Dynamic Job Market Signaling and Optimal Taxation

André Sztutman*

MIT

December 6, 2022

For the latest version, click here

Abstract

How are optimal taxes affected by reputation building and imperfect information in labor markets? In this paper, I build a model of labor markets with incomplete and asymmetric information where job histories play a crucial role in transmitting information about workers’ productivity, which allows us to better understand the efficiency and distributive consequences of imperfect monitoring and screening in labor markets, and the tradeoffs the government faces when setting taxes. Optimal taxes are described by generalized versions of standard redistributive and corrective taxation formulas, which depend crucially on labor wedges: the ratios of the marginal contribution to output over the increases in lifetime earnings that result from supplying one extra unit of labor at each period. Combining estimates from the literature and new estimates using data from the Health and Retirement Study, I find that, the corrective component of taxes is large, especially at the top of the income distribution.

*MIT Department of Economics. Email: sztutman@mit.edu. I am grateful for the guidance and support of my advisors Jonathan Gruber, Robert Townsend and Iván Werning. I also thank Daron Acemoglu, Alessandro Bonatti, Oguzhan Celebi, Alden Cheng, Jonathan Cohen, Arnaud Costinot, Peter Diamond, Dave Donaldson, Lucas Finamor, Joel Flynn, Nicole Immorlica, Caio Lorecchio, Anh Nguyen, James Poterba, Joseph Shayani, Diana Svertlin, John Sturm, Martina Uccioli, Christian Wolf, Bryant Xia, MIT Macro, Public Finance, and Applied Micro Lunch and Seminar participants for their comments and suggestions.
1 Introduction

It takes time for a highly productive worker to be recognized as such in labor markets. For a worker with little experience, it is hard for employers to infer their abilities from their short resumes. However, as workers advance in their careers, increasing the length of their resumes, they make information about their abilities available to potential employers. In line with this idea, there is evidence that measures of ability that are not observed by firms become increasingly more predictive of workers’ salaries as they progress in their careers.\(^1\)

These informational imperfections drive workers to exert effort throughout their careers. In the beginning, foreseeing higher salaries in the future, they work hard 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, allowing workers to reap the benefits of their past effort. Those dynamic career concerns may not in general induce workers to exert the efficient amount of effort.\(^2\)

Besides shaping the incentives to exert effort, these informational imperfections affect the income distribution. Whenever employers cannot differentiate among workers of distinct productivities at some point in their careers, those workers may receive the same wages. These informational imperfections, thus, may make the higher productivity workers implicitly subsidize the lower productivity workers. Because these informational asymmetries affect incentives and the distribution of income, it is natural to ask what can be done to counterbalance those effects, and how the presence of these dynamic informational asymmetries affects the efficiency and redistribution tradeoffs the government faces when setting taxes.

This paper proposes a simple model of career concerns, building on an otherwise standard dynamic model of labor supply and demand, that addresses the connection between information transmission in labor markets, taxes, and inequality. The model builds on the signaling logic from Spence (1973), shifting the focus from one-time investment decisions on education to the dynamics of job market experience accumulation and the effort decisions throughout the lifetime of a worker. Concretely, the model is built around two ideas. First, it is easy for the firms to see and pay for the execution of clearly specified tasks, or deliverables, but it is much harder for them to assess the individual contribution of each worker to the firm’s total output, or the value of the firm. Second, firms assess the productivity of workers by looking at the resumes, which parsimoniously summarize the history of deliverables the

---

\(^1\)For example, Armed Forces Qualification Test (AFQT) exam scores become increasingly more predictive of workers’ salaries (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007).

\(^2\)Indeed, Holmström (1999) shows that dynamic career concerns introduce labor supply distortions and may only partially replace the incentives that pay-for-performance contracts could provide for workers to exert effort.
workers have produced so far in their careers.\footnote{The model is a counterpart of the Arrow (1962) learning-by-doing model, where instead of affecting human capital, the history of completion of deliverables affect, in equilibrium, employers’ perceptions of the worker productivity. This connection is especially clear in the special case where the resume is defined as the total experience, or the cumulative sum of deliverables a worker has produced. For this reason, we can call it a “signaling-by-doing model.”}

In equilibrium, working hard has dual benefits: it generates larger payments today, and it also establishes workers’ reputations, signaling their productivity to employers. The logic behind this is the following: employers do not know the productivities of the workers they are hiring, but they can see how many tasks they have completed. Under the assumption that for those who are productive it is less costly to complete deliverables, the firms can infer that those who have longer resumes are also more productive. This generates a “rat race,” and pushes workers to exert additional effort to signal to employers their productivities. While doing that, they build up their resumes, and progressively earn more. There is a positive return to experience even without human capital accumulation, which is entirely due to signaling.

The simple structure of the model will also allow us to derive comparative statics and to incorporate and understand the dual role of nonlinear taxes in correcting for the labor market imperfections and redistributing income. This understanding is brought by a series of results that i) delineate what the optimal tax base is, ii) write optimal tax formulas in terms of simple sufficient statistics that can be estimated and compared to standard optimal tax formulas, which would ignore the role of information imperfections, and iii) derive comparative statics on taxes and welfare from changes in the degree of information asymmetries.

Optimal lifetime income taxes are described by a generalization of standard Mirrleesian nonlinear taxation formulas. This generalization accounts for the labor market imperfections, and introduces a Pigouvian component to the standard formulas. Intuitively, optimal taxes can be thought of in two steps. First, they correct for the informational friction at the margin where it matters by making sure that, for any extra unit of effort, the marginal benefits to workers is equal to their marginal product of labor. Then, simultaneously, redistributive taxes are imposed on top of these taxes following standard Mirrleesian formulas, where the costs and benefits of redistribution are expressed in terms of the shape of the income distribution, compensated and income elasticities, and welfare weights.\footnote{Moreover, these generalized formulas apply independently of the model of competition with imperfect information or of the source of the friction that makes workers to receive salaries that are not their marginal products, provided some simple conditions are satisfied, pretax earnings are an invertible function of the supply of deliverables, and firms make zero profits.}

Surprisingly, in a simple benchmark case, to correct for the career concerns distortion is not necessary to keep track of the ratio of marginal productivities over salaries throughout
the lifetime of the worker.\textsuperscript{5} It is enough to look at this ratio only around the time of retirement. Moreover, under the assumption that the willingness to provide deliverables and the unobserved productivities are positively correlated, and holding constant the usual sufficient statistics for optimal taxes (elasticities and the distribution of income), taking career concerns into account pushes towards higher marginal taxes. The reason behind this is the following: employers do not know who are the employees retiring at the time of contracting, so those who are retiring are offered the same wage as those who have the same experience but are more productive and are willing to work even more. By the assumption that the willingness to complete deliverables and the unobserved productivities are positively correlated, those who are retiring are the least productive in the pool of workers with the same resume. This explains the wedge between their salaries and marginal product. This wedge, more precisely one minus the wedge, can be interpreted as the negative externality of the “rat race”. That is, holding the remuneration schedule of firms constant, if a worker would work more and earn one more dollar, they would generate less than one dollar of output for the firm. This difference can be thought of as a negative production externality, and can be corrected with Pigouvian taxes. For these reasons, taking into account these dynamic informational imperfections pushes towards higher marginal taxes over lifetime income.\textsuperscript{6}

Recent technological changes are reshaping how workers are monitored and screened.\textsuperscript{7} This paper shows that the elimination of informational frictions is a force that: i) pushes lower marginal corrective taxes, ii) pushes towards higher marginal redistributive taxes, iii) has a generally ambiguous net effect on taxes, but in sensible simple cases the corrective component dominates, and iv) it is a force that hurts society from the point of view of a redistributive planner. On one hand, the corrective taxes diminish because, as the strength of the “rat race” diminishes, it takes lower taxes to correct for the fact that workers work too much over their lifetimes. On the other hand, because better information benefits disproportionally more the more productive workers, the shape of the income distribution may change

\textsuperscript{5}Nor it is sufficient: another simple benchmark case shows that salaries can be equal to marginal productivities at all points in the career of a worker, but the timing of labor supply decisions can be severely distorted.

\textsuperscript{6}Optimal taxes do not depend only on labor wedges but also on elasticities and the shape of the income distribution. This paper shows that from the point of view of optimal taxation formulas, similarly to results in Scheuer and Werning (2017) common elasticity estimates are biased downwards. The reason for it is that increases in marginal retention induce the marginal types – who are the less productive types – to work more, and therefore reduce pretax salaries, making the effective change in post-tax wages smaller. Thus, the estimated elasticities of taxable income are lower in magnitude than the elasticities that enter optimal taxation formulas, which keep pretax salaries fixed. This is a countervailing force that makes accounting for career concerns push less towards higher marginal taxes.

\textsuperscript{7}These changes appear in the form of increasing availability of data and new tools to analyze it (Chalfin et al., 2016; Autor, 2019; Acemoglu et al., 2020; Bales and Stone, 2020), and changes in task composition of jobs from the automation of routine tasks (Autor et al., 2003), and the advent of “new work” (Autor, 2019).
and push towards a higher redistributive component of taxes. Relatedly, whenever redistribution towards the poorer is valued by society, society is hurt when there is less information asymmetry because it becomes easier for high productivity workers to separate themselves from lower productivity workers, undoing part of the implicit cross subsidies that happen in labor markets from those high productivity workers towards the lower productivity workers.

From an empirical point of view, to calibrate the optimal taxation formulas, a key statistic that needs to be estimated is the ratio of the marginal productivity of workers at retirement over their salaries, as a function of their lifetime income. In principle, this labor wedge could be inferred from the shape of the lifetime income distribution and from the signaling component of the return to experience. Although there are estimates for the shape of the income distribution and the growth rate of salaries across the income distribution (for example, Guvenen et al. (2021)), decomposing this growth rate into its signaling and human capital accumulation is a considerable challenge. There are not estimates for each of those in regards to the return to experience, but a growing literature has decomposed the return to schooling into its different components. A simple calibration exercise combining estimates of the growth rate of salaries across the income distribution from Guvenen et al. (2021) and assuming that the signaling share of the return to experience is equal to the signaling share of the return to schooling from Aryal et al. (2019), indicates that the Pigouvian component of taxes could be as high as 25% for top earners, and on average would be around 6%.

To get more direct estimates of the signaling component of the return to experience and the wedge between salaries and marginal productivities at retirement, this paper develops an empirical strategy relying on tax changes as a source of exogenous variation in wages and using data from the Health and Retirement Study survey. This empirical strategy adapts results from the literature that has quantified the degree of adverse selection in insurance markets by leveraging exogenous variation in prices (as in Einav et al. (2010); Einav and Finkelstein (2011); Cabral et al. (2022)). In the context of imperfect information and career concerns in labor markets, the key idea is that with a source of exogenous variation in wages for a specific labor contract, one can non-parametrically trace the shape of the labor demand curve by looking at average productivities as a function of salaries. However, inferring productivities in labor markets is a difficult endeavor. To circumvent that challenge two complementary approaches are adopted.

The first strategy is to assume that, although there are informational assymetries, labor markets are competitive and, thus, we can treat hourly salaries as the average marginal productivity of workers. Under this assumption by observing individual salary changes before and after a tax change, and observing the number of people who retired in response to the tax change, we can infer the productivity of those who were originally almost indifferent
between retiring or not. Applying this strategy to data from the Health and Retirement Study survey shows that, for an average worker, the Pigouvian component of taxes is of the order of 5%, while for high earners it ranges from 10% to as high as 60%. Since optimal taxes can be thought of as the product of a Pigouvian and a Mirrleesian component, the fact that the Pigouvian component is so high implies that the redistributive component of taxes is potentially quite small or close to zero. In other words, the current tax system is significantly less redistributive than otherwise might be thought, if imperfect information was not taken into account.

The second strategy leverages the rich set of questions asked in the Health and Retirement Study and allows for more direct tests for the mechanism highlighted in this paper. In particular, the Health and Retirement Study includes data on cognitive scores, assessed at each interview from questions involving counting, naming, and vocabulary tasks. We find that the pool of individuals who keep working after a tax increase has better cognitive scores (as measured before the tax change), in line with the idea that the changes in pretax salaries induced by tax changes are due to selection, or changes in the composition of the pool of individuals who are still working. This effect is also larger at the top of income distribution, in line with the idea that those informational imperfections are more pronounced for high-paying occupations.

