) - w(I(a)) = \int_a^1 w'(I(\tilde a))\, \dot I(\tilde a)\, d\tilde a, \tag{28}$$

where $\dot I(\tilde a) = \varphi(\tilde a,\tilde a)\tilde h(\tilde a) + \int_0^{\tilde a} \partial_2 \varphi(z,\tilde a)\, \tilde h(z)\, dz$ is the equilibrium drift of the resume. Collecting terms,

$$\int_a^1 w'(I(\tilde a)) \left[\frac{q(\tilde a)\, \varphi(a,\tilde a)}{q(a)}\, \tilde h(\tilde a) - \dot I(\tilde a)\right] d\tilde a = 0 \qquad \text{for all } a.$$

By richness, evaluate the display within the extensive-margin family for a type active through $[a,1]$: the bracket is fixed across these economies, while by Lemma 4 the display holds for every admissible wage profile, and differences of the associated gradients span the continuous functions on $[a,1]$. By the fundamental lemma of the calculus of variations the bracket must vanish for almost every $\tilde a \geq a$:

$$\frac{q(\tilde a)\, \varphi(a,\tilde a)}{q(a)}\, \tilde h(\tilde a) = \dot I(\tilde a) \qquad \text{for a.e. } \tilde a \geq a. \tag{29}$$

The right-hand side does not depend on a; hence $\varphi(a,\tilde a)/q(a)$ is independent of a on $a \leq \tilde a$, so $\varphi(a,\tilde a) = q(a)\, k(\tilde a)$ for some function $k > 0$: the weight on an age-a deliverable must be proportional to its discount factor. Substituting back, $I(\tilde a) = k(\tilde a)h(\tilde a)$ and $\dot I(\tilde a) = k'(\tilde a)h(\tilde a) + k(\tilde a)q(\tilde a)\tilde h(\tilde a)$, while (29) requires $\dot I(\tilde a) = k(\tilde a)q(\tilde a)\tilde h(\tilde a)$; therefore $k'(\tilde a)h(\tilde a) = 0$, and since $h > 0$ at interior ages, k is constant: $\varphi(a,\tilde a) = c\, q(a)$. □

A.5 Lifetime Income Base (Proof of Proposition 5)

Proof. By the taxation principle, the equilibrium outcome of any anonymous tax system is an incentive compatible, resource-feasible allocation and, under the selection maintained in Section 2.2, any such allocation can be implemented. Fix an optimal allocation, and keep each type's pair (C_θ, H_θ). Replace each type's flows by the solutions of

$$e(C) = \min_{\tilde{c}} \int_0^1 \tilde{c}(a)\, da \quad \text{s.t.} \quad C(\tilde{c}) \geq C, \tag{30}$$

$$e_H(H) = \max_{\tilde{h}} \int_0^1 \tilde{h}(a)\, da \quad \text{s.t.} \quad H(\tilde{h}) \leq H, \tag{31}$$

the least-cost consumption path attaining C_θ and the labor path maximizing lifetime labor at disutility H_θ ($q \equiv 1$ in the length case). With C and H strictly increasing, both values are strictly increasing; we take the constraint sets closed, so the extrema are attained (failing that, an ε-approximation delivers the same conclusions). By Assumption 1(a), utilities depend on the flows only through (C, H), so the utility of every report is unchanged and the transformed allocation is incentive compatible. Resources are weakly saved type by type: expenditure at each C_θ is minimized and, since production in the length case is $v(\theta)$ times lifetime labor, the labor-maximal path at each H_θ is also the production-maximal path. The transformed allocation is therefore feasible and attains the same welfare as the optimum; it is itself optimal.

It remains to implement it with a tax on lifetime income alone. Compute the equilibrium salaries generated by the transformed allocation: with $q \equiv 1$, the pool at length z consists of the types whose lifetime labor $h(\theta) = e_H(H_\theta)$ weakly exceeds z, with the stationary weights of Lemma 1, so $w(z) = \mathbb{E}[v \mid h(\theta) \geq z]$, and lifetime income $y(h) = \int_0^h w(z)\, dz$ is strictly increasing, since salaries are bounded below by $\underline{v} > 0$. Because e_H is strictly increasing, equal lifetime labor implies equal H and, by incentive compatibility, equal consumption expenditure $e(C_\theta)$. The menu $\{(e(C_\theta), h(\theta))\}$ therefore defines a function of lifetime income, $R(y(h(\theta))) = e(C_\theta)$, completed confiscatorily off the menu. Facing R, each worker chooses within the menu; incentive compatibility delivers the assigned pair and, given (C, H), workers are indifferent over timing, so the labor-maximal flows can be selected. The salaries above are computed from the transformed allocation itself, so they are consistent with these choices. The transformed optimal allocation is thus the equilibrium outcome of a tax system that conditions only on lifetime income. $\square$

A.6 Lifetime Discounted Aggregates

Lemma 5 (Lifetime discounted aggregates)**.** *Suppose Assumption 1(a) holds with C and H continuous and strictly increasing, the discount rates q are positive and common across workers, and lifetime post-tax earnings depend on the labor supply flows only through the discounted lifetime labor supply $\int_0^1 q(a)\tilde{h}(a)\,da$, as in the "length-of-the-resume" case with lifetime income taxes. Then there exists $\hat{U}(\cdot,\cdot;\theta)$, strictly increasing in its first argument and strictly decreasing in its second, such that the worker's optimal flows maximize*

$$\hat{U}\left(\int_0^1 q(a)\,\tilde{c}(a)\,da,\ \int_0^1 q(a)\,\tilde{h}(a)\,da;\ \theta\right)$$

subject to the lifetime budget constraint.

Proof. Write $x = \int_0^1 q(a)\tilde{c}(a)\,da$ for lifetime discounted consumption and $y = \int_0^1 q(a)\tilde{h}(a)\,da$ for lifetime discounted labor supply. Define

$$C^*(x) = \max_{\tilde{c}} \left\{ C(\tilde{c}) : \int_0^1 q(a)\tilde{c}(a)\,da \le x \right\}, \qquad H^*(y) = \min_{\tilde{h}} \left\{ H(\tilde{h}) : \int_0^1 q(a)\tilde{h}(a)\,da = y \right\}.$$

Both are well defined under continuity of the aggregators and $q > 0$; C^* is strictly increasing in x and H^* is increasing in y, by strict monotonicity of C and H in the flows. Set $\hat{U}(x, y; \theta) \equiv U(C^*(x), H^*(y), \theta)$, strictly increasing in x and strictly decreasing in y.

Consider any feasible plan $(\tilde{c}, \tilde{h})$ with aggregates (x, y). Because lifetime post-tax earnings depend on the labor flows only through y, replacing $\tilde{h}$ by the minimizer in the definition of $H^*(y)$ leaves the budget constraint unaffected and weakly raises utility; and replacing $\tilde{c}$ by the maximizer in the definition of $C^*(x)$ leaves discounted expenditure unchanged and weakly raises utility. Hence for every (x, y) attainable within the budget, the best attainable utility is $\hat{U}(x, y; \theta)$, and the worker's problem reduces to choosing (x, y) to maximize $\hat{U}$ subject to the lifetime budget constraint. Any optimal plan attains the optimal (x, y) with flows solving the two subproblems. $\square$

A.7 Reduction of Time-Separable Preferences (Lemma 6)

Lemma 6 (Reduction of case (b) to case (a))**.** *Let preferences be time separable with flow utility additively separable between consumption and labor,*

$$U = \int_0^1 e^{-\rho a}\big[m(\tilde{c}(a);\theta) - g(\tilde{h}(a);\theta)\big]\,da,$$

with the discount rate ρ common across workers. If both felicities are increasing affine transformations of common functions,

$$m(\tilde{c};\theta) = \gamma(\theta)\,m_0(\tilde{c}) + \kappa(\theta), \qquad g(\tilde{h};\theta) = \alpha(\theta)\,g_0(\tilde{h}) + \beta(\theta), \tag{32}$$

with $\gamma(\theta) > 0$ and $\alpha(\theta) > 0$, then preferences satisfy Assumption 1(a).

Proof. Under (32), define the common aggregators $C(\tilde{c}) = \int_0^1 e^{-\rho a} m_0(\tilde{c}(a))\,da$ and $H(\tilde{h}) = \int_0^1 e^{-\rho a} g_0(\tilde{h}(a))\,da$. Then $U = \gamma(\theta)C - \alpha(\theta)H + (\kappa(\theta) - \beta(\theta))\int_0^1 e^{-\rho a}da$, a function of (C, H, θ), strictly increasing in C and strictly decreasing in H. □

The conditions cover the parametric example of Section 4.3, where $g_0(\tilde{h}) = \tilde{h}^{1+1/\epsilon}/(1+1/\epsilon)$ and $\alpha(\theta) = b(\theta)^{-(1+1/\epsilon)}$.

A.8 Solving in Post-Tax Space (Proof of Lemma 2)

Proof. Because $y'(h) = \mathbb{E}[v \mid h(\theta) \geq h]$ is a well-defined, strictly positive function of the allocation, $y(h)$ exists and is strictly increasing, so the inverse $h^{-1}(y)$ exists. Given any solution $\tilde{R}(h)$ of problem (18), define $R(y)$ by $R(y(h)) = \tilde{R}(h)$: this recovers the income tax schedule and the equilibrium salaries prevailing in the economy where the planner solved for $\tilde{R}$ directly. Conversely any $R(y)$ induces $\tilde{R}(h) = R(y(h))$ attaining the same allocation and the same resource cost, since firms break even: $\mathbb{E}[y(h(\theta))] = \mathbb{E}[v(\theta)h(\theta)]$ in equilibrium. □

A.9 Optimal Tax Formula (Proof of Theorem 1)

Proof. By Lemma 2, derive the necessary condition for the problem in post-tax space,

$$\max_{\tilde{R}}\ \mathbb{E}[W(V(\tilde{R};\theta),\theta)] \quad \text{s.t.} \quad \mathbb{E}[v(\theta)h(\tilde{R},\theta) - \tilde{R}(h(\tilde{R},\theta))] \geq 0.$$

Let $\mathcal{L}(\tilde{R}) = \mathbb{E}[W(V(\tilde{R};\theta),\theta)] + \mu\,\mathbb{E}[v(\theta)h(\tilde{R},\theta) - \tilde{R}(h(\tilde{R},\theta))]$, with μ the multiplier on resources. Consider a one-bracket variation $\tilde{R}^* = \tilde{R} + \epsilon\Delta$ with

$$\Delta(x,\ell) = \begin{cases} 0 & x \le h \\ x - h & h \le x \le h + \ell \\ \ell & x \ge h + \ell, \end{cases}$$

and define, for any functional $\mathcal{F}$, $\frac{d\mathcal{F}}{d\tilde{r}_h} \equiv \lim_{\ell\to 0^+} \frac{1}{\ell}\frac{d}{d\epsilon}\Big|_{\epsilon=0}\mathcal{F}(\tilde{R}^*)$ and $\frac{d\mathcal{F}}{dI} \equiv \frac{d}{d\epsilon}\Big|_{\epsilon=0}\mathcal{F}(\tilde{R} + \epsilon\Delta_I)$ with $\Delta_I \equiv 1$ (a uniform transfer). Optimality requires $d\mathcal{L}/d\tilde{r}_h = 0$, that is, denoting $\lambda(\theta) = \frac{dW(V,\theta)}{dV}$ and assuming lifetime labor responds at the intensive margin,

$$\mathbb{E}\left[\lambda(\theta)\,\frac{dV(\tilde{R},\theta)}{d\tilde{r}_h}\right] = -\mu\,\mathbb{E}\left[v(\theta)\,\frac{dh(\theta)}{d\tilde{r}_h} - \tilde{r}(h(\theta))\,\frac{dh(\theta)}{d\tilde{r}_h} - \mathbf{1}\big(h(\theta) \ge h\big)\right]. \tag{33}$$

The limit $\ell \to 0^+$ separates the behavioral responses by the worker's position relative to the bracket. A worker with $h(\theta) > h$ keeps an unchanged marginal retention rate and receives the transfer $\epsilon\ell$: for these workers the perturbation converges to a uniform transfer, so $\frac{dh(\theta)}{d\tilde{r}_h} = \frac{dh(\theta)}{dI}$ and, by the envelope theorem, $\frac{dV}{d\tilde{r}_h} = \frac{dV}{dI}$. Workers in the bracket have mass $f(h)\,\ell + o(\ell)$, where f is the density of lifetime labor. For workers in the bracket, the perturbation changes the marginal retention rate by ϵ and grants a mechanical transfer of order ℓ. By the Slutsky decomposition, as $\ell \to 0$ their labor response converges to the *compensated* response $\frac{dh^c}{d\tilde{r}_h}$, that is, the constant-utility response of lifetime labor to its own marginal retention rate, because the gap between the two responses is an income effect scaled by a transfer that vanishes. Collecting the two groups, dividing by μ, and writing

$$\lambda(\tilde{h}) \equiv \frac{1}{\mu}\,\frac{dW(V,\theta)}{dV}\,\frac{dV(\tilde{R},\theta)}{dI} \qquad \text{evaluated at the type supplying } \tilde{h},$$

for the normalized welfare weight, the marginal social value of a unit of lifetime income to that worker in units of public funds, (33) becomes

$$\big(v(h) - \tilde{r}(h)\big)f(h)\,\frac{dh^c}{d\tilde{r}_h} = \int_h^\infty f(\tilde{h})\big(1 - \lambda(\tilde{h})\big)\,d\tilde{h} + \int_h^\infty \big(v(\tilde{h}) - \tilde{r}(\tilde{h})\big)f(\tilde{h})\,\frac{d\tilde{h}}{dI}\,d\tilde{h}, \tag{34}$$

where f is the density of lifetime labor and $v(h)$ the productivity of the worker supplying h. In elasticity form, with $\epsilon^c_{\tilde{r}}(h) \equiv \frac{dh^c}{d\tilde{r}_h}\frac{\tilde{r}(h)}{h}$ and $\eta_I(h) \equiv \frac{dh}{dI}\,\tilde{r}(h)$:

$$\left(\frac{v(h)-\tilde{r}(h)}{\tilde{r}(h)}\right) f(h)\, h\, \epsilon^c_{\tilde{r}}(h) = \int_h^\infty f\,(1-\lambda)\, d\tilde{h} + \int_h^\infty \frac{v-\tilde{r}}{\tilde{r}}\, f\, \eta_I\, d\tilde{h}. \tag{35}$$

Finally, convert to lifetime income. Marginal retention satisfies $\tilde{r}(h) = r(y(h))\, y'(h)$; densities, $f(h) = g(y(h))\, y'(h)$; behavioral responses, $\frac{dy}{dr} = y'(h)^2 \frac{dh}{d\tilde{r}}$ and $\frac{dy}{dI} = y'(h)\frac{dh}{dI}$; and $dh = dy/y'(h)$. Substituting into (35) and writing $\chi(y) \equiv v(y)/y'(h(y))$ delivers (19). □

A.10 Corrective Layers (Proof of Proposition 6)

The proof proceeds through three lemmas: optimal allocations feature efficient timing and lifetime optimality (Lemma 7); salaries are increasing in the signal (Lemma 8); and earnings histories reveal labor supply histories (Lemma 9), which permits decentralization.

Lemma 7 (Efficient timing and lifetime optimality)**.** *Under Assumption 1(a), any optimal incentive compatible allocation satisfies: (i) for any H and C, the flows solve*

$$\tilde{h}(\cdot; H) = \arg\max_{\tilde{h}} \int_0^1 q(a)\tilde{h}(a)\, da \quad s.t. \quad H(\tilde{h}) = H,$$
$$\tilde{c}(\cdot; C) = \arg\min_{\tilde{c}} \int_0^1 q(a)\tilde{c}(a)\, da \quad s.t. \quad C(\tilde{c}) = C;$$

and (ii) the Mirrleesian condition (35) holds in terms of $e_H = \max\{\int q\tilde{h} : H(\tilde{h}) = H\}$ and $\tilde{R}(e_H) = e_C$.