To add realism to an otherwise stylized model, several extensions to the basic model are presented, including on-the-job learning, richer heterogeneity in elasticities of labor supply, and richer signal structures. This paper shows that, in these extensions, the key insights from the generalized optimal taxation formulas and their empirical implications still hold with minor caveats. When there is human capital accumulation on the form of on-the-job learning, the return to experience features both a signaling and human capital accumulation component; however the same optimal lifetime income tax formula applies. When there is rich heterogeneity in elasticities, the average labor wedge at each lifetime income level should be weighted by those lifetime income elasticities. When resumes include richer exogenous signals which cannot be observed by the government, the benefits and costs in optimal tax formulas should be weighted by the sensitivity of post-tax salaries to tax changes. When the signal the firm sees is a richer function of the history of deliverables, there may be additional distortions to be corrected, but the same optimal lifetime income tax formula applies.

Related Literature

This paper is related and contributes to several strands of the literature, including a public finance optimal taxation literature, the empirical and the theoretical literature on
imperfect information in labor markets, and an empirical literature on dynamic labor supply choices.

First, the taxation results in this paper build on the optimal taxation literature that goes back to the seminal contributions of Mirrlees (1971); Diamond (1998); Saez (2001), and more precisely contribute to a growing literature on optimal taxation with richer models of labor markets (Hariton and Piaser, 2007; Stantcheva, 2014, 2017; Bastani et al., 2015; Scheuer and Werning, 2016, 2017; da Costa and Maestri, 2019; Costinot and Werning, 2018; Craig, 2020; Hummel, 2021; Guerreiro et al., 2022). Methodologically perhaps the closest papers are Scheuer and Werning (2017) and Scheuer and Werning (2016), who show that standard optimal taxation formulas apply quite generally, including a broad range of models where wages are endogenous. Relative to these papers, this paper adds further generality to optimal taxation formulas, enriching them to cover situations where labor market frictions introduce labor market distortions. That is, the generalized taxation formulas in this paper hold when there are additional labor market inefficiencies, independently of their nature. The results on welfare in this paper speak to welfare theorems for economies with informational frictions from Prescott and Townsend (1984) and generalize results from Stantcheva (2014). In the latter, welfare comparisons are drawn between economies with “double adverse selection” – in the form of non-linear screening as in Miyazaki (1977) – to economies where firms know the productivity of workers. This paper extends the comparison to more general and arbitrary frictions, as well as to other intermediary levels of informational frictions.

Second, this paper contributes to the literature on imperfect information in labor markets, which goes back as far as the seminal contributions of Spence (1973), to the models of the “rat race” as in Akerlof (1976) and Miyazaki (1977), and in Stantcheva (2014), and the dynamic career concerns model of Holmström (1999), extended and further analyzed more recently by Bonatti and Hörner (2017), Cisternas (2018), and Hörner and Lambert (2021). Relative to this literature, this paper provides a new model of dynamic signaling combining elements from both of these classes of models. The assumption in this paper, that firms see and pay for the execution of clearly specified tasks, or deliverables, borrows from static competitive screening models of the labor markets as in Miyazaki (1977), and bypasses a key limitation from the canonical career concerns setup in Holmström (1999), where firms can see the individual contribution that each worker makes to the firms’ profits but cannot pay for performance. Conversely, the assumption that firms assess the productivity of workers by looking at resumes borrows from Holmström (1999), and bypasses a key limitation from the static setup in Miyazaki (1977), where firms learn the productivity of the workers through a one-time interaction, and resumes play no role in transmitting information. Furthermore, the idea that firms see a simple public signal builds on the motivational rating setup from...
Hörner and Lambert (2021), and bear resemblance to aggregation and linearity results from Holmstrom and Milgrom (1987).

Third, a related, but more empirically focused literature, has looked at how firms learn about the productivity of workers and whether there are information asymmetries in labor markets, including Jovanovic (1979); Farber and Gibbons (1996); Acemoglu and Pischke (1998, 1999); Altonji and Pierret (2001); Lange (2007); Kahn and Lange (2014); Cella et al. (2017); Aryal et al. (2019). Relative to this literature, this paper provides new evidence for the importance of informational asymmetries in labor markets, in particular for workers in later stages of their careers. This complements the evidence from Kahn and Lange (2014) who found that firms have substantial uncertainty over the productivities of older workers. This paper also contributes to the literature on technological changes and their impacts on labor markets (Autor et al., 2003; Brynjolfsson and Mitchell, 2017; Brynjolfsson et al., 2018; Autor, 2019; Acemoglu et al., 2020; Acemoglu, 2021; Autor et al., 2022), delineating key welfare and normative implications from changes in the technologies for monitoring and screening workers.

There is a large literature that has looked at dynamic labor supply decisions, human capital accumulation and on-the-job learning (Heckman, 1976; Eckstein and Wolpin, 1989; Shaw, 1989; Altuğ and Miller, 1998; Keane and Wolpin, 2001; Imai and Keane, 2004; Keane, 2011; Altonji et al., 2013). One key insight from that literature is that workers, when making their labor supply decisions today, would consider the impact of those decisions today on their future salaries, and that in this context the workers’ opportunity cost of time may not be equal to their current wages. The same effect is present in this paper, where the dependence of future salaries on current effort decisions are alternatively explained by signaling effects.

The remainder of the paper is structured as follows. Section 2 presents the dynamic signaling model. Section 3 presents a simple example, and discusses intuitions and properties of the model. Section 4 introduces taxes and presents the main normative results. Section 5 discusses the existing empirical evidence, the empirical strategy, and the empirical results. Section 6 discusses extensions, including human capital accumulation, heterogeneous elasticities, and richer signal structures. Section 7 concludes.

2 Signaling-by-doing model

This paper adapts a standard dynamic model of labor supply, demand and taxation, by adding the constraint that firms have limited information over the workers’ productivities and can only contract based on a subset of the observed activities workers perform. The
general setup encompasses a standard neoclassical life-cycle labor supply problem, and with minor modifications can encompass several models of screening and imperfect information in labor markets as special cases. These special cases include a) models where individual contributions to output are observed and workers get paid a fixed salary independent of the realized output, as in Holmström (1999), as well as b) models where hours are observed but output is not, and workers are screened through the total amount of hours or observable effort they commit to offer, as in Miyazaki (1977) or in Akerlof (1976). For exposition, we focus on one specialized version of the general model that allows to derive clear comparative statics and simple optimal tax formulas, while retaining the essential economic assumptions that describe dynamic job market signaling. That is, employers will see a worker’s resume – how much in terms of deliverables a worker has provided so far in their career – and will pay workers for the execution of these deliverables. Workers, aware of how their resumes will be read, will choose their labor supply balancing costs and benefits in terms of current and future wages. Section 6 will consider several departures from this simple model, including richer signal structures and richer heterogeneity in preferences.

2.1 Preferences and Technology

The household block of the model consists of a continuum overlapping generations of workers who live for a continuum of periods going from zero to one. These workers have arbitrary preferences over labor and consumption flows, and are forward-looking: they understand that their labor supply choices can affect the information firms will have about them in the future. The production block of the model is described by competitive firms with linear production functions.

More formally, workers are indexed by their types $\theta$, which determines their productivity and their preferences, and there are different cohorts of workers. Each worker lives for a continuum of periods that goes from zero to one, where zero corresponds to the time the worker is born and one corresponds to the time their life ends. At each period, workers of all ages coexist. They supply labor ($\tilde{h}(\cdot)$) and consume ($\tilde{c}(\cdot)$) at each period. That is, $\tilde{c}(\cdot)$ denotes the flow of consumption function $\tilde{c} : [0, 1] \mapsto \mathbb{R}^+$, and $\tilde{h}(\cdot)$ denotes the flow of labor supply function $\tilde{h} : [0, 1] \mapsto \mathbb{R}^+$. An individual of type $\theta$ has a productivity $v(\theta) > 0$, and production is linear, that is the flow of production is equal to the product of productivities and the labor supply $\tilde{h}$. Preferences are denoted $U(\tilde{c}(\cdot), \tilde{h}(\cdot), \theta)$, where $\tilde{c}(\cdot)$ is the time-flow of consumption and $\tilde{h}(\cdot)$ is the time-flow of labor supply.

---

8 Although it is not necessary and it is a especially restrictive in the context of continuous time choices, assuming time-separability of preferences can help us understand the trade-offs workers face when deciding their labor supply flows. For an example with time-separability, see Section 4.3.
The worker problem is standard: they maximize a utility functional, subject to a lifetime budget constraint, where flows in the future, at age \( a \), are discounted at the rate \( q(a) \). However, salaries \( w \) depend on the information the firms have about the worker \( I(\tilde{h}(\cdot), a, \theta) \), which will be specified below, but more generally could be a function of the flow of labor supply across all the periods \( \tilde{h}() \), the type \( \theta \) and age \( a \) of the worker. This captures the possibility that the workers may want to change their labor supply to influence their future salaries, by changing the employers’ perceptions of their productivities.

\[
V = \max_{\tilde{c}(\cdot),\tilde{h}(\cdot)} U(\tilde{c}(\cdot),\tilde{h}(\cdot),\theta) \text{ st. } \int_0^1 q(a)\left(\tilde{c}(a) - w(I(\tilde{h}(\cdot), a, \theta))\tilde{h}(a)\right)da \leq 0 \quad (1)
\]

Notice that a standard life cycle labor supply problem features as a special case of this, where wages are equal to the productivity of the worker, and this productivity would be independent of their labor supply decisions. Notice as well that human capital accumulation of the form of on-the-job learning would generate analogous concerns for the worker, with labor supply decisions affecting future wages through real increases in their productivities.

To illustrate the mechanics of reputation building, here we have abstracted away from the fact that there are taxes, and that the government could use taxes to shape incentives and redistribute income. Those taxes will be introduced again in section 4.

### 2.2 Contracts and Information

Firms are constrained to offer infinitesimal contracts, which is a contract for one unit of the labor supply. Firms are unable to commit to long-term contracts; they cannot promise that they will not try to renegotiate labor contracts once more information becomes available. This makes the firm problem essentially static. Finally, there is free entry and exit, therefore firms make zero profits.