Proof. (i) follows from an argument analogous to the production-efficiency theorem of Diamond and Mirrlees [1971]. Consider the planning problem over flows subject to resources and incentive compatibility in (C, H). Fix any candidate solution and replace its flows with the solutions of the two subproblems above, holding each worker's (C_θ, H_θ) fixed: incentive compatibility is unaffected (utilities depend on flows only through the aggregates) while resources are saved. If the candidate did not already solve the subproblems, the resource constraint becomes slack, and a uniform increase of the implicit retention map $\tilde{R}$ produces a Pareto improvement—contradicting optimality. (ii) then follows from the variational argument in the proof of Theorem 1, applied to the problem written in terms of e_H, assuming no bunching at the optimum. $\square$

Lemma 8 (Salaries and the signal, conditional on age)**.** *Assume that (i) the assignment of labor supply paths to lifetime aggregates is pointwise monotone, so that paths attaining a higher H have weakly higher $\tilde{h}(a)$ at every age, and (ii) $MRS(C, H, \theta) \equiv -U_H/U_C$ is strictly decreasing in θ. Then, for $I = \int_0^a \varphi(\tilde{a}, a)\tilde{h}(\tilde{a})\, d\tilde{a}$ with $\varphi \geq 0$, and conditional on any age $a > 0$, the expected productivity $\mathbb{E}[v \mid I, a]$ is nondecreasing in I, strictly increasing where the signal strictly increases in the type.*

Proof. By (ii), for any strictly increasing $\tilde{R}(H)$—a property of the optimal allocation by Lemma 7—higher types choose higher H. By (i), higher H implies weakly higher $\tilde{h}(a)$ at every age, so conditional on age the signal is nondecreasing in the type, and the conditional expectation of v given the signal and age is nondecreasing in the signal. $\square$

Two comments are in order. Condition (i) is a comparative-statics property of the common subproblem of choosing the path that maximizes lifetime pay at a given aggregate; supermodularity of lifetime pay in the flows together with submodularity of the aggregator delivers it by standard monotone comparative statics [Topkis, 1998]. And monotonicity conditional on age aggregates to monotonicity of $w(I)$ whenever the age composition of the pool does not vary adversely with the signal. In the two benchmark cases it holds directly, in the length case by Lemma 1 and in the pace case because salaries equal marginal products; for general signals it is part of the equilibrium selection of Section 2.2, which focuses on equilibria with salaries nondecreasing in the signal.

Lemma 9 (Earnings reveal labor supply)**.** *Suppose the weights are bounded by $\bar{\varphi}$, labor supply histories are bounded by $\bar{h}$, and salaries, as a function of the signal, are bounded by $\bar{w}$, bounded below by $w_{\min} > 0$, and Lipschitz with constant L. Then no two continuous labor supply histories $\tilde{h}$ and $\tilde{h}'$ map into the same earnings history $\tilde{y}$.*

Proof. The idea is that earnings at each age equal current labor times a salary that depends only on past labor, so the history can be recovered forward from age zero, dividing earnings by the salary implied by the already-recovered past; the bounds make the recovery unambiguous. Formally, suppose $\tilde{h}$ and $\tilde{h}'$ generate the same earnings history $\tilde{y}$. Since salaries are bounded away from zero, $\tilde{h}(a) = \tilde{y}(a)/w(I(\tilde{h}^a))$, and likewise for $\tilde h'$. Write $D(a) = |\tilde h(a) - \tilde h'(a)|$ for the gap between the two histories. The two signals differ by at most $\bar\varphi \int_0^a D(\tilde a)\, d\tilde a$, and hence the two salaries by at most L times that amount. Using $\tilde y(a) \le \bar h \bar w$,

$$D(a) = \tilde y(a)\, \frac{\big|w(I(\tilde h^a)) - w(I(\tilde h'^a))\big|}{w(I(\tilde h^a))\, w(I(\tilde h'^a))} \;\le\; K \int_0^a D(\tilde a)\, d\tilde a, \qquad K \equiv \frac{\bar h \bar w L \bar\varphi}{w_{\min}^2}.$$

A continuous gap with this property is zero. On the interval $[0, 1/(2K)]$, let M be its maximum; the inequality gives $D(a) \le KaM \le M/2$ at every a in the interval, so $M \le M/2$ and $M = 0$. Repeating the argument on successive intervals of length $1/(2K)$ covers the career, so $D \equiv 0$ and the two histories coincide. □

Proof of Proposition 6. By Lemmas 7–9, the planner can solve for optimal post-tax wages $\tilde R$ in the problem

$$\max_{\tilde R}\; \mathbb{E}[W(V(\tilde R, \theta), \theta)] \quad \text{s.t.} \quad \mathbb{E}\Big[\int_0^1 v(\theta) q(a) \tilde h(a; \tilde R, \theta)\, da - \tilde R\big(\tilde h(\cdot; \tilde R, \theta)\big)\Big] = 0,$$

and decentralize it with history-dependent earnings taxes, as if a first corrective layer guaranteed that the lifetime gain from increasing labor supply equals the contribution to output:

$$\frac{\partial \tilde R_p(\tilde h)}{\partial \tilde h(a)} = v(\tilde h)\, q(a). \tag{36}$$

The separation hypothesis makes the right-hand side well defined, since on the equilibrium path the labor supply history identifies the type and hence v. Define the corrective layer globally as a functional of the labor history through discounted lifetime labor: let $E(\tilde h) = \int_0^1 q(a)\, \tilde h(a)\, da$, let $\theta(e)$ denote the type whose equilibrium discounted lifetime labor equals e (well defined on the equilibrium range by separation and the strict monotonicity of E in the type, and extended continuously off that range), and set

$$\tilde R_p(\tilde h) = \Phi\big(E(\tilde h)\big), \qquad \Phi(E) = \int_{\underline{E}}^{E} v(\theta(e))\, de.$$

Then $\tilde R_p$ is differentiable in every direction, on and off the equilibrium menu, with $\partial \tilde R_p / \partial \tilde h(a) = v(\theta(E(\tilde h)))\, q(a)$, so condition (36) holds along the equilibrium path, where $\tilde R_p(\tilde h_\theta) = \int_{\underline{\theta}}^{\theta} v(t)\, dE(t)$. For the factorization, types with equal $\tilde R_p$ have equal

E; by Lemma 7 the optimal flows are timing-efficient, so $E = e_H(H)$ along the optimum with e_H strictly increasing, hence equal E implies equal aggregate H; and by Assumption 1(a) a labor history with aggregate H imposes the same disutility on every type, so mimicry across equal-H menu items is costless and incentive compatibility forces equal post-tax income. The optimum therefore factors through $\tilde{R}_p$, and only the redistributive layer is completed confiscatorily off the equilibrium menu. In terms of earnings, $R_p(\tilde{y}(\tilde{h})) = \tilde{R}_p(\tilde{h})$, and by the chain rule

$$v(\tilde{h})\, q(a) = \frac{dR_p(\tilde{y})}{d\tilde{y}(a)}\, w(\tilde{h}^a) + \int_a^1 \frac{dR_p(\tilde{y})}{d\tilde{y}(\tilde{a})}\, \frac{dw(\tilde{h}^{\tilde{a}})}{d\tilde{h}(a)}\, \tilde{h}(\tilde{a})\, d\tilde{a}, \tag{37}$$

which is conditions 1–2 of the proposition: the corrective layer neutralizes both the timing distortion and the lifetime distortion. On top of it, by Lemma 7(ii), the redistributive layer R_m satisfies the standard Mirrleesian formula (21) in terms of R_p. When there are no pre-tax intertemporal distortions, $\frac{dR_p}{d\tilde{y}(\tilde{a})}\frac{1}{q(\tilde{a})} = \frac{dR_p}{d\tilde{y}(a)}\frac{1}{q(a)}$ for all $a, \tilde{a}$, and (37) simplifies to

$$\frac{dR_p(\tilde{y})}{d\tilde{y}(a)}\frac{1}{q(a)} = \frac{v(\tilde{h}^a)\, q(a)}{q(a)w(\tilde{h}^a) + \int_a^1 \frac{dw(\tilde{h}^{\tilde{a}})}{d\tilde{h}(a)}\tilde{h}(\tilde{a})q(\tilde{a})\, d\tilde{a}}.$$

□

A.11 Age-Dependent Corrective Taxes (Proof of Corollary 2)

Proof. In the pace case with the parametric assumptions of Section 4.3, each worker supplies labor at a constant rate, pre-tax salaries depend only on the pace $\bar{h}(a) = \int_0^a \tilde{h}(\tilde{a})/a\, d\tilde{a}$, and in equilibrium equal productivities. Equation (37) becomes

$$w(\bar{h})\, e^{-\rho a} = \frac{dR_p(\tilde{y})}{d\tilde{y}(a)}\, w(\bar{h}(a)) + \int_a^1 \frac{dR_p(\tilde{y})}{d\tilde{y}(\tilde{a})}\, \frac{dw(\bar{h}(\tilde{a}))}{d\tilde{h}(a)}\, \bar{h}\, d\tilde{a}.$$

Using $\frac{dw(\bar{h}(\tilde{a}))}{d\tilde{h}(a)} = w'(\bar{h}(\tilde{a}))\frac{d\bar{h}(\tilde{a})}{d\tilde{h}(a)} = \frac{w'(\bar{h})}{\tilde{a}}$ and defining $r_y(a)$ by $\frac{dR_p(\tilde{y})}{d\tilde{y}(a)} = r_y(a)e^{-\rho a}$, the equation becomes

$$e^{-\rho a} = r_y(a)\, e^{-\rho a} + \int_a^1 r_y(\tilde{a})\, \frac{e^{-\rho \tilde{a}}}{\tilde{a}}\, \frac{w'(\bar{h})\, \bar{h}}{w(\bar{h})}\, d\tilde{a}.$$

Under the parametric assumptions the wage elasticity is constant, $\frac{w'(\bar{h})\bar{h}}{w(\bar{h})} = \gamma = \frac{\delta}{1+\epsilon-\delta}$. Differentiating with respect to a gives

$$r_y'(a) = \frac{\gamma}{a} r_y(a) + \rho\big(r_y(a) - 1\big),$$

with terminal condition $r_y(1) = 1$ from the equation at $a = 1$. The homogeneous solutions are proportional to $a^\gamma e^{\rho a}$, and varying the constant gives the unique solution

$$r_y(a) = a^\gamma e^{\rho a}\left[e^{-\rho} + \rho \int_a^1 s^{-\gamma} e^{-\rho s}\, ds\right],$$

hence $\frac{dR_p(\tilde{y})}{d\tilde{y}(a)} = a^\gamma\big[e^{-\rho} + \rho \int_a^1 s^{-\gamma} e^{-\rho s}\, ds\big]$, which reduces to a^γ when $\rho = 0$. Two properties used in the text follow from the differential equation. First, $r_y < 1$ at every interior age: if r_y reached one before retirement, the equation would give $r_y' = \gamma/a > 0$ there, so r_y would cross one from below and, since $r_y' > 0$ whenever $r_y \geq 1$, could never return to its terminal value. Second, r_y is strictly increasing: at an interior critical point $r_y'' = -\gamma r_y/a^2 < 0$, so every critical point is a local maximum; a decrease would have to be followed by a return to the terminal value through a local minimum, which is impossible, so $r_y' > 0$ throughout. □

A.12 Welfare (Proof of Proposition 12)

Definition 3. *The planner has redistributive preferences if $W(V,\theta) = W(V)$ and $W'(V)U_C$ is positive and decreasing.*

Proof. By Lemma 2 (or Lemma 9 in the general case), the planner solves for the allocation directly, in terms of consumption and deliverables, subject to incentive compatibility and resources. Writing effort as $\ell(\theta) = h(\theta)/b(\theta)$, the problem is

$$\max_{\ell(\theta), c(\theta)} \int W\big(u(c(\theta), \ell(\theta))\big) f(\theta)\, d\theta \quad \text{s.t.} \quad u\big(c(\theta), \ell(\theta)\big) \geq u\Big(c(\theta'), \ell(\theta')\tfrac{b(\theta')}{b(\theta)}\Big) \;\; \forall \theta, \theta',$$

$$\int \big(c(\theta) - v(\theta) b(\theta) \ell(\theta)\big) f(\theta)\, d\theta \leq 0.$$

Replace global with local incentive constraints, $u'(\theta) = U_\theta$. Under single crossing (MRS strictly decreasing in the type), local constraints plus monotonicity of the allocation imply the global constraints. Since we maintain the hypothesis that the relaxed solutions are monotone under both information structures, the replacement is without loss, and the comparison below is between the two full optima rather than between their relaxations. Write the Lagrangian with costate $\mu(\theta)$ on the local constraints and multiplier κ on resources, integrate by parts, and take first-order conditions in u and h. Following Seade [1982] and Werning [2000], the conditions imply

$$\mu\, U_C/\kappa = \frac{f\,(v-r)}{-\,\partial \log MRS/\partial\theta},$$

and $\mu(\theta) \geq 0$ everywhere. To see this, suppose toward a contradiction that $\mu(\theta) < 0$ somewhere. Then there exist $\theta_1 < \theta_2$ with $\mu(\theta_1) = \mu(\theta_2) = 0$ and $\mu'(\theta_1) < 0 < \mu'(\theta_2)$. At such points, the display above implies the workers are undistorted, $v = r$, so $e_u = 1/U_C$ at θ_1 and θ_2; and the first-order condition in u evaluated where $\mu = 0$ gives

$$e_u(\theta_i) = \frac{1}{\kappa}\left(\lambda(\theta_i) - \frac{\mu'(\theta_i)}{f(\theta_i)}\right), \qquad i = 1, 2.$$

Since λ is decreasing (redistributive preferences) and $-\mu'(\theta_1) > 0 > -\mu'(\theta_2)$, we get $e_u(\theta_1) \geq e_u(\theta_2)$, hence $U_C(\theta_2) \geq U_C(\theta_1)$. Under convex preferences with leisure a normal good, higher marginal utility of consumption at θ_2 implies $u(\theta_2) \leq u(\theta_1)$—which is impossible, since θ_2 is more productive and could attain strictly more than $u(\theta_1)$ by imitating θ_1's allocation, violating incentive compatibility. Hence the local downward incentive constraints bind ($\mu \geq 0$), and $v - r \geq 0$.

Now decrease information asymmetries per Definition 4: $v^n = v/\Delta$, $b^n = b\cdot\Delta$ with Δ increasing. The resource constraint, written in terms of effort, $\int (e(u,\ell) - \theta\,\ell) f\, d\theta \leq 0$ with $\theta = vb$ unchanged, is unaffected. The local incentive constraints,

$$u'(\theta) = -\frac{\partial U}{\partial \ell}\,\ell(\theta)\,\frac{b'(\theta)}{b(\theta)} \;\longrightarrow\; -\frac{\partial U}{\partial \ell}\,\ell(\theta)\left(\frac{b'(\theta)}{b(\theta)} + \frac{\Delta'(\theta)}{\Delta(\theta)}\right),$$

become tighter, since $\Delta'/\Delta > 0$ adds to the information-rent term and the multipliers $\mu(\theta)$ are nonnegative. A tightening of binding constraints with an unchanged objective and resource constraint weakly reduces welfare. The reduction is strict: with $W'(V)U_C$ strictly decreasing, the costate $\mu(\theta) = \int_\theta (1-\hat\lambda)\, f\, dt$ is strictly positive at every interior type, and with $\Delta'/\Delta > 0$ the added rent term is strictly positive wherever $\ell > 0$, so the first-order welfare effect of the tightening, $-\int \mu(\theta)\,(-U_\ell)\,\ell(\theta)\,\frac{\Delta'(\theta)}{\Delta(\theta)}\,d\theta$, is strictly negative at any allocation with interior labor supply on a set of positive measure. The comparison between the two information structures follows by integration: for $t \in [0,1]$, the intermediate structure $\Delta_t \equiv \Delta^t$ satisfies $\Delta_t'/\Delta_t = t\,\Delta'/\Delta$ and, by hypothesis, the conditions of the proposition; the argument above applied at each $t > 0$ gives $dW(t)/dt < 0$, so $W(1) < W(0)$. □

A.13 The Lifetime Wedge and the Information Order (Proof of Lemma 13)