Firms do not observe types or productivities, but instead, they observe a signal of their past experience. In the simplest case, we can think of that signal as how much labor a worker has supplied so far, which is denoted by \( I(\tilde{h}(\cdot), a, \theta) = h(a) = \int_0^a \tilde{h}(\tilde{a})q(\tilde{a})d\tilde{a} \), or the length of the resume when the worker has age \( a \). More generally, as covered in Sections 4.2 and 6, \( I(\tilde{h}(\cdot), a, \theta) \) can accommodate richer signal specifications, including more general functions of past experience \( I(\tilde{h}(\cdot), a, \theta) = \int_0^a \phi(\tilde{a}, a)\tilde{h}(\tilde{a})d\tilde{a} \) (with \( \phi(\tilde{a}, a) > 0 \), continuous in both arguments), or exogenous signals \( I(\tilde{h}(\cdot), a, \theta) = (h(a), z(\theta)) \). The more general idea is

---

9These discount rates are assumed to be exogenously given, i.e. there is a technology for transferring resources across periods at rates \( q(\cdot) \), and that these rates are such that budget and resource constraints are well defined; that is, the present value of resources in the economy is finite. While the first assumption is not essential, the fact that the present value of resources is finite is important to guarantee that the economy is dynamically efficient.
that the resume is an imperfect measure of the past history of the completion of deliverables and their timing.\textsuperscript{10}

An important point to emphasize is that we should interpret what has been referred to as labor supply $\tilde{h}(\cdot)$ not necessarily as hours or effort, but as what will be referred to in this paper as deliverables. That is, $\tilde{h}(\cdot)$ represents the specific piece of the information on which the firms can condition their contracts. This will allows us to consider the possibility that changes in technology make these deliverables a better or worse measure of output. In the extreme case where $\tilde{h}(\cdot)$ is the flow of output, there is no information asymmetry problem between the firms and workers, and workers will get paid for their marginal contributions to the value of the firm output.\textsuperscript{11} Firms, moreover, do not observe output, or cannot individually assess the contribution of each worker to output, or the firm’s profits. They can, however, infer the expected productivity given the signal of experience, either because they have hired many employees with the same experience and have seen how much output these employees have generated on average or because they know the economy-wide distribution of productivities and labor supply.\textsuperscript{12}

Figure 1 represents diagrammatically the flows of production, payments, and information. Workers complete deliverables for the firms, and the firm sees the completion of the deliverables and the workers’ resume. Each worker’s resume keeps track of their deliverables

\textsuperscript{10}This detailed history, if it was perfectly observed by firms and if workers’ types are single-dimensional, in most sensible cases, would contain enough information to allow firms to infer the workers’ type almost instantaneously. While explicitly introducing stochastic noise on preferences or on the information could equivalently capture this imperfect inference problem, this formulation avoids dealing with technical difficulties arising from multidimensional screening problems, where in this case both the type space and the space of goods would be high-dimensional. This formulation can also flexibly accommodate other concerns that would lie outside the immediate scope of the model, such as information becoming harder to retrieve for work experiences that are further in the past, or firms assigning disproportional weights to some experiences because of limited attention or because they have the wrong model of how the economy works and mistakenly attribute increases in productivity to human capital accumulation instead of signaling. Special cases of this formulation include observing the total experience $\int_{0}^{a} q(\tilde{a}) \tilde{h}(\tilde{a}) d\tilde{a}$, and the pace of experience accumulation $\int_{0}^{a} \frac{\tilde{h}(\tilde{a})}{a} d\tilde{a}$. Section 4.3 includes two simple examples where the signal is defined as either the total experience or the pace of experience accumulation.

\textsuperscript{11}To match this setup to canonical models of imperfect information in labor markets notice that, in the Miyazaki (1977) model the only difference would be that the information set of firms would consist of not how much a worker has supplied so far $h(a)$, but how much the worker would supply over the lifetime $h(1)$. In the Holmström (1999) model, labor supply $\tilde{h}(a)$ would be a two-dimensional vector of effort and an indicator function for whether the individual decides to work at any period. Workers would get paid only by the second component of $\tilde{h}(a)$. The information set of firms would consist of a stochastic function of the history of effort and the type of the agent.

\textsuperscript{12}The fact that firms observe hours but not individual output can also be thought of through the lens of team production (Alchian and Demsetz, 1972). Production needs to take place inside a firm, that aggregates the work of multiple workers. For example, the firm production function could be $F=\prod_{i=1}^{n} \mathbb{1}(h_{i} > 0) \cdot \sum_{i} h_{i} \cdot v_{i}$, that is, production inside the firm is linear but it needs multiple ($n$) workers to be present. In the limit of a large number of workers ($n$), the firm would assess the productivity of workers only from their expected productivity.
completion history, adding those deliverables up. Payments are based on what the firm observes, that is, resumes and the execution of the deliverable. Output and profits are realized, but the firm cannot individually assess the contribution of each unit of deliverable to the firm output.

### 2.3 Equilibrium Definition

Equilibrium is described by workers choosing optimal consumption and labor supply flows taking salaries as given, and anticipating the effect that their labor supply decisions have on their future salaries, as stated in 1, while firms simultaneously set salaries according to the zero profit condition 2, presented below, paying each worker for their expected productivity conditional on their resumes. This zero profit condition is justified by competition among firms to enter the market and hire labor.

**Definition.** For each marginal unit of labor, the firms’ free entry condition is described by an Akerlof (1970) lemons condition, where salaries are equal to productivity of the workers with the same resume:\(^{13}\)

---

\(^{13}\)In this definition, it is assumed that \( I \) and \( \tilde{h}(\cdot) \) are continuously distributed, so that the expectation is the same if it is conditioned on the workers who accept the contract.
\[ w(I(\tilde{h}(\cdot), a, \theta)) = \mathbb{E}[v(\theta)|I(\tilde{h}(\cdot), a, \theta)] \]  

Equation 2 lies at the core of the model. It simply states that at any period people with equivalent resumes will be paid equally, and that firms will on average break-even.

Equilibrium, thus, is described by workers choosing optimal consumption and labor supply flows taking salaries as given, and anticipating the effect that their labor supply decisions have on their future salaries, as stated in 1, while firms simultaneously are setting salaries according to the condition 2, paying each worker for their expected productivity conditional on their resumes.

3 Positive analysis

Here, we evaluate some of the positive implications of the model presented in the previous section. We first focus on the case where the public signal that the firms see is defined as the “length-of-the-resume”.

**Lemma 1.** If the information of firms is defined as the “length-of-the-resume”, that is, \( I(\tilde{h}(\cdot), a, \theta) = \int_{a}^{a} q(\tilde{a})\tilde{h}(\tilde{a})d\tilde{a} \), then the Akerlof lemons condition (Equation 2) is equivalent to:

\[ w(h) = \mathbb{E}[v(\theta)|h(\theta) \geq h] \]  

**Proof.** See Appendix Section A.1.

Where \( h(\theta) \) denotes the length of the resume of type \( \theta \) at the end of their career, that is \( h(\theta) \equiv h(1, \theta) \). Lemma 1 states that salaries as a function of the length of resume \( h \) are the average productivity of all the types who eventually reach a point in their careers where their resumes are longer than \( h \). This is the case, because at any period, for every type who eventually reaches a longer than \( h \) resume, there is someone, potentially from a different generation, who has a resume of length \( h \) today.\(^{14}\) If workers’ labor supply decisions are heterogeneous only at the extensive margin (differing in how long they stay in the labor force) or if the deliverables were defined as working additional years, then the set of workers that would be pooled together would consist only of workers of the same cohort, as in the first example from Section 4.3. Section 4.2 analyzes the cases where the information set of firms is more flexibly defined, as well as cases where the detailed timing of deliverables is observed and no inter-cohort pooling is allowed.

\(^{14}\)The same argument could be applied to any specification of \( I \) of the form \( I(\tilde{h}(\cdot), a, \theta) = \int_{a}^{a} \phi(\tilde{a})\tilde{h}(\tilde{a})d\tilde{a} \).
Given equation 3, which describes salaries as a function of lifetime income, it is useful to note that lifetime income can be written quite simply. We can change variables to express the worker lifetime income as

\[ y(h) = \int_{0}^{h} w(z) dz, \]

that is we can express lifetime income as instead of integrating over time, as integrating over increases in the length of the resume \((dz)\). Its derivative, or how much more the worker would receive for providing one extra unit of the deliverable over their lifetime, is equal to the payment for the last unit of the deliverable, that is \(y'(h) = \mathbb{E}[v(\theta) | h(\theta) \geq h]\).

Equation 3 helps simplify the analysis behind this dynamic model of career concerns, making it as simple as a static model. In the economically-sensible case where it is less costly for the more productive people to supply the deliverables, it will be shown that salaries increase with experience. Workers at the beginning of their careers are willing to work more, relative to the myopic trade-off between current salaries and current effort, because they expect higher future salaries as an outcome of building up their experiences, signaling to employers that they have higher productivities. The property that exerting effort increases future salaries, by affecting employers’ perceptions of the ability of the worker, will be shared by a large range of informational structures, as described in more detail in Section 4.2.

As in Arrow (1962) learning-by-doing model, where productivity is a function of the cumulative use of a factor, firms’ perceptions about each workers productivity in this model are also a function of their cumulative labor supply. Relatedly, firms could mistake the return to experience as human capital accumulation instead of changes in the composition of the pool of workers. That is, the firm could hire many workers with different lengths of resumes, and observe that the pool of workers with longer resumes is, on average, more productive. The firm does not need to know why the pool is more productive, attributing it to human capital accumulation instead of signaling, without changing the model predictions. In other words, firms could think the change in the composition of workers who happen to achieve a higher degree of experience is a real return to human capital when it is not, and still post the same salaries and make zero profits.\(^{15}\) Worker as well may not know whether they will become more productive by working more, or whether they would just signal to employers they are more productive. From their perspective, to make their labor supply choices, it only matters how experience accumulation and labor supply decisions today will likely impact their future salaries.

Given this structure of labor demand and supply, a natural question in this setup is whether there exists an equilibrium, as defined in Section 2.3. Proposition 10 in the Appendix Section A.3, shows that it does exist under some standard assumptions.\(^{16}\)

\(^{15}\)A more general version of the model that has both ingredients is presented and discussed in Section 6.

\(^{16}\)Namely, these assumptions state that the distribution of types, productivities and marginal rates of
3.1 An Illustrative Example

In this section, a simple example illustrates basic mechanics of the model, where workers enter a “rat race” building up their resumes and working excessively over their lifetimes. The example also shows that, in this case, although for each worker salaries increase over time, intertemporal labor supply decisions are not distorted. Further, inequality increases when information becomes more symmetric.

At the beginning of their working life, for the first job they can get, all individuals face the same wage. Employers cannot distinguish between workers who have no experience at all. As worker advance in their careers, completing tasks and increasing their lifetime supply of labor, the length of the resume of hard working individuals works to separate them from the other workers that execute fewer tasks and have shorter resumes. These more productive workers initiate their career subsidizing the less productive, but at each new task they execute, some less productive workers are left behind with a shorter resume. For this reason, the remuneration that the more productive workers receive for the execution of tasks becomes progressively higher. More generally, as will be shown in Section 4.2, there is a positive signaling return to experience whenever it is easier for more productive workers to complete deliverables, and the signal the firms see is increasing in the completion of deliverables.

The size of the return to experience depends on how many people are being left behind by these more productive workers as they advance in their career, and on how much more productive they are relative to those with shorter resumes. That is, the return to experience depends on the joint distribution of preferences and productivities.

Fortunately, there are simple and sensible assumptions on preferences and heterogeneity that allow us to derive equally simple expressions for the return to experience and the shape of the income distribution, as well as to analyze how they would change in response to information becoming more symmetric. Towards this goal, let’s assume that preferences over lifetime labor supply and lifetime consumption are such that there is a constant elasticity of substitution are continuous, and that marginal rates of substitution are smooth as a function of lifetime consumption and lifetime income, with bounded derivatives.
lifetime labor supply and those preferences are quasilinear in lifetime consumption.\footnote{Note that given the result from Proposition 1, and assuming exogenous discount rates, it is without loss to specify preferences in terms of lifetime labor supply and lifetime consumption. Section 4.3 presents two additional simple examples that further clarify the mechanics of this model, one with heterogeneity only at the extensive margin (when to retire), and the other with time-separable preferences, and heterogeneity only at the intensive margin.}

\[
U(c, h, \theta) = c - \left( \frac{h}{b(\theta)} \right)^{1 + \frac{1}{\epsilon}} \left( 1 + \frac{1}{\epsilon} \right)^{-1}
\]

where \( b(\theta) = \theta^{1-\delta}, v(\theta) = \theta^\delta, \theta \sim \text{Pareto with shape parameter } \alpha > 1, \) and \( 0 \leq \delta < 1. \) In this example, \( \delta \) governs the amount of information asymmetry: a higher \( \delta \) means more heterogeneity comes from unobserved productivities \((v(\theta))\) instead of observable productivities \((b(\theta))\). Under this formulation \( b(\theta) \) should be thought of as how many deliverables a worker can provide per unit of effort \( l \), that is, \( h = l \cdot b(\theta) \). \( v(\theta) \) should be interpreted as how much output the worker generates per unit of the deliverable, that is, \( y = h \cdot v(\theta) \). The total productivity as a function of effort then is just the product of \( v(\theta) \) and \( b(\theta) \), and it is equal to \( \theta \). Using the equilibrium definition and the convenient properties of Pareto distributions, a worker of type \( \theta \) works less than all types above it, and faces a salary for their marginal unit of labor equal to

\[
w(h(\theta)) \equiv \mathbb{E}[v(\hat{\theta})|\hat{\theta} \geq \theta] = \frac{\alpha}{\alpha - \delta} \cdot \theta^\delta
\]

As a function of labor, we can guess and verify that salaries are also a power function, so wages and experience follow a log-linear relationship:

\[
\log(w) = \left( \frac{\delta}{1 - \delta + \epsilon} \right) \cdot \log(h) + \left( \frac{(1 + \epsilon)(1 - \delta)}{(1 + \epsilon)(1 - \delta) - \delta \epsilon} \right) \cdot \log\left( \frac{\alpha}{\alpha - \delta} \right)
\]

There is a constant proportional return to experience that is entirely driven by selection or employers learning about the types who are willing to take the jobs they are offering. This return to experience is larger when heterogeneity comes mostly from unobserved productivities, that is when \( \delta \) is higher, as there is more heterogeneity to be screened out by experience. When the elasticity parameter \( \epsilon \) is low, the return to experience is also larger: in this case, a higher experience is really indicating that the worker is more productive. When the elasticity is low, the difference between how a marginal increase in labor supply hurts the less productive relative to the more productive workers becomes larger.
### 3.2 General Properties

The example presented in the previous section was stylized, and relied on a particular assumption on the information structure and on parametric assumptions on preferences and productivities. This section shows that properties of salaries and the return to experience from the previous example hold under a broader set of assumptions.

Under the assumption that the public signal the firms look at is the cumulative discounted sum of deliverables, i.e. \( I(\tilde{h}(\cdot), a, \theta) = \int_{0}^{a} \tilde{h}(\tilde{a}) q(\tilde{a}) d\tilde{a} \), as in Lemma 1 salaries are equal to the expected productivity of the workers with the same resume \( w(h) = \mathbb{E}[v(\theta)|h(\theta) \geq h] \). For any preferences and distribution of productivities workers with above-average productivity starts their career earning less than their marginal product, as the average productivity is everyone’s initial salary. All workers have to climb the same career ladder and before they can show anything in their resume they are all indistinguishable for employers. Provided that those who have higher unobservable productivities are also those for which it is easier to complete deliverables, every worker, however, will end their career getting paid more than their marginal products: at the time of retirement, each will be ending up with the shortest lifetime resume and have the smallest productivity among everyone that has gotten as far as them in their career. In fact, the log return to experience can be described as the product of a shape parameter of the distribution of the length of resumes and one minus the ratio of productivities of the marginal worker and the average productivity of those who have longer resumes, that is:

\[
\frac{d \log w(h)}{d \log h} = \alpha_{h}(h) \cdot (1 - \chi(h)),
\]

where \( \alpha_{h}(h) = \frac{f(h)h}{1 - F(h)} \) and \( \chi(h) = \frac{\mathbb{E}[v(\theta|h(\theta) \geq h)]}{\mathbb{E}[v(\theta)|h(\theta) \geq h]} \). Intuitively, tells us how much salaries increase for the workers as a function of their experience from how many workers are being left out of their pool as these workers increase their experience, and how much more productive these workers who remain at the pool are relative to those who are being left out.\(^{18}\)

More generally, for information sets \( I(\tilde{h}(\cdot), a, \theta) = \int_{0}^{a} \tilde{h}(a) \phi(\tilde{a}, a) d\tilde{a} \), with \( \phi(\tilde{a}, a) > 0 \), bounded and continuous, workers start their careers earning the unconditional average productivity, so highly productive workers start earning less than their marginal productivity. At any point, assuming more productive types are more willing to provide the deliverables at all periods, exerting more effort increases future salaries. Workers at the peak of their careers (when they reach the highest value of the index \( \int_{0}^{a} \tilde{h}(\tilde{a}) \phi(\tilde{a}, a) d\tilde{a} \)) will earn more than

\(^{18}\)The same arguments can be applied if the signals are more generally as the cumulative weighted sum of deliverables, with arbitrary continuous weights \( \phi(\tilde{a}) > 0 \), i.e. \( I(\tilde{h}(\cdot), a, \theta) = \int_{0}^{a} \tilde{h}(\tilde{a}) \phi(\tilde{a}) d\tilde{a} \), and if more productive types are more willing to provide flows of deliverables.
their marginal productivity because they will be pooled together only with the types that are willing to provide higher amounts of deliverables across all periods.

Relative to Holmström (1999), this model imposes a different assumption on the contract space. In Holmström (1999) firms cannot pay for performance, firms do not observe effort but observe output, and overworking in a short period of time is a way for workers to trick employers into believing they are more productive. Here, firms must pay for the observable component of effort, the deliverables, and the fulfillment of these deliverables add to the resumes, which is a publicly available signal to employers. At the cost of other simplifying assumptions, the fact that firms in the model can condition payments on the performance measures that are available to them – as in fact many of them do – adds realism to the way that career concerns are modeled. Interestingly, employer learning under this assumption can be seen as a problem in itself instead of a side effect of a insurance and incentives tradeoff.

This matters not only from a positive perspective, but also from a normative point-of-view. Workers enter a “rat race” as in Akerlof (1976) and Holmström (1999), but interestingly, this “rat race”, in the case where resumes are defined as the cumulative discounted sum of deliverables, does not generate an intertemporal distortion, but rather, generates a lifetime labor supply distortion, as it will be shown more formally in the next section. Workers work too much, but the timing of their labor supply decisions is not distorted. For building a resume, in this simple information structure, there is no advantage of concentrating efforts in a particular period of their lives. The discounted sum of payments a worker receives from increasing the length of the resume by one unit is the same independently of when the worker decides to increase it, and at the same time the contribution to the total output of the economy from increasing the length of the resume by one unit is also the same independently of when the worker decides to increase it.

This also illustrates that the difference between current salaries and productivities does not in itself represent a distortion. Rather, distortions come from differences between productivities and the lifetime gains from exerting effort, where those lifetime gains are the sum of two components: current payments and increases in future payments from reputational effects. In this particular case, salaries are initially low but increase with experience, while the size of the distortion is constant over the lifetime of the worker. In Section 4.3, a simple example – where the resume instead of being defined as the cumulative sum of deliverables is defined as the pace of production of deliverables over time (i.e. $I(\tilde{h}(\cdot), a, \theta) = \int_0^a \frac{\tilde{h}(\tilde{a})}{\tilde{a}} d\tilde{a}$) – illustrates that the opposite case is also a possibility. Salaries can be exactly equal to the productivities of the workers at all times and, at the same time, there could be large intertemporal distortions coming from reputational effects and how effort today affects future payments when a worker is younger.

18
4 Taxation

In this section, we introduce a government and explain how it sets taxes. As in the previous section, we first focus on the case where the public signal that the firms see is defined as the “length-of-the-resume”, or the discounted cumulative sum of what the worker has produced, and then we extend the results to more general information structures.

As in Mirrlees (1971) the government does not observe the workers’ types, and because firms also do not observe the types, this model features what has been called “double adverse selection” (Stantcheva, 2014). We are going to place special focus on the case where the government sets taxes on lifetime income ($y$) to maximize a welfarist functional of utilities. This focus is motivated by two sets of reasons. On the practical side, many taxes and transfers actually condition on lifetime income: most notably, the key determinant of US social security benefits is a measure of average earnings over the lifetime of a worker. Furthermore, for many taxes the timing of earnings can be manipulated: the realization of capital gains can be delayed indefinitely to avoid taxes, and there is flexibility in reporting the timing of income coming from C-corporations. On the theoretical side, an extension of the Atkinson and Stiglitz (1976) result shows that it is still optimal to tax lifetime income even with “double adverse selection”.

Proposition 1. (AS extension): If preferences take the form $U(C(\tilde{c}(\cdot)), H(\tilde{h}(\cdot)), \theta)$, then it is optimal to tax lifetime income, even with “double adverse selection”.

Proof. See Appendix Section A.4.

In this Proposition $C(\tilde{c}(\cdot))$ and $H(\tilde{h}(\cdot))$ are common aggregators of flows of consumption and labor supply, respectively. The assumption on preferences says that if preferences across households are homogenous over the timing of consumption and labor supply decisions, that is, if preferences can be written as a function of common indexes $(C, H)$ that aggregate the flows of consumption and labor supply, and these indexes are the same across households, then it is always optimal to use only lifetime income as the tax base. It means that dynamic job market signaling, under the baseline assumptions here in this paper, does not introduces intertemporal distortions and under the benchmark case where preferences over timing of consumption and labor flows are homogeneous across households, there would be no reason to introduce further distortions on intertemporal decisions. This resembles a special case in Holmström (1999), where there are also no intertemporal distortions, namely when there is no discounting and productivities follow a random walk process. More generally, in Holmström (1999), reputation concerns in general induce workers to exert relatively more effort earlier in their careers, and progressively exert less effort as their reputations are consolidated.
This result echoes tax smoothing results as in Werning (2007a), and the tax smoothing ideas that go as back as Vickrey (1947), and adds another reason to use income averaging rules, as they are present for example in the calculation of social security benefits. Importantly, although the income base should be the lifetime income of a worker, those taxes do not need to be raised only at the end of a worker’s life, and can be raised annually, as long as taxes each year depend on the history of earnings of each worker up to the current date.

Moreover, the result is general in an important way. It essentially says that post-tax lifetime earnings, when it is possible, should be a function of lifetime labor supply. In the simple model we just presented, pretax lifetime earnings are a function of lifetime labor supply, and thus taxes on lifetime income are enough to guarantee that post-tax lifetime earnings are a function of lifetime labor supply. But in extensions of the model where pretax lifetime earnings would cease to be a function of lifetime labor supply (for example if the signal the firm sees about a worker is a different function of the flows of labor supply instead of the discounted sum of these flows as discussed in later in this section), then the optimal policy would entail taxes meant to undo these intertemporal distortions, and to make post-tax lifetime earnings a function of discounted lifetime labor supply.

### 4.1 Optimal Taxes

In this section, we derive necessary conditions for optimal taxes in terms of sufficient statistics, as in Saez (2001). We assume that the government maximizes a welfarist functional of worker utilities, $W(V(R;θ))$.\(^\text{19}\) The government solves:

$$\max_{R(y)} \mathbb{E}[W(V(R;θ))] \text{ s.t. } \mathbb{E}[y(h(R;θ), R) - R(y(h(R;θ)))] ≥ 0$$

and subject to the constraint that pretax wages are determined by the differential equation $y'(h) = \mathbb{E}[v(θ)|h(R;θ) ≥ h]$, with initial condition $y(0) = 0$.

Solving directly for $R(y)$ is complicated - changing taxes at an income level $y$ has cascading effects on salaries of everyone earning $y$ or more, by shifting the composition of workers. However, the following proposition allows us to simplify the problem by allowing the planner to keep salaries fixed when taxes change. In other words, we can frame the problem as the planner solving directly for post-tax salaries, effectively ignoring how pretax salaries are set.