Proof. Under Definition 4 the ranking of types and their final resumes is preserved, so the conditioning event $\{\tilde{\theta} \geq \theta\}$ is unchanged. For $\tilde{\theta} \geq \theta$, Δ increasing gives $\Delta(\tilde{\theta}) \geq \Delta(\theta)$, hence $v(\tilde{\theta})/\Delta(\tilde{\theta}) \leq v(\tilde{\theta})/\Delta(\theta)$ pointwise on the conditioning event, and taking conditional expectations,

$$\mathbb{E}\big[v(\tilde{\theta})/\Delta(\tilde{\theta}) \mid \tilde{\theta} \geq \theta\big] \leq \frac{1}{\Delta(\theta)}\,\mathbb{E}\big[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta\big].$$

Therefore

$$\bar{\chi}^n(\theta) = \frac{v(\theta)/\Delta(\theta)}{\mathbb{E}[v(\tilde{\theta})/\Delta(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]} \geq \frac{v(\theta)/\Delta(\theta)}{\mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]/\Delta(\theta)} = \bar{\chi}(\theta),$$

with strict inequality wherever Δ is strictly increasing above θ on a set of positive probability. □

A.14 Algebra Behind the Parametric Example of Section 2.3

Salaries equal the expected productivity of those who supply at least h; in terms of types, using the Pareto distribution, $w(h(\theta)) = \mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta] = \frac{\alpha}{\alpha-\delta}\,\theta^{\delta}$. The worker's first-order condition under (7) implies

$$w(h) = h^{1/\epsilon}\, b(\theta)^{-(1+1/\epsilon)}. \tag{38}$$

Guessing and verifying that the wage is a power function, $w(h) = kh^{\gamma}$, we conclude that

$$w(h) = \frac{\alpha}{\alpha-\delta}\, k^{\frac{-\delta\epsilon}{(1+\epsilon)(1-\delta)}}\, h^{\frac{\delta(1-\epsilon\gamma)}{(1+\epsilon)(1-\delta)}}, \tag{39}$$

and therefore $w(h) = kh^\gamma$ with

$$k = \left(\frac{\alpha}{\alpha-\delta}\right)^{\frac{(1-\delta)(1+\epsilon)}{(1-\delta)(1+\epsilon)+\delta\epsilon}} = \left(\frac{\alpha}{\alpha-\delta}\right)^{\frac{(1-\delta)(1+\epsilon)}{1-\delta+\epsilon}}, \qquad \gamma = \frac{\delta}{1-\delta+\epsilon},$$

as claimed in the main text. For the lifetime wedge, $\bar{\chi}(\theta) = v(\theta)/\mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta] = \theta^\delta / \left(\frac{\alpha}{\alpha-\delta}\theta^\delta\right) = \frac{\alpha-\delta}{\alpha}$, constant across types, so the Pigouvian correction is $\tau_p = 1 - \bar{\chi} = \delta/\alpha$, as used in Sections 3.4 and 4.2.

A.15 Existence of Equilibrium

Proposition 7 (Existence)**.** *Assume $v(\theta)$ is smoothly increasing in types; $MRS(c, h, \theta) \equiv -U_h/U_c$ is smoothly decreasing in types and increasing in consumption and labor; the distribution of types has compact support and a continuously differentiable density with nonvanishing hazard λ, so that $w(\theta) = \mathbb{E}[v \mid \tilde{\theta} \geq \theta]$ is continuously differentiable, with $w' = \lambda(w - v)$; $MRS_c + MRS_h/w$, where $w(\theta) \equiv \mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]$, is bounded away from zero; MRS_θ, MRS_h, MRS_c are uniformly Lipschitz continuous in c and h; and, for the lowest type, $MRS(h\,\mathbb{E}[v], h, \underline{\theta}) - \mathbb{E}[v]$ is continuous in h and changes sign on $(0, \infty)$. Then an equilibrium exists.*

The role of the last condition is to give the shooting argument below a starting point. It is satisfied whenever the marginal disutility of labor vanishes at zero labor supply and grows without bound, which is the case in the parametric example of Section 2.3.

Proof. Consider the allocation $(c(\theta), h(\theta))$ solving the system

$$\frac{c'(\theta)}{h'(\theta)} = MRS\big(c(\theta), h(\theta), \theta\big), \qquad \frac{c'(\theta)}{h'(\theta)} = \mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta],$$

with boundary conditions $MRS(c(\underline{\theta}), h(\underline{\theta}), \underline{\theta}) = c(\underline{\theta})/h(\underline{\theta})$ and $c(\underline{\theta}) = h(\underline{\theta})\,\mathbb{E}[v(\theta)]$. The first equation is local incentive compatibility. The two boundary conditions jointly require $h_0 > 0$ with $MRS(h_0\,\mathbb{E}[v], h_0, \underline{\theta}) = \mathbb{E}[v]$; by the sign-change condition and the intermediate value theorem such an h_0 exists, and we set $(c(\underline{\theta}), h(\underline{\theta})) = (h_0\,\mathbb{E}[v], h_0)$. As for monotonicity, c and h are increasing in θ because $\mathbb{E}[v \mid \tilde{\theta} \geq \theta]$ increases in θ while MRS decreases in θ and increases in c and h; by single crossing, local incentive compatibility plus monotonicity imply global incentive compatibility.

Feasibility holds automatically: workers are paid $\mathbb{E}[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]$ and firms break even in expectation. Rearranging, the system can be written as an initial value problem,

$$c'(\theta) = \frac{\frac{d}{d\theta}\mathbb{E}[v \mid \tilde{\theta} \geq \theta] - MRS_\theta}{MRS_c + MRS_h \,/\, \mathbb{E}[v \mid \tilde{\theta} \geq \theta]}, \qquad h'(\theta) = \frac{c'(\theta)}{\mathbb{E}[v \mid \tilde{\theta} \geq \theta]},$$

whose right-hand sides are continuous in θ, since w is C^1 on the compact support, and locally Lipschitz in (c, h) on the relevant compact sets under the maintained assumptions, so a solution exists by standard existence theorems for initial value problems [Coddington and Levinson, 1955]. $\square$

B Extensions and Additional Results

B.1 Model Extensions

Several extensions add realism to the basic model; the key insights from the generalized optimal taxation formulas hold in each, with caveats noted below.

B.1.1 On-the-Job Learning

The assumption of constant lifetime productivity is extreme; part of the observed return to experience surely reflects on-the-job learning. Let productivity depend on experience itself, $v = v(\theta, h)$. Revisiting the parametric example of Section 2.3 with productivities growing proportionally with experience, $v(\theta, h) = \tilde{v}(\theta)h^\beta$, wages become $w(h) = h^\beta\, \mathbb{E}[\tilde{v}(\tilde{\theta}) \mid \tilde{\theta} \geq \theta(h)]$, and the log-linear wage profile has a slope that combines the human capital component β with a signaling component attenuated by learning through the equilibrium sorting of experience (assuming $\beta\epsilon < 1$, so that experience remains increasing in the type):

$$\log w = \left(\beta + \frac{\delta(1-\beta\epsilon)}{1-\delta+\epsilon}\right)\log h + \left(\frac{(1-\delta)(1+\epsilon)}{(1-\delta)(1+\epsilon)+\delta\epsilon}\right)\log\left(\frac{\alpha}{\alpha-\delta}\right). \tag{40}$$

Signaling and learning are observationally equivalent in the wage profile—but not in the tax formula.