**Lemma 2.** *Without loss, we can solve directly for $\bar{R}(h) = R(y(h))$, and then find $y(h)$, and $R(y)$. That is, the planner can solve the simpler problem:*

\(^{19}\)This formulation is quite general, and can be converted to Pareto efficiency tests as in Werning (2007b), by picking linear functionals of the form $W(V(R;θ)) = \lambda(θ) · V(R;θ)$, making it the dual of a revenue maximization problem subject to a minimum utility requirement for each type.
\[
\max_{\tilde{R}(h)} \mathbb{E}[W(V(\tilde{R}; \theta))] \quad \text{s.t.} \quad \mathbb{E}[v(\theta) h(\tilde{R}; \theta) - \tilde{R}(h(\tilde{R}; \theta))] \geq 0
\]

**Proof.** See Appendix Section A.5.

This result applies more generally to other models of labor market frictions provided that firms make zero profits,\(^{20}\) and, given any allocation, \(y(h)\) is well defined and invertible. An important example where these conditions hold is the Azevedo and Gottlieb (2017) model of competition with adverse selection. More generally, there could be other production externalities, imperfect competition generating compressed wages, or monopolistic screening by a single firm. This feature allows us to conceptually separate what are the relevant externalities coming from the information frictions that the planner would like to correct from what are just regular transfers and innocuous price adjustments. Importantly, the same idea will be applied in Section 4.2, to more general information structures, allowing us to separate issues of efficiency from issues of redistribution quite generally.\(^{21}\)

**Proposition 2.** (For single-dimensional \(\theta\)) If a tax schedule is optimal, then it satisfies the following optimal tax formula:

\[
\left(\frac{\chi(y) - r(y)}{r(y)}\right) \epsilon^c_r(y) g(y) y = \int_y^\infty \left(1 - \lambda(\tilde{y})\right) g(\tilde{y}) d\tilde{y} + \int_y^\infty \left(\frac{\chi(\tilde{y}) - r(\tilde{y})}{r(\tilde{y})}\right) \eta_I(\tilde{y}) g(\tilde{y}) d\tilde{y} \quad (4)
\]

**Proof.** See Appendix Section A.6.

where \(r(y)\) denotes the marginal retention function; \(\epsilon^c_r(y) = \frac{dy}{dr(y)}\) are the local compensated elasticities; \(g(y)\) denotes the density of \(y\); \(\lambda(y) \equiv \frac{W'(V)}{\mu}\) is the marginal value the planner places on transfer to a worker earning \(y\); \(\eta_I(y) = \frac{dy}{df} r(y)\) is the income elasticity; and \(\chi(y) \equiv \frac{v(y)}{y'(h(y))}\), and \(v(y)\) is the productivity of the worker with lifetime income \(y\).

This equation is almost the same as the standard first-order condition that appears in Saez (2001) and Werning (2007b), with one additional ingredient: \(\chi(y)\). The equation says that if a tax schedule is optimal, then three sorts of effects should balance each other: compensated effects, mechanical and welfare effects, and income effects. When a planner considers increasing marginal retention over a small region (holding post-tax salaries fixed everywhere else), there are compensated effects (left-hand side of the equation) coming from

---

\(^{20}\)Or, more generally, there is full taxation of profits, or, still, profits are uniformly shared between workers, as in Scheuer and Werning (2017).

\(^{21}\)Innocuous in the sense of the first and second welfare theorems, where changes in fundamentals could result in changes in relative prices which no planner that has access to individual-specific transfers would not like to undo.
the fact that people would work more. These effects are proportional to the densities and compensated elasticities, and they ease the feasibility constraint not by $1 - r(y)$ as in the standard Mirrlesian model, but by $\chi(y) - r(y)$, that is, a worker who earns an extra dollar in their lifetime generates $\chi(y)$ units of output per dollar earned (and retains $r(y)$). The mechanical and welfare effects (first term on the right-hand side of the equation) are the same as before: the planner is giving one dollar to those who earn at least $y$, and this has mechanical costs of one dollar per person and welfare effects that are weighted by the marginal value of a dollar that the planner places on each of these people. Finally, there are income effects that affect everyone who earns at least $y$. These people are induced to work less (if income effects are negative), and, as in the compensated effects, they damage the feasibility constraint by $\chi(y) - r(y)$.

The same equation can alternatively be written in two blocks: One block that translates the redistributive motive, and that looks exactly as the standard Mirrleesian formula, and a second block that defines retention as the product of two components, the Mirrleesian, and a Pigouvian component that corrects for the production externality:

$$
\left(\frac{1 - \rho_m(y)}{\rho_m(y)}\right) \epsilon^*_s(y) g(y) y = \int_y^\infty \left(1 - \lambda(\tilde{y})\right) g(\tilde{y}) d\tilde{y} + \int_y^\infty \left(1 - \rho_m(\tilde{y})\right) \rho(\tilde{y}) g(\tilde{y}) d\tilde{y}
$$

$$
r(y) = \rho_m(y) \cdot \chi(y)
$$

This decomposition allows us to understand what exactly are the externalities that career concerns make workers impose on each other. The key insight is that, from the point of view of individual workers, their individual actions do not affect the whole remuneration schedule of other workers, but they do not get paid their marginal products. When increasing their lifetime labor supply so that they get paid one extra dollar, they are contributing to the economy not one dollar but $\chi(y)$ dollars. In other words, it is as if they would produce one dollar but generate negative production externalities of the size of $1 - \chi(y)$ dollars.

It is interesting to compare how (a) the economy without any information asymmetries, (b) the economy with information asymmetries only between the government and the workers, as in the standard Mirrleesian model, and (c) the economy with information asymmetries between the firms, workers and government compare in terms of their utility possibilities frontier. In (b), when there are no information asymmetries in the labor markets, the planner could set taxes to zero, and then by the first welfare theorem, we know this is a first best allocation. Hence, there is a common point in the first and second best utility possibilities frontier, i.e., in the utility possibilities frontier of (a) and (b). Now, what is perhaps surpris-

---

22 As it appears for example in Scheuer and Werning (2017), and Saez (2001).
ing is that there is a common point in the utility possibilities frontier of (a) and (c), that is, between the third best utility possibilities frontier and the first best utility possibilities frontier, which is achieved by setting the Mirrleesian component of taxes to zero, so that \( r(y) = \chi(y) \), as shown in Proposition 13, in the Appendix Section A.8. This further justifies the decomposition above as one between a Mirrleesian and a Pigouvian component.

Employers when facing workers with the same resume, do not know who are the workers who will retire and will not extend their resumes even further. By the assumption that those who are more willing to extend their resumes are those who are more productive in the unobservable dimension, the workers who are retiring have the smallest productivity among those with the same resume. Therefore, we have that \( \chi(y) \leq 1 \), workers at retirement get paid more than their marginal products. From the tax formula above we can see that taking into account these career concerns unambiguously pushes towards higher marginal taxes at every income level, holding estimates of elasticities and densities of the income distribution constant.

These optimal taxes do not depend only on labor wedges but also on elasticities and the shape of the income distribution. Appendix Section A.7 shows that from the point of view of optimal taxation formulas, similarly to results in Scheuer and Werning (2017) common elasticity estimates are biased downwards. The reason for it is that increases in marginal retention induce the marginal types – who are the less productive types – to work more, and therefore reduce pretax salaries, making the effective change in post-tax wages smaller. Thus, the estimated elasticities of taxable income are lower in magnitude than the elasticities that enter optimal taxation formulas, which keep pretax salaries fixed. This is a countervailing force that makes accounting for career concerns push less towards higher marginal taxes.

These formulas can also be read as Pareto efficiency tests, as in Werning (2007b). If a tax schedule is Pareto efficient, then there are weights \( \lambda(y) \geq 0 \) such that, given the current tax rates, the income distribution, and the estimated elasticities and labor wedges \( \chi(y) \), the formula above holds. Relative to the standard Pareto efficiency test, the inclusion of \( \chi(y) \), holding the other estimated statistics fixed, makes the formula easier to be satisfied, that is, higher marginal tax rates can be rationalized.

### 4.2 Taxes Under Richer Signal Structures

So far in this Section, we have assumed that firms summarize a resume by looking at its length, which we have defined as the discounted cumulative sum of deliverables the worker has completed. This section relaxes this assumption and considers a richer set of signal structures under the same framework.
4.2.1 Conditioning on the “Strength of the Resume”

In this subsection, we will allow for a broad class of signals, that take the form of
\[ h_\phi(\tilde{h}(\tilde{a}), a) = \int_0^a \phi(\tilde{a}, a) \tilde{h}(\tilde{a}) d\tilde{a}, \]
with \( \phi(\tilde{a}, a) > 0 \), continuous in \( a \) and \( \tilde{a} \).\(^{23}\) The general idea behind this formulation is that the signal the firm sees can be thought of as the “strength of the resume”. This measure is an imperfect signal of the past history of deliverables of the worker, but completing more deliverables always makes the resume stronger.

From this perspective, the overlapping generation structure of the model can be thought of as a parsimonious way of introducing noise in how employers assess what has been done and when. While doing that, it implies that employers may pool together workers of different generations, whenever their past experiences are assessed as equivalent. It allows firms to see and put different weights on past experiences, and those weights may depend on the current age of the employee.

This formulation can capture in a reduced form different possibilities: firms may be interested in the pace at which the worker has produced deliverables, and thus may use \( \phi(\tilde{a}, a) = 1/a \). The firms may want to look at the total experience, as a proxy of human capital, and use \( \phi(\tilde{a}, a) = q(\tilde{a}) \). It may become hard to verify experiences in the distant past, so that \( \phi(\tilde{a}, a) < q(\tilde{a}) \). The firms may have all those concerns at the same time, as long as, when put together, they can be summarized by the idea that firms would evaluate the experience through the lens of an index of the form
\[ h_\phi(\tilde{h}(\tilde{a}), a) = \int_0^a \phi(\tilde{a}, a) \tilde{h}(\tilde{a}) d\tilde{a}. \]
This formulation bypasses the need to introduce explicitly all those elements, and to postulate explicit stochastic processes for shocks for each of those elements and considerations. Not introducing those shocks directly makes the analysis of optimal tax systems tractable, and avoids technical issues arriving from multidimensional screening problems, as well as from failures of the homogeneity assumption on preferences over the timing of consumption and labor supply flows.\(^{24}\)

Despite this apparent complexity, optimal taxation formulas will take a simple structure. We will follow the same logic from Proposition 2, and assume preferences are homogeneous over the timing of labor supply and consumption as in Proposition 1, and that more productive types are more willing to provide the deliverables.\(^{25}\) Under these assumptions, it

\(^{23}\)This assumption guarantees that expectations of productivities are well defined, and further it will be shown that those expectations are increasing in the completion of tasks \( \tilde{h} \). But, more generally, we could consider any informational structure with these properties. Later in this section, the full history of labor supply decisions will be assumed to be observed, and there will be richer heterogeneity in preferences to guarantee that those expectations are well-defined.

\(^{24}\)One may worry that those weights should be endogenous to the tax system. However, as it will be shown, the same optimal taxation formulas will hold even if those weights were to depend on the tax system.

\(^{25}\)Which in this case follows from additionally assuming that \( \frac{d^2 H(h(a))}{da(a')da(a)} < 0 \).
will be shown that, as in the previous section, taxes can be described as the composition of i) corrective taxes that guarantee that the workers benefits from each marginal increase in labor supply is equal to the their contribution to the output of the firm, and ii) redistributive taxes that are described by the same Mirrleesian optimal taxation formulas.

To show that, we first set aside the issue of implementation, and characterize the set of optimal incentive compatible allocations. Those allocations need to satisfy two properties, more formally stated in Lemma 3, in Appendix Section A.11. These properties state that, ii) intertemporally, labor supply decisions will not be distorted, and ii) an analogous equation to the standard Mirrleesian formula should hold in terms of the lifetime labor disutility index $H$.

Then, towards the goal of implementing these optimal allocations with a system of taxes and transfers, we establish that, quite generally, there is a positive return to experience – formally stated in Lemma 4 (in Appendix Section A.11). In other words, exerting more effort is a way to signal to employers that you are a more productive worker, and thus, will impact future salaries positively. Intuitively, the signals considered in this section are increasing in the effort decisions, and under the assumption that those who are willing to exert more effort are those who are more productive, employers can infer that higher signals translate into higher expected productivities. Besides being of interest in itself, this result is used to show that the planner can infer labor supply decisions from earnings histories, as formally stated in Lemma 5 (in Appendix Section A.11).

This result allows us to decentralize the incentive compatible and efficient allocations, analogous to the way that Lemma 2 was used to derive the optimal lifetime income taxation formulas from Proposition 2. That is, we can think of the planner as solving for the optimal allocation as in Lemma 3, or equivalently, solving for post-tax wages $R(\tilde{h}(\cdot)^1_0)$, and then implementing this post-tax retention function with history-dependent earnings taxes.

Finally, we summarize these taxation results in Proposition 3, below, which states that the tax system should be such that i) history-dependent taxes ($R_p$) should be used to correct for labor wedges, and ii) after correcting for these distortions, lifetime income redistributive taxes should be imposed on top of these taxes, according to standard redistributive formulas.

**Proposition 3.** If $R(\tilde{y}(\cdot)^1_0)$ is optimal, then, there exists $R_m, R_p$ with $R(\tilde{y}(\cdot)^1_0) = R_m(R_p(\tilde{y}(\cdot)^1_0))$, such that $R_m$ and $R_p$ satisfy the following conditions:

1. Intertemporal, Pigouvian: for any $\tilde{a}, \tilde{g}$, and $H(\tilde{h}(\tilde{a})^1_0) = H$, switching the timing of labor supply decisions and holding fixed lifetime labor supply and $H$ should leave lifetime
earnings unaffected:

\[
\int_a^1 \frac{dR_p}{d\tilde{y}(\tilde{a})} \frac{d\tilde{y}^{(\tilde{a})}_0}{dH} d\tilde{a}q(a) da = \int_a^1 \frac{dR_p}{d\tilde{y}(\tilde{a})} \frac{d\tilde{y}^{(\tilde{a})}_0}{dH} d\tilde{a}q(a) da
\]

2. Lifetime, Pigouvian: increasing \( H^* \) should increase lifetime earnings proportionally to the increase in output:

\[
\tilde{R}_p'(H) = \int_0^1 \frac{dR_p}{d\tilde{y}(\tilde{a})} \frac{d\tilde{y}^{(\tilde{a})}_0}{H^*} d\tilde{a}q(a) = v(H) \int_0^1 q(a) \frac{d\tilde{h}(a)}{dH} da
\]

3. Lifetime, redistributive: Define the retention that workers face as \( R_m(R_p(\tilde{y}(\cdot))) \), and \( r_m = R'_m(R_p) \). After correcting for distortions, then \( R_m \) should satisfy standard Mirrleesian formulas:

\[
\left( \frac{1 - r_m(R_p)}{r_m(R_p)} \right) g(R_p) R_p \epsilon^*_p(R_p) = \int_{R_p}^{\infty} g(\tilde{R}_p) \left( 1 - \lambda(\tilde{R}_p) \right) d\tilde{R}_p + \int_{R_p}^{\infty} \left( \frac{1 - r_m(\tilde{R}_p)}{r_m(\tilde{R}_p)} \right) g(\tilde{R}_p) \eta(\tilde{R}_p) d\tilde{R}_p,
\]

Notice that in this Proposition we describe taxes in “layers”, and there is a choice in describing taxes as one layer of Pigouvian taxes and another layer of Mirrleesian taxes. An alternative formulation could have one layer of intertemporal taxes, and another double layer of Pigouvian and Mirrleesian taxes. The first would keep lifetime income unchanged as a function of \( H \), and the second would feature an analogous lifetime Pigouvian component and a redistributive component as in Proposition 2. The following remark presents an alternative simple way of describing the Pigouvian component of taxes.

Remark 1. We can define \( R_p \) to be such that:

\[
\frac{dR_p(\tilde{y}(\cdot)_0)}{d\tilde{y}(\cdot)} = \frac{v(H)q(a)}{\int_a^1 \frac{d\tilde{y}(\tilde{a})q(\tilde{a})}{dH} da} = \frac{v(H)q(a)}{q(a)w(\tilde{h}(\cdot)_0)} + \int_a^1 \frac{h(\tilde{a})dw(\tilde{h}(\tilde{a})_0)q(\tilde{a})}{dH} da
\]

where \( v(H) \) is the marginal productivity of the type that supplies the level \( H \) of labor, and where for ease of notation the dependence on \( \tilde{y}(\cdot)_0 \) is omitted. That is, the formula should be read as a function of earnings flows \( \tilde{y}(\cdot) \), through the inverse operator \( \tilde{h}(\tilde{y}(\cdot)_0) \).

Notice that correcting for intertemporal distortions is a significantly more complicated endeavor: taxes should be history-dependent, and depend on how much a change in labor supply today translates into higher lifetime earnings, not only through its impact on current earnings, but additionally through its indirect impact on future salaries. For workers of different earnings histories, the tax rate they would face on the next dollar at a given period would depend on their future earnings and past earnings (which can be mapped to their labor supply choices), and how an increase of one unit of their labor supply today would
impact current and future earnings, with the latter capturing the private benefits from the signaling effects from that increase in labor supply.

Notice that if there is a tax system in place that already corrects for intertemporal distortions $R_p$, then Proposition 3 tells us that we are back to the simpler case covered in Proposition 1. Moreover, as a corollary of Proposition 3, we know that if the information structure does not introduce intertemporal distortions, then the simpler optimal taxation formulas 4 hold.

**Corollary 1.** *If the information structure is such that no intertemporal distortions are generated, then if taxes are optimal they satisfy condition 4.*

### 4.2.2 Conditioning on the Full History of Deliverables

The second case of more general endogenous information structures we consider is the case in which firms observe the full history of deliverables, and the detailed timing of the execution of those. That is, for a worker of age $a$, the firm will see the history of labor supply decisions from the time the worker was born up to their current age $a$, which is denoted $\tilde{h}(\cdot)^a_{0}$.\(^{26}\)

We assume in this section that preferences now take the form $U(C, \tilde{h}(\cdot), \theta)$, and that $\theta$ is high-dimensional. To talk meaningfully about expectations of productivities conditional on some history $\tilde{h}(\cdot)^a_{0}$, we need some workers to be willing to provide this rich set of histories, which would not be possible for preferences of the form $U(C, H(\tilde{h}(\cdot)), \theta)$, or with a low dimensional type space. To keep this section simple, we additionally assume there is a single consumption good $C$.\(^{27}\) We retain the assumption that more productive types are more willing to provide the deliverables.\(^{28}\)

Under those assumptions, we will show that optimal taxes can also be thought of as the composition of i) corrective taxes that guarantee that, for each additional unit of labor, workers receive lifetime benefits equivalent to their contribution to output, and ii) redistributive taxes, which, in this case, do not necessarily take the simple Mirrleesian form of the previous section.

Towards showing that, analogously to the previous section, first we establish that if the planner could choose the allocation, while being restricted to the set of incentive compatible allocations, any optimal allocation would lie at the frontier of production possibilities set

---

\(^{26}\)In this case, workers of different ages are never pooled together, as any history that ends up at age $a$ is different than a history that ends at age $a'$, when $a \neq a'$.

\(^{27}\)This is without loss, since discount rates $q(\cdot)$ are assumed to be exogenous.

\(^{28}\)Which in this case can be stated as: if $v(\theta) > v(\theta')$ then for any $\tilde{h}(\cdot)^a_0$, $C$, $MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot), \theta) < MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot), \theta')$, where $MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot), \theta) = -\frac{U_{\tilde{h}(\cdot),\theta}^{\tilde{h}(\cdot),\theta}}{U^{\tilde{h}(\cdot),\theta}}$.  

---

27
(Lemma 6 in the Appendix Section 4.2.2). Then, we show that the planner can use Pigouvian taxes to achieve the frontier of the production efficient set of allocations. This will be the case because the planner will be able to infer labor supply choices from the history of earnings, that is, sequences of \( \tilde{h}(\cdot)_0 \) will map to sequences of \( \tilde{g}(\cdot)_0 \) one-to-one, as in Lemma 2 and Lemma 5, from the previous sections.

To show that the planner can infer labor supply choices from the history of earnings, an important intermediate result is that there is a positive signaling return to experience. This result is interesting on its own, and it shows that the career concerns logic carries through this more general environment (see Lemma 7 in the Appendix Section 4.2.2, which is the analogous counterpart to Lemma 4). This signaling return to experience is driven by the fact that higher productivity types are those who are willing to provide labor supply paths with a larger number of deliverables. That property is used to show that the planner can infer labor supply decisions from earnings histories (see Lemma 8 in the Appendix Section 4.2.2), which in turn is crucial for the ability of the planner to implement allocations at the frontier of production possibilities set. That is, analogously to Lemma 7 and Lemma 4, from the previous Sections, Lemma 8 in this Section allows the planner to decentralize feasible, incentive compatible allocations that lie at the efficient possibilities frontier of the economy with the use of taxes on earnings. In doing that, Pigouvian taxes play an important role, as stated in the following Proposition.

**Proposition 4.** The planner can guarantee that the allocation would lie at the frontier of the production possibilities set by using Pigouvian taxes.

These Pigouvian taxes take the same general form as in the previous section’s Remark 1. Thus, although this economy may look quite complicated, the same principles of tax design can be applied. There is a caveat though. Because we have unrestricted preferences, and multiple goods, now the design of optimal redistributive taxes, after correcting for the Pigouvian distortions, is more complicated, and without further normative assumptions, we cannot point to lifetime income taxation as the preferred form of redistribution. From an implementation perspective, analogously to the case analyzed in the previous Section, the structure of taxes can be thought of as a composition of two layers of taxes. First, Pigouvian taxes correct for the distortions. Second, on top of these taxes, another layer of taxes is imposed to take care of redistribution.\(^{29}\)

\(^{29}\)Multidimensional heterogeneity can also make the second layer interact with the first layer, while retaining the general structure of a Mirrleesian formulas. For an example, see Section A.7.
4.3 Pigouvian Taxes: Two Simple Examples

In this section, two simple polar examples are presented. In the first case, all the heterogeneity in preferences come from how long each worker is willing to stay in the market. The signal the firms look at is the length of the resume, which in this case can simply be thought of as how many years of experience each worker has. In the second case, all the heterogeneity in preferences comes from how fast each worker is willing to supply the deliverables. The signal the firms look at is the pace, or how many deliverables per year the worker has provided. In both cases, we focus on a particular allocation, where Pigouvian taxes guarantee that the production is efficient, and there is no further redistributive taxes.

4.3.1 Length of Resumes and Heterogeneity on the Extensive Margin

In this example, all the heterogeneity in preferences is at the extensive margin, and the signals are defined as before, as the length of resumes. In this example payments per unit of the deliverable will increase with the experience of the worker, but the corrective component of taxes will be constant. There are no intertemporal distortions to be corrected, as whenever workers are on the job market they exert a constant amount of effort. There is also no pooling across generations.

Workers make a decision about when to retire, so they face a stopping problem. Given exogenous interest rates, and given choosing when to retire is the only decision they make, we can think of their preferences as $U(c, h, \theta)$, where $c$ and $h$ are lifetime measures of labor supply and consumption. We further assume that preferences are linear in consumption and with constant elasticities of substitution, that is:

$$U(c, h, \theta) = c - \frac{h^{1 + \frac{1}{\epsilon}}}{1 + \frac{1}{\epsilon}} b(\theta)^{-(1 + \frac{1}{\epsilon})},$$

where $b(\theta)$ is the cost of a worker of type $\theta$ to work longer and provide more deliverables. We denote productivities by $v(\theta)$, and assume production is linear in lifetime labor supply. Productivities and preferences are parameterized so that $v(\theta) = \theta^\delta$, $b(\theta) = \theta^{1-\delta}$, with $0 < \delta < 1$, and $\theta$ is Pareto distributed with a lower bound $\bar{\theta}$ and shape $\alpha > 1$.

The optimal allocation with Pigouvian taxes and no redistributive guarantees that the labor supply of each worker is such that marginal productivities are equal to the marginal cost of supplying more labor, that is it should satisfy the condition:

$$v(\theta) = h^{1/\epsilon} b(\theta)^{-(1 + \frac{1}{\epsilon})}.$$

We also know that salaries, with a signal structure of that form, are the average produc-
tivity of workers that remain on the market: \( w(h) = \mathbb{E}[v(\theta)|h(\theta) \geq h] \).

Marginal Pigouvian taxes should guarantee that workers at the margin are paid for their marginal product, i.e. post tax payments for a marginal unit of labor \((r(h))\) should be such that \( r(h) = v(\theta) \). We can translate this condition as a restriction on marginal retention on earnings \((r(y))\), which state that \( r(y) = \frac{v(\theta(y))}{\mathbb{E}[v(\theta)|y(\theta) \geq y]} \). In this case, under the parametric assumptions above, \( r(y) = \frac{\alpha - \delta}{\alpha} \). Because higher types are more productive and are those are who are willing to work longer, wages increase over time, and at the time retirement each worker would be facing a higher salary than their productivity. Pigouvian taxes are positive, and correct for that distortion. Notice that this example features exactly the same equations from the example in Section 3.1, with the subtle difference that here there is no pooling across generations, and because workers are heterogeneous only at the extensive margin, all workers that are pooled together are from the same generation.