Remark 1 (Behavior without corrective taxes)**.** *The dependence of the corrective taxes on age suggests that, without them, younger workers would exert a lot more* *effort and work longer hours, and paces would not be constant over the career. This prediction is in line with Landers, Rebitzer, and Taylor [1996], who show that, in law firms, hours decrease with tenure and associates work too many hours relative to their desires as well as relative to partners. It is, however, at odds with the commonly found inverted U-shaped pattern for hours over the life cycle: younger and less experienced workers work fewer hours, hours are mostly constant for workers between the ages of 25 and 55, and they quickly fall for older workers [Card, 1994, Kaplan, 2012]. Importantly, both pieces of evidence could alternatively be explained by a combination of age-dependent preferences and age-dependent patterns of human capital accumulation and depreciation. The normative implications of human capital accumulation and signaling, however, are very distinct. For that reason, Section 5 gauges the strength of the mechanism from the signaling share of the returns to schooling rather than from hours profiles.*

Proposition 8 (Human capital)**.** *Suppose productivity depends on experience and unobserved type. The total contribution to output of a worker is* $\int_0^{h(\tilde{R},\theta)} v(\theta,\tilde{h})\,d\tilde{h}$*, and, writing the retention schedule as* $\tilde{R}(h) = \tilde{R}(0) + \int_0^h \tilde{r}(z)\,dz$ *with intercept* $\tilde{R}(0)$*, the planner's problem becomes*

$$\max_{\tilde{R}(h)} \mathbb{E}[W(V(\tilde{R};\theta),\theta)] \quad s.t. \quad \mathbb{E}\left[\int_0^{h(\tilde{R},\theta)} \big(v(\theta,\tilde{h}) - \tilde{r}(\tilde{h})\big)d\tilde{h} - \tilde{R}(0)\right] \geq 0. \tag{41}$$

A necessary condition for optimality is formula (19)*, with* $\chi(y) = v(y)/y'(h(y))$*, where* $v(y) = v(\theta(y), h(\theta(y)))$ *is the productivity* at retirement *of the worker with lifetime income* y*.*

The formula is unchanged, but its inputs are reinterpreted, with two practical consequences. First, productivities of workers far from retirement cannot be used to infer productivity at retirement. Second, the degree of informational asymmetry can no longer be read off the return to experience alone—which is exactly why the calibration of Section 5 must take a stance on the signaling share of the return to experience, imported from the schooling context, rather than infer it from the slope of the wage profile, and why its second route prices the dispersion within pools directly. More generally, the corrective layer of Proposition 6 extends by interpreting v broadly as the present value of the contribution of a marginal deliverable to output, including its effect on future productivity, when learning is costless.

B.1.2 Signals the Government Does Not See

Firms may extract signals the government cannot condition on: some workers are visibly on different career tracks, but codifying track distinctions into the tax code may be impossible or manipulable. Let firms observe an exogenous signal z in addition to the resume. Then Lemma 2 does not apply directly—there is one pre-tax salary function per signal realization but a single nonlinear tax instrument. Considering a variation of the income retention schedule and tracking how it moves post-tax salaries across careers yields a weighted version of the basic formula.

Proposition 9 (Firms see additional signals)**.** *If a tax schedule is optimal, then*

$$\begin{aligned}
&\mathbb{E}_z\left[\int_y^\infty \left(\frac{\chi(\tilde{y},z)-r_{\tilde{y}}}{r_{\tilde{y}}}\right)\epsilon_r^c(\tilde{y},z)\,g(\tilde{y}|z)\,\tilde{y}\,\frac{dr_{h(\tilde{y};z)}}{dr_y}\,d\tilde{y}\right] \\
&\quad= \mathbb{E}_z\left[\int_y^\infty\int_{\tilde{y}}^\infty \big(1-\lambda(\tilde{\tilde{y}};z)\big)g(\tilde{\tilde{y}}|z)\,\frac{dr_{h(\tilde{y};z)}}{dr_y}\,d\tilde{\tilde{y}}\,d\tilde{y}\right] \\
&\qquad+ \mathbb{E}_z\left[\int_y^\infty\int_{\tilde{y}}^\infty \left(\frac{\chi(\tilde{\tilde{y}},z)-r_{\tilde{\tilde{y}}}}{r_{\tilde{\tilde{y}}}}\right)\eta_I(\tilde{\tilde{y}})\,g(\tilde{\tilde{y}}|z)\,\frac{dr_{h(\tilde{y};z)}}{dr_y}\,d\tilde{\tilde{y}}\,d\tilde{y}\right],
\end{aligned} \tag{42}$$

where $\chi(\tilde{y},z) = v(h(\tilde{y},z);z)/y'(h(\tilde{y},z);z)$ and $\frac{dr_{h(\tilde{y};z)}}{dr_y}$ is the response of post-tax salaries of a worker initially earning y with signal z to the variation in income retention. Formally, consider the perturbation that raises marginal retention on a small earnings bracket at y; the weight is the Gateaux derivative of the marginal retention per unit of labor faced at earnings level $\tilde{y}$ on track z with respect to that perturbation, normalized so that, with no information asymmetry and unchanged within-track salary schedules, it is the identity operator (a Dirac kernel at $\tilde{y}=y$), in which case (42) *reduces track by track to* (19)*.*

Proof. Consider a small variation in marginal retention as a function of earnings; it translates into a $\frac{dr_h}{dr_y}$-weighted variation in retention as a function of effort within each career track z. Repeating the steps of the proof of Theorem 1 with these weights delivers (42). □

Optimal policy is a weighted version of the standard condition, with weights given by how much post-tax wages move when income taxes change on each career path. Absent income effects the income-effect block drops; the kernel itself is diagonal only under the stability conditions below. The weights $\frac{dr_{h(y;z)}}{dr_y}$ are equilibrium pass-through responses: they embed the adjustment of within-track wage schedules and any reallocation of workers across tracks, and the proposition places no restriction on them. With no information asymmetry the pass-through is one for one. When, in addition, within-track wage schedules are stable and the reform induces no reallocation across tracks, the weights lie between zero and one, and the heterogeneity attenuates, without overturning, the force by which informational asymmetries push toward higher marginal taxes; outside these conditions the direction of the adjustment is not determined by the formula alone.

B.1.3 Conditioning on the Full History of Deliverables

Finally, let firms observe the entire history $\tilde{h}^a$, including the detailed timing of execution.[22] For expectations conditional on rich histories to be meaningful, the type space must be rich: we assume preferences $U(C, \tilde{h}, \theta)$ with high-dimensional θ, a single consumption aggregate, and maintain that more productive types are more willing to supply deliverables at every age.

Under these assumptions, optimal taxes are still the composition of (i) corrective taxes guaranteeing that each additional unit of labor yields lifetime benefits equal to its contribution to output, and (ii) redistributive taxes—which, however, no longer take the simple Mirrleesian form, since with unrestricted preferences and rich heterogeneity no normative argument singles out lifetime income as the redistributive base.

Lemma 10 (Production efficiency)**.** *If the planner chooses the allocation subject to incentive compatibility, any optimal allocation lies at the frontier of the production possibilities set.*

Proof. Analogous to the production-efficiency argument of Diamond and Mirrlees [1971]. By the taxation principle, incorporate incentive compatibility into indirect utilities and write the problem as choosing $R(\tilde{h}_\theta)$ subject to resources. If production were interior, a uniform increase in R—feasible for a small enough step, by continuity of labor supply choices—would raise every worker's utility and the planner's objective, a contradiction. □

Proposition 10 (Corrective taxes with full histories)**.** *Suppose the optimal allocation is separating, and that expected productivity is constant on level sets of discounted*

[22] Workers of different ages are then never pooled: histories ending at different ages differ.

lifetime labor along the optimum: $v(\theta) = \psi(E_\theta)$ *with* $E_\theta = \int_0^1 q(a)\tilde{h}_\theta(a)\,da$ *and* ψ *increasing (automatic when types are ranked by a scalar index in which* E *is strictly increasing). The planner can guarantee that the allocation lies at the frontier of the production possibilities set using corrective taxes of the form of Proposition 6.*

The argument parallels Section 4.3 and rests on two lemmas, the full-history counterparts of Lemmas 8 and 9.

Lemma 11 (Positive return to experience with full histories)**.** *Assume, in addition, that productivity and the willingness to supply deliverables are affiliated across the rich type space, so that higher labor supply at any age is associated with higher* v *in the sense of positive selection. In any optimal allocation, salaries* $w(\tilde{h}^a) = \mathbb{E}[v(\theta) \mid \tilde{h}^a]$ *are increasing in the labor supply choices* $\tilde{h}(\tilde{a})$, $\tilde{a} \leq a$.