Marginal taxes are increasing in the amount of information asymmetry \( \delta \), but are constant as a function of lifetime income. In this case, the Pigouvian corrections can be implemented with linear taxes. Furthermore, because those taxes are linear, they do not need to be a function of lifetime income per se and can be implemented with history-independent annual income taxes.\(^{30}\)

### 4.3.2 Pace of Resumes and Heterogeneity on the Intensive Margin

In this example all the heterogeneity in preferences are at the intensive margin, and the signals are defined as “pace of the resume”, that is, the ratio of the total amount of deliverables a worker has produced and the age of the worker. In this example payments per unit of the deliverable will be constant thoughout the career of the worker, and equal to their marginal product of labor. However, the corrective component of taxes will be positive and will correct for intertemporal distortions and the incentives to work too much when young.

Workers make decisions of how much to work at each period. They have additively separable preferences over time, and those are heterogeneous. In particular, we assume that preferences take the form:

\[
U = \int_0^1 e^{-p a} \left( \tilde{c}(a) - \frac{\tilde{h}(a)^{1+\frac{1}{\epsilon}}}{1 + \frac{1}{\epsilon}} b(\theta)^{-\left(1+\frac{1}{\epsilon}\right)} \right) \, da,
\]

where \( b(\theta) \) is the cost of a worker of type \( \theta \) to work longer and provide more deliverables. We denote productivities by \( v(\theta) \), and assume production is linear in the flow of labor.

\(^{30}\)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.
supply. Productivities and preferences are parameterized so that \( v(\theta) = \theta^\delta, \ b(\theta) = \theta^{1-\delta}, \) with \( 0 < \delta < 1. \) Under these assumptions, we can think of all the heterogeneity as coming from how each worker evaluates different paces of work, and not from how long they would like to stay in the labor market.

The optimal allocation has a constant labor supply over time and, for each worker, the flow needs to satisfy the following first order condition:

\[
\tilde{h}(\theta)^\frac{1}{\delta} b(\theta)^{(1+\frac{1}{\delta})} = v(\theta)
\]

Denoting a particular level of a resume pace as \( h_\alpha, \) using the condition above we can conclude that pre-tax salaries as a function of the pace would satisfy \( w(h_\alpha) = \mathbb{E}[v(\theta)]\int_a^\infty \tilde{h}(\tilde{a}) \tilde{a} \ d\tilde{a} = h_\alpha^{\frac{\delta}{1+\epsilon-\delta}}. \) Notice these are exactly the productivity of the workers for whom it is optimal to supply a certain pace \( h_\alpha. \) That is, salaries at each period are equal to the marginal productivity of the workers. This does not imply that there should be no corrective taxes, in fact the relevant labor wedge is not the static ratio between productivities and current salaries, but the ratio of productivities to the sum of current salaries and the increases in future lifetime earnings from reputation building effects.

In fact, for a worker who has supplied labor at the pace \( h_\alpha \) up until the age \( a, \) increasing their labor supply today and then reverting back to their constant flow increases their lifetime earnings by:

\[
\frac{dy}{dh(a)}(h_\alpha, a) = e^{-\rho a} h_\alpha^{\frac{\delta}{1+\epsilon-\delta}} + \left( \frac{\delta}{1+\epsilon-\delta} \right) \int_a^1 e^{-\rho a} \frac{1}{a} da \cdot \left( h_\alpha \right)^{\frac{\delta}{1+\epsilon-\delta}}^{-1}
\]

Where the first term is the current salary, and the second term are the reputation effects, that is the increase of future payments for the completion of future deliverables (holding the future flow at the constant pace \( h_\alpha. \)) Because the first term is exactly the productivity of the workers, corrective taxes as a function of labor supply \( (\tau) \) should be exactly equal to the second term, that is:

\[
\tau(h_\alpha, a) = \left( \frac{\delta}{1+\epsilon-\delta} \right) \int_a^1 e^{-\rho a} \frac{1}{a} da \cdot \left( h_\alpha \right)^{\frac{\delta}{1+\epsilon-\delta}}^{-1}
\] (5)

where \( h_\alpha \) can be inferred from the history of earnings.\(^{31}\) In this simple case, those Pigouvian

\(^{31}\)In terms of marginal retention functions \( R_p(\tilde{y}), \) we can write it as:

\[
\frac{dR_p}{d\tilde{y}} = \frac{e^{-\rho \tilde{a}} h_\alpha^{\frac{\delta}{1+\epsilon-\delta}}}{e^{-\rho \tilde{a}} h_\alpha^{\frac{\delta}{1+\epsilon-\delta}} + \left( \frac{\delta}{1+\epsilon-\delta} \right) \int_a^1 e^{-\rho a} \frac{1}{a} da \cdot \left( h_\alpha \right)^{\frac{\delta}{1+\epsilon-\delta}}^{-1}}
\]
taxes can be implemented as relatively simple age and history-dependent taxes. For workers of age $a$, if their lifetime earnings has been increasing at a ratio that is consistent with a constant labor supply flow equal to $h_a$, then the marginal taxes on labor they should face should be equal to those in equation 5. Those Pigouvian taxes correct not for the difference between current payment and productivities (which in this example is zero!), but for the difference between the lifetime gains from a marginal increase in effort today and the productivities. Those lifetime gains come from the positive reputation effects of exerting higher effort, which result in higher salaries in the future.

These large intertemporal corrective taxes are suggestive that, without them, younger workers would exert a lot more effort and work longer hours. This prediction is in line with the evidence from Landers et al. (1996) who show that, in law firms, hours decrease with tenure and associates work too many hours relative to their desires and relative to partners. This prediction however is 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 to 55, and they quickly fall for older workers (Card, 1991; Kaplan, 2012).

4.4 Changes in Information Processing Technologies

As new technologies and richer datasets begin to affect the workplace (Chalfin et al., 2016; Autor, 2019; Acemoglu et al., 2020; Bales and Stone, 2020), an important question to consider is the impact of technological changes in monitoring and screening workers on incentives, the distribution of income and the tradeoffs the government faces when setting taxes. Those changes should affect how workers get paid and how resumes are read, arguably making information imperfections less pronounced. On the other hand, the rising automation of routine tasks ( Autor et al., 2003), and the advent of “new work” ( Autor, 2019), that is, novel jobs reflecting changes in technology and tastes, have contributed to changes in the task composition of jobs. As non-routine cognitive tasks become more prevalent, it may become harder to monitor and assess the productivity of workers. As a result, these imperfections in the information transmission process may become more salient. In this section, we look at what are the consequences of changes in fundamentals that make information asymmetry problems worse or better.

We define information asymmetry problems in labor markets to decrease if deliverables become a better measure of product, affecting how firms pay workers and how firms read resumes. More precisely, we adopt the following definition.

---

32 Notice that Lemma 5 guarantees that quite generally in this setup the government can infer the labor supply flows from the earning flows.
Definition. Let preferences be denoted $U(c, h, \theta) = \tilde{U}(c, h/b(\theta), \theta)$. If new productivities and tastes are such that $v^\theta(\theta) = v(\theta)/\Delta(\theta)$ and $b^\theta(\theta) = b(\theta) \cdot \Delta(\theta)$, $\Delta > 0$ and increasing in $\theta$, order preserving (meaning that if $\theta > \theta'$, then $v(\theta) > v(\theta')$, $v^\theta(\theta) > v^\theta(\theta')$, $MRS_{c,h}^\theta > MRS_{c,h}^{\theta'}$, and $MRS_{c,h}^M > MRS_{c,h}^{M'}$), then information asymmetries in labor markets decrease.

Under this definition, decreasing information asymmetries keeps the first best utilities possibilities frontier unchanged. It also increases the ratio $v(\theta)/\mathbb{E}[v(\tilde{\theta})|\tilde{\theta} > \theta]$ for every type of worker, and preserves the relative ranking of types in terms of productivities and marginal rates of substitution.

A perhaps surprising result is that information asymmetries in labor markets, very generally, increase welfare in this setup, as stated in Proposition 5. This result implies that those frictions are, in a particular way, good for redistribution, and attenuate inequality. That is, if how the government evaluates inequality is expressed in how it sets marginal tax rates, then an increase in information asymmetries in labor markets increases welfare, or, in other words, improves the distribution of outcomes from the point-of-view of a redistributive planner. Information asymmetries help with redistribution, because they make harder for high productivity workers to separate themselves from low productivity workers. Moreover, because under the assumptions of Lemma 2, (or the more general Lemma 2) the optimal post-tax wages do not depend on the specifics of the model of asymmetric information, the impact on welfare is the same across different models, provided that taxes are set optimally and following the same preferences for redistribution.

In other words, this result does not depend on the particular assumption we imposed on the information structure that determines how salaries are set by firms. It relies on the planner being able to solve for the allocation directly, which is possible whenever there is a one-to-one mapping between earnings and labor supply decisions. This property is satisfied by different information structures as shown in Section 4.2. Further, the result also holds across different models of career concerns, provided they satisfy the invertibility conditions behind Lemma 2 or the more general Lemma 5, that allow for labor supply decisions of workers to be perfectly inferred from their history of earnings. In fact, this result is a generalization of Stantcheva (2014) Proposition 13, which compares welfare in an economy under a Miyazaki-Wilson-Spence (MWS) model of the labor market relative to a Mirrleesian economy.\(^{33}\)

Proposition 5. If the original tax schedule is optimal and the planner has strongly enough redistributive preferences, and labor is a normal good, then decreasing information asymme-

\(^{33}\)Another similar setup where these conditions are satisfied is the Azevedo and Gottlieb (2017) model of competitive screening.
tries in labor markets decreases welfare.

Proof. See Appendix Section A.9.

The idea behind this result is that if the planner has redistributive tastes, then it sets positive marginal redistributive taxes to transfer resources from higher types from lower types, an the incentive compatibility constraints bind downwards (Seade, 1982; Werning, 2000). Whenever the degree of informational asymmetry in labor markets decreases, these downward incentive compatibility constraints become tighter, as it becomes less costly for the high types to imitate the low types. With less information asymmetries in labor markets, they can use more of their previously unobserved productivities to imitate the deliverable production of the lower productivity workers.

The impact on taxes is more subtle. When resumes are defined as the cumulative discounted sum of deliverables, the corrective component unambiguously falls for each type, because $\chi(\theta) = \frac{v(\theta)}{E[v(\tilde{\theta}) | \tilde{\theta} \geq \theta]}$ increases proportionally more in the numerator relative to the denominator. In that vein, we can think of an decrease in information asymmetries in labor markets as a force towards lower taxes. To understand how the Mirrleesian component and total marginal taxes would be affected is useful to write optimal taxes in terms of types $\theta$, as in the following proposition.

**Proposition 6.** Optimal taxes as a function of types $\theta$ must satisfy the following equations:

$$r_y(\theta) = r_m(\theta) \cdot \chi(\theta)$$

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

Proof. See Appendix Section A.10.

Changes in the degree of informational asymmetry affect two key ingredients in the formula above $\chi$ and $\frac{\partial \log MRS}{\partial \theta}$, and each of them, respectively, affects directly the Pigouvian and the Mirrleesian component of taxes. They also operate very differently. The Pigouvian component of taxes for a type $\theta$ is affected by the changes in the unobservable component of the productivity of everyone that ends up with a lifetime income higher than $y(\theta)$. As the type of labor those people supply become more easily measurable, type $\theta$ ends up getting a smaller subsidy from them, and $\chi(\theta)$ increases, approaching one. The Mirrleesian component on the other hand depends on how type $\theta$ is more or less willing to provide more deliverables relative to their local neighboring types. Thus, for any given $\theta$, one can imagine a decrease in
informational asymmetries that can have impacts on the Pigouvian or Mirrleesian component of taxes of arbitrarily different magnitudes. That is, the net effect on total marginal taxes, taking into account both the Pigouvian and Mirrleesian components, is in general ambiguous.