Proof. Affiliation delivers the monotonicity of the conditional expectation in each coordinate of the history; richness of the type space ensures that the expectations are well defined for every path $\tilde{h}$. □

Lemma 12 (Earnings reveal labor supply with full histories)**.** *Suppose labor supply histories are bounded and equilibrium salaries* $w(\tilde{h}^a) = \mathbb{E}[v(\theta) \mid \tilde{h}^a]$ *are bounded below by* $w_{\min} > 0$ *(as when* v *is bounded away from zero), depend on the history through its restriction to* $[0, a)$*, and are Lipschitz in that restriction in the* L^1 *norm, uniformly in* a*. Then the planner can infer labor supply choices from earnings: no two continuous* $\tilde{h}$ *map to the same* $\tilde{y}$.

Proof. As in Lemma 9: earnings at each age equal current labor times a salary that depends only on past labor, so the planner recovers the history forward in age, dividing earnings by the salary implied by the already-recovered past. Formally, if $\tilde{h}$ and $\tilde{h}'$ generate the same earnings history, then $\tilde{h}(a) = \tilde{y}(a)/w(\tilde{h}^a)$ and likewise for $\tilde{h}'$, and the boundedness of earnings together with the Lipschitz property give, for the gap $D(a) = |\tilde{h}(a) - \tilde{h}'(a)|$, the inequality $D(a) \leq K\int_0^a D(\tilde{a})\,d\tilde{a}$ for a constant K. As in the proof of Lemma 9, a continuous gap with this property is zero on successive intervals of length $1/(2K)$, hence everywhere. □

Proof of Proposition 10. Consider the planning problem over $R(\tilde{h})$ subject to resources and incentive compatibility (Lemma 10). Its solution is a wage schedule $R(\tilde{h})$, a functional of labor supply decisions, which by Lemma 12 can be written as a functional of the history of earnings, $R(\tilde{y})$, and hence implemented with history-dependent earnings taxes, delivering each type's assigned post-tax income on the equilibrium path and confiscating any surplus off it, as in the taxation principle.

Suppose there is separation, so that the map $\theta \mapsto \tilde{h}_\theta$ is one to one on the equilibrium path, and let $E(\tilde{h}) = \int_0^1 q(a)\,\tilde{h}(a)\,da$. Define the corrective layer globally by

$$\tilde{R}_p(\tilde{h}) \;=\; \Phi\big(E(\tilde{h})\big), \qquad \Phi(E) = \int_{\underline{E}}^{E} \psi(e)\,de,$$

using the hypothesis that expected productivity is constant on level sets of discounted lifetime labor, $v(\theta) = \psi(E_\theta)$. Then $\tilde{R}_p$ is differentiable in every direction with $\partial \tilde{R}_p(\tilde{h})/\partial \tilde{h}(a) = \psi(E(\tilde{h}))\,q(a)$, equal to $v(\theta)\,q(a)$ along the equilibrium path. With unrestricted preferences, the assignment need not factor through a scalar; instead, because the government observes full earnings histories and separation makes histories identify types, the redistributive layer is indexed by the equilibrium history directly, $\tilde{R}(\tilde{h}_\theta) = \tilde{R}_p(\tilde{h}_\theta) + s(\tilde{h}_\theta)$, with both layers completed confiscatorily off the equilibrium path: a corrective layer of the form of Proposition 6, with a history-indexed redistributive layer on top.[23]

□

B.2 First Best with Purely Corrective Taxes

Proposition 11 (Common point of the first- and third-best frontiers)**.** *Under the assumptions of Proposition 7 (with $MRS_c + MRS_h v$ bounded away from zero), a first-best allocation can be achieved with $r(y) = \chi(y)$.*

Proof. Consider the direct mechanism offering $(c(\theta), h(\theta))$ solving

$$\frac{c'(\theta)}{h'(\theta)} = MRS\big(c(\theta), h(\theta), \theta\big), \qquad \frac{c'(\theta)}{h'(\theta)} = v(\theta),$$

[23] If there is no separation, the solution can still be written as $\tilde{R}(\tilde{h}) = R_m(\tilde{R}_p(\tilde{h}(\cdot;\theta)), \theta)$ with $\partial R_p(\tilde{h})/\partial \tilde{h}(a) = v(\theta)q(a)$ along the path. This is an accounting decomposition of the assignment rather than a tax system: the layers are indexed by the unobserved type, and no functional of observable histories has the prescribed derivative for pooled types with different productivities. It isolates, type by type, the part of assigned pay that mimics a corrected price system; the residual type dependence measures the corrective error that pooling forces on any implementable system. The decomposition becomes implementable when pooled types share the same productivity, or if the corrective price is set at the expected productivity conditional on observable histories, a modification we do not develop.

with boundary conditions $\int(c(\theta) - v(\theta)h(\theta))f(\theta)\,d\theta = 0$ and $MRS(c(\underline{\theta}), h(\underline{\theta}), \underline{\theta}) = v(\underline{\theta})$. As in Proposition 7, monotonicity and single crossing deliver global incentive compatibility, and existence follows from standard ODE theorems, with the initial value pinned down by continuity: making the lowest types consume their own product leaves the resource constraint slack, making the highest types do so violates it, so an intermediate initial value balances it. Along this allocation $\int_{\underline{\theta}}^{\theta} v(\tilde{\theta})h'(\tilde{\theta})\,d\tilde{\theta} < v(\theta)h(\theta)$, so $c(\underline{\theta}) > 0$: higher types subsidize lower types and consumption is positive everywhere. Marginal retention satisfies $r(y) = \chi(y)$, and the allocation corresponds to workers facing linear budgets, paid their marginal products, with lump-sum transfers $c(\theta) - v(\theta)h(\theta)$: a first best. $\square$

B.3 Comparative Statics of the Labor Wedges

Beyond the parametric family, the labor wedges obey general monotone comparative statics. The right order on primitives is the definition of decreasing information asymmetries from Appendix B.4, under which heterogeneity shifts from the unobserved productivities v toward the observable component b, holding total productivities and the ranking of types fixed.

Lemma 13 (Labor wedges and information asymmetries)**.** *Consider the length case, in which the retirement pool of type θ is the upper tail $\{\tilde{\theta} \geq \theta\}$ with population weights, and fix an allocation in which higher types accumulate strictly higher final lengths. If information asymmetries decrease in the sense of Definition 4—$v^n = v/\Delta$, $b^n = b \cdot \Delta$ with $\Delta > 0$ increasing and order-preserving—then the labor wedge at retirement rises pointwise toward one:*

$$\chi^n(\theta) = \frac{v(\theta)/\Delta(\theta)}{\mathbb{E}\big[v(\tilde{\theta})/\Delta(\tilde{\theta}) \mid \tilde{\theta} \geq \theta\big]} \;\geq\; \frac{v(\theta)}{\mathbb{E}\big[v(\tilde{\theta}) \mid \tilde{\theta} \geq \theta\big]} = \chi(\theta),$$

with strict inequality wherever Δ is strictly increasing above θ on a set of positive probability.

Proof. See Appendix A. The inequality is one line: for $\tilde{\theta} \geq \theta$, $\Delta(\tilde{\theta}) \geq \Delta(\theta)$ implies $v(\tilde{\theta})/\Delta(\tilde{\theta}) \leq v(\tilde{\theta})/\Delta(\theta)$ inside the conditional expectation. $\square$

For the intertemporal distortions, the relevant comparison is between signal weights that reward effort earlier or later in the career. A clean statement is available holding the equilibrium wage schedule and allocation fixed; the full equilibrium adjustment is discussed below.

Lemma 14 (Signal weights and intertemporal distortions)**.** *Fix a career path of salaries $W(\tilde{a}) = w(I(\tilde{a}))$ and of salary gradients $G(\tilde{a}) = w'(I(\tilde{a})) > 0$ along the equilibrium path, and an allocation with $\tilde{h} > 0$. Consider two weighting functions with $\varphi_1(a, \tilde{a}) \geq \varphi_2(a, \tilde{a})$ for all $\tilde{a} > a$, with strict inequality on a positive-measure set. Then the marginal lifetime returns evaluated at the common path,*

$$m_i(a; \theta) = W(a) + \frac{1}{q(a)} \int_a^1 q(\tilde{a})\, G(\tilde{a})\, \varphi_i(a, \tilde{a})\, \tilde{h}(\tilde{a})\, d\tilde{a},$$

satisfy $m_1(a; \theta) \geq m_2(a; \theta)$ at every interior age with strict inequality where the weights differ, while $m_1(1; \theta) = m_2(1; \theta) = W(1)$. Hence $\iota_1(a; \theta) \leq \iota_2(a; \theta)$ at every interior age: the intertemporal distortions become larger.

Proof. Immediate from the display: with the path (W, G) fixed, the reputational term is increasing in the future weights $\varphi(a, \tilde{a})$, $\tilde{a} > a$, and vanishes at $a = 1$ regardless of the weights. □

The comparison holds the units of the signal fixed: the two kernels are compared at a common parametrization, against the same path of salaries and salary gradients. It is a statement about reweighting the timing of deliverables, not about rescaling the signal, which would rescale the gradient inversely and leave the returns unchanged.

Lemma 14 is a statement about the direct effect: it evaluates the returns at a fixed path of salaries and salary gradients. A change in φ also changes the realized signals, and with them the equilibrium wage paths and allocations, so the total effect is not signed in general. The pace case shows that the direct effect can dominate. There, the equilibrium adjustment leaves salaries equal to marginal products at every instant, and yet the intertemporal distortions are large.

B.4 Welfare and Information Technologies

New monitoring and screening technologies—richer data on workers, algorithmic evaluation of output—change how workers are paid and how resumes are read, arguably making deliverables a better measure of product. Their welfare effects, however, are not what partial-equilibrium intuition suggests.

Definition 4 (Less information asymmetry)**.** *Let preferences be $U(C, H/b(\theta))$. Information asymmetries* decrease *if new productivities and tastes satisfy $v^n(\theta) = v(\theta)/\Delta(\theta)$ and $b^n(\theta) = b(\theta) \cdot \Delta(\theta)$ for some $\Delta > 0$ increasing and order-preserving (types keep their ranking in productivities and marginal rates of substitution).*

Proposition 12 (Less asymmetry, lower welfare)**.** *Suppose the original tax schedule is optimal, the planner has redistributive preferences ($W'(V)U_C$ positive and strictly decreasing), workers' preferences are convex (convex upper contour sets over consumption and leisure), and leisure is a normal good; and suppose that, under both information structures, the solution of the planner's problem with local incentive constraints is monotone in the type, so that local and global incentive compatibility coincide and there is no bunching; let Δ in Definition 4 be continuously differentiable with $\Delta'/\Delta > 0$; and suppose the conditions above hold at every intermediate structure $\Delta_t \equiv \Delta^t$, $t \in [0, 1]$. Then decreasing information asymmetries in labor markets strictly decreases welfare.*

Proof. See Appendix A. The proof is self-contained: it reproduces the argument that only local downward incentive constraints bind [following Seade, 1982, Werning, 2000], so that tightening them reduces welfare. □

Under the order of Definition 4 the first-best frontier is unchanged, since total productivities vb are untouched; what shifts inward is the constrained frontier. The intuition is that a redistributive planner transfers resources from high to low types, so downward incentive constraints bind; when information asymmetries shrink, high types can use more of their previously invisible productivity to imitate low types, tightening exactly the constraints that bind. In this sense, information asymmetries in the labor market work as implicit cross-subsidies from high- to low-productivity workers, and they do part of the planner's work. The result generalizes Proposition 13 of Stantcheva [2014] to intermediate degrees of informational friction and to any model satisfying the invertibility conditions of Lemma 2. The impact on *taxes* is more subtle. The corrective factor moves toward one, a force toward lower corrective taxes, while the Mirrleesian factor moves ambiguously; Appendix B.5 characterizes which effect dominates.

B.5 Effects of Information Technologies on Taxes

To understand how the Mirrleesian component and total marginal taxes respond to changes in information asymmetries (Appendix B.4), write optimal taxes in terms of types.

Proposition 13 (Optimal taxes as a function of types)**.** *Let the environment be the length case with single-dimensional types, an atomless distribution F with density f, and a monotone optimal allocation without bunching; $\eta(\theta)$ denotes the income effect on the labor supply of type θ, the type-space counterpart of the statistic η_I of (19). Optimal taxes satisfy*

$$r_y(\theta) = r_m(\theta) \cdot \chi(\theta), \tag{43}$$

$$\frac{1 - r_m(\theta)}{r_m(\theta)} f(\theta) \left(-\frac{\partial \log MRS}{\partial \theta} \right)^{-1} = \int_\theta^\infty \big(1 - \hat{\lambda}(\tilde{\theta})\big) f(\tilde{\theta})\, d\tilde{\theta} + \int_\theta^\infty \frac{1 - r_m(\tilde{\theta})}{r_m(\tilde{\theta})} \eta(\tilde{\theta}) f(\tilde{\theta})\, d\tilde{\theta}. \tag{44}$$

Proof. Adapting the derivation of Scheuer and Werning [2017] to this environment: set up the planner's problem over $(u(\theta), h(\theta))$ with local incentive constraints $u'(\theta) = U_\theta$ and the resource constraint; form the Lagrangian with costate $\mu(\theta)$ and multiplier κ; integrate by parts and take first-order conditions in u and h (as in the proof of Proposition 12); substitute $e_u = U_c^{-1}$, define $\hat{\mu} = \mu U_c / \kappa$, $\hat{\lambda} = \lambda U_c / \kappa$, and rearrange to obtain (44). □

Changes in the degree of informational asymmetry move χ and $\partial \log MRS / \partial \theta$, the ingredients of the Pigouvian and the Mirrleesian components respectively, and move them very differently. The Pigouvian component of type θ responds to the unobservable productivity of everyone with lifetime income above $y(\theta)$: as their labor becomes more measurable, θ receives smaller implicit subsidies and $\chi(\theta)$ rises toward one. The Mirrleesian component responds to how willing θ is to supply deliverables relative to *local* neighboring types. The net effect on total marginal taxes is in general ambiguous.

A simple example shows when the Pigouvian effect dominates. With no income effects, preferences as in Section 2.3, and Pareto types,

$$\chi = \frac{\alpha - \delta}{\alpha}, \qquad \frac{1 - r_m}{r_m} = \frac{1 - \bar{\lambda}}{\frac{\epsilon}{1+\epsilon} \frac{\alpha}{1-\delta}} \qquad \Longrightarrow \qquad r = r_m \cdot \chi = \frac{\frac{\epsilon}{1+\epsilon} \frac{\alpha - \delta}{1-\delta}}{(1 - \bar{\lambda}) + \frac{\epsilon}{1+\epsilon} \frac{\alpha}{1-\delta}}, \tag{45}$$

where $\bar{\lambda}(\theta) = \mathbb{E}[\hat{\lambda}(\tilde{\theta}) \mid \tilde{\theta} \geq \theta]$ is assumed constant above some $\hat{\theta}$. The corrective component rises with the degree of asymmetry δ and the redistributive component falls, so the net effect turns on parameters. Simplifying, $r = \frac{\epsilon}{1+\epsilon}(\alpha - \delta)\big/\big[(1-\bar{\lambda})(1-\delta) + \frac{\epsilon}{1+\epsilon}\alpha\big]$, and differentiating, r is nonincreasing in δ if and only if

$$(1-\bar{\lambda})(\alpha - 1) \;\leq\; \frac{\epsilon}{1+\epsilon}\,\alpha,$$

that is, if and only if the welfare weight at the top is not too small, $\bar{\lambda} \geq 1 - \frac{\epsilon\alpha}{(1+\epsilon)(\alpha-1)}$; the decrease is strict when the inequality is strict, and at equality the Pigouvian and redistributive effects exactly offset. Since lifetime income is proportional to $\theta^{1+\epsilon}$ in this parametrization, finite aggregate income requires $\alpha > 1+\epsilon$, and the condition then fails at $\bar{\lambda} = 0$: with no social weight at the top, the redistributive effect dominates and marginal taxes decrease with the degree of asymmetry. With moderate weights the ranking reverses; for instance, with $\alpha = 2$ and $\epsilon = 0.3$, any $\bar{\lambda}$ above 0.54 makes marginal retention decrease in δ: the Pigouvian effect dominates, and marginal taxes increase with the degree of informational asymmetry.