A simple example shows that we may expect the Pigouvian component to dominate over the Mirrleesian component in certain circumstances. Assuming, there are no income effects, and that preferences take the simple form as in the example from section 3.1, and types are Pareto distributed, we can solve for optimal marginal retention as a function of types as:

\[ \chi = \frac{\alpha - \delta}{\alpha}, \quad \frac{1 - r_m}{r_m} = \frac{1 - \lambda}{1 + \epsilon} \cdot \frac{\alpha}{1 - \delta} \quad \Rightarrow \quad r = r_m \cdot \chi = \frac{\epsilon}{1 + \epsilon} \cdot \frac{\alpha - \delta}{1 - \delta} \cdot \frac{1 - \lambda}{1 + \epsilon} + \frac{\epsilon}{1 + \epsilon} \cdot \frac{\alpha}{1 - \delta}\]

where \( \bar{\lambda}(\theta) = \mathbb{E}[\hat{\lambda}(\tilde{\theta})|\tilde{\theta} \geq \theta] \), the marginal of a one dollar transfer to the types above \( \theta \), which is assumed to be constant in \( \theta \) starting from some level \( \hat{\theta} \). \( \chi \) is decreasing in the degree of informational asymmetry, corrective taxes increase as informational asymmetries increase. \( r_m \) is increasing in the degree of informational asymmetry, redistributive component of taxes decrease. Provided that \( \frac{\alpha}{1 + \epsilon} \geq 1 \) (which incidentally guarantees that output is finite), holding fixed this marginal value of transfers \( \lambda \), marginal retention \( r \) decreases with the degree of informational asymmetry \( \delta \), that is, marginal taxes increase as the degree of informational asymmetry increases. The Pigouvian component dominates over the Mirrleesian component, calling for higher marginal taxes as the degree of informational asymmetry increases.

5 Empirical Evidence

The previous section has offered an understanding of the tradeoffs a government faces between incentives and redistribution through the lens of simple sufficient statistics. In this sense, a key new element that needs to be estimated is the ratio of marginal productivities over salaries for the last unit of labor that workers supply.\(^{34}\)

5.1 Data

To evaluate the magnitude of the signaling return to experience and the Pigouvian component of taxes we use Health and Retirement Study data. The survey is a representative sample of the US population older than 50 and a biannual panel covering the period from

---

\(^{34}\)More generally, we would like to estimate the ratio of marginal productivities over the sum of current salaries and all future earnings increases that result from supplying one unit of labor, at each point in the career of the workers, as discussed in Section 6. While estimating marginal productivities over salaries is a considerably hard challenge, estimating the latter is an even harder challenge. Direct attempts at estimating the latter are left for future work.
1992-2018. It follows around 20,000 workers and it is rich on covariates, including job histories, hours worked, education, cognition, measures of lifetime income, and geographic, industry, and occupation variables.

There were several state and federal tax reforms over the period, which we are going to explore as a source of exogenous variation in wages. The federal tax reforms include the Omnibus Budget Reconciliation Act of 1993, which affected mostly top income earners; the Economic Growth and Tax Relief Reconciliation Act of 2001, which affected those at the bottom and at top of the income distribution; the Jobs and Growth Tax Relief Reconciliation Act of 2003, which affected tax rates for middle and top income earners; the American Recovery and Reinvestment Act of 2009, with tax changes across different parts of the income distribution; and the American Taxpayer Relief Act of 2012, which changed tax rates at the top of the income distribution.

There were also several state tax reforms. These were widely dispersed across the US, as shown in Figure A1, although they were more prevalent in some states such as California, Connecticut, Delaware and Idaho, and were almost completely absent in Alaska, Florida, Nebraska, North Dakota, South Dakota, Tennessee, Texas, Washington, and Wyoming. A more detailed description of state tax reforms is presented in the Appendix Sections B.2 and B.5.

To explore the variation induced by those changes in marginal rates, we construct simulated changes in marginal tax rates at initial incomes using the NBER tax simulator. We assign a wide range of income, consumption and demographic variables from HRS to the 32 inputs in the NBER tax simulator, adapting and extending to our period of analysis (1992-2018) the procedure developed by Pantoja et al. (2018).  

5.2 Empirical Strategy and Results

The key statistic we would like to estimate is the ratio of the marginal productivity of workers at retirement over their salaries, as a function of their lifetime income. Our approach relies on tax changes as a source of exogenous variation in wages, to quantify the degree of informational asymmetry. The key idea is that, looking at the labor markets around the time of retirement, the change in average productivities induced by the exogenous variation in wages is informative of the marginal productivity of those who are almost indifferent between working more or retiring.

This approach builds on the literature that has quantified the degree of adverse selection in markets described by simple Akerlof (1970) lemons conditions, such as Einav et al. (2010).

35A more detailed description of the mapping between the variables is presented in Appendix Section B.1.
In the context of health insurance markets they analyze, the key argument is that with a source of exogenous variation in prices, one can non-parametrically trace the shape of the cost curve for the insurance contract by looking at the average cost as a function of prices. The average cost, for a given price, can be inferred from data on insurance claims. Translated into our context, with a source of exogenous variation in wages (for a specific labor contract), one can non-parametrically trace the shape of the labor demand curve by looking at average productivities as a function of salaries.

However, inferring productivities in labor markets is a more difficult endeavor than inferring costs in insurance markets. In insurance markets, detailed data on insurance claims can be used to compute reasonably precise measures of expected costs for the insurance contracts. In labor markets, most often direct data on productivities is not available. To circumvent this challenge, two complementary approaches are adopted. The first assumes that labor markets are competitive and therefore leverages the observation that wages would be equal to the average productivity of workers with the same resume, as in the model presented in this paper. The second takes advantage of the rich set of covariates available from the Health and Retirement Study and looks at cognitive measures as proxies for productivities.

5.2.1 Salary Changes and Retirement Decisions

At the general level, to infer the degree of information asymmetry from salary changes, consider an increase in marginal taxes for those who are near retirement and how it affects a pool of workers with equivalent resumes and the same original salaries. This tax change may induce some of the individuals in this pool to actively retire. The pool of workers who remain in the workforce in the next period is different, because it no longer contains those who were almost indifferent between retiring or not. If those who were almost indifferent were also less productive in the unobservable dimension, as in the model, then wages should increase. Moreover, by observing wages before and after the change in marginal rates, we can find the average productivity of those who are close to being indifferent between retiring or not. That is, we can decompose wages before and after as a weighted sum of the productivities of those who are almost indifferent and those who are not, that is:

\[ E[v_{before}] = E[v_{after}](1 - s_{mg}) + v_{mg}s_{mg}, \]

where, \( E[v_{before}] \) and \( E[v_{after}] \) denotes the average productivity (or salaries) of the workers with a common resume, respectively before and after the tax increase; \( s_{mg} \) is the share of people who are induced to retire by the increase in marginal taxes. This simple relationship can be written in terms of labor wedges (\( \chi \)), elasticities of salaries (\( \epsilon^{w} \)) and semi-elasticities.
of labor market participation ($\eta_{pr}$). That is:

$$\chi = \frac{v_{mg}}{E[v_{after}]} = 1 - \frac{E[v_{after}] - E[v_{before}]}{E[v_{after}]}$$

$$\implies \chi = 1 + \frac{\epsilon_w}{\eta_{pr}}$$

Thus, from observing salaries before and after, and the share of people who retire as a result of a small tax increase, one can infer the productivity of the people who are marginally indifferent between retiring or not. Labor wedges $\chi$, thus, can be inferred from elasticities of salaries ($\epsilon_w$) and semi-elasticities of labor market participation ($\eta_{pr}$). Data from the Health and Retirement Study is particularly well suited for recovering these elasticities, as it includes carefully calculated measures of lifetime income, the timing of retirement decisions, salaries, hours of work, and detailed income measures allowing us to approximate the marginal tax rates faced by workers using the NBER taxsim model.

While this explanation has focused on salary levels, and a simple before and after comparison, in practice we will be looking at salary changes at the individual level, and we will be pooling together different tax reforms, across different state, time periods, and parts of the income distribution, while at the same time controlling for year-fixed effects, initial hourly wages, and marital status. The key identification assumption for the results that rely on changes in salaries is that future productivity changes, and elasticities of participation, are independent of each other and of tax changes conditional on the set of controls.\(^\text{36}\)

The Pigouvian externality ($\chi - 1$) can also be interpreted as the coefficient of a two sample instrumental variable regression of changes in log salaries on changes in participation, where changes in taxes are the instrument for changes in participation. Intuitively, the coefficient on that regression tells how salaries change when the marginal worker is forced to stay in the labor force, and thus is informative of their productivities.

To obtain estimates for the elasticities of wages, we regress changes in log hourly salaries on simulated changes in the log of marginal retention rates, including different sets of controls $X_{it}$, as in equation 6. These controls aim to capture i) the possibility that changes in the tax schedule and in hourly wages may both respond to business cycles fluctuations (thus the inclusion of year-fixed effects), ii) the possibility that tax changes may have targeted different income groups and wages may evolve differently for those different groups (thus the inclusion of log hourly wages, and other non-linear functions of hourly wages), and iii) similarly, tax changes may have targeted differentially people of different marital status, for whom wages may evolve differentially as well (thus the inclusion of marital status indicator variables).

\(^{36}\)This assumption becomes weaker when we estimate heterogeneous elasticities by different groups as in 5.2.3, as then only within groups the elasticities of participation need to be independent of future productivity changes.
\[ \Delta \log w_{it} = \epsilon \Delta \log r_{it} + \gamma' X_{it} + u_{it} \] (6)

The results of this set of regressions are presented in Figure 8 and Table 3 in the Appendix Sections B.4 and B.5. For the main specification, which includes year fixed effects, marital status dummies, and a 10-piece linear spline on hourly wages as controls, the estimated elasticity of wages is -0.16, with a standard deviation of 0.1, implying that a 1% increase in marginal retention between years 0 and 2 causes a 0.16% decrease in salaries between years 0 and 4. The effects are stronger at the 6-year horizon, with a point estimate of -0.27, and revert back at the 8-year horizon to -0.14, when estimates also get noisier. This is in line with the idea that marginal tax increases push the people who were almost indifferent between retiring or not into retirement, and those people are less productive than the average worker although they were receiving the same salaries. Salaries then would increase as employers realize that there was a change in the productivity composition of the pool of workers still on the labor force.

It is a common wisdom that workers experience most of their salary changes when they change jobs. To more precisely capture the effects of changes in marginal tax rates on the wages of workers, we also consider the effects of restricting the sample to only those who switch jobs over the relevant time period. These results are presented in Figure 9 and Table 4 in the Appendix Section B.4. These elasticities are higher in magnitude, in line with the idea that wages are more flexible when workers switch jobs. The results from the main specification (including the full set of controls and a 10-piece spline on hourly wages) imply a 1% increase in marginal retention causes wages to fall by 0.34% over the 4-year horizon, and by 0.43% over the 6-year and 8-year horizons.

To obtain estimates for the semi-elasticities of participation, analogously, we regress changes in participation\(^{37}\) on simulated changes in the log of marginal retention rates, including different sets of controls, as in equation 7. Again, these controls aim to capture the possibility that changes in the tax schedule and in labor market participation may both respond to business cycles fluctuations (thus the inclusion of year-fixed effects), the possibility that tax changes may have targeted different income groups and labor market participation may evolve differently for those different groups (thus the inclusion of log hourly wages, and other non-linear functions of hourly wages), and similarly, tax changes may have targeted differentially people of different marital status, for whom labor market participation may evolve differentially as well (thus the inclusion of marital status indicator variables).

\[ \Delta p_{it} = \eta p \Delta \log r_{it} + \gamma' X_{it} + u_{it} \] (7)

\(^{37}\)That is, changes in the indicator variable that is equal to one whenever the individual is working.
The estimated semi-elasticities of participation when including the full set of controls and the 10-piece spline on hourly wages are of the order of 0.10 at the 2-year horizon, 0.01 at the 4-year horizon and 0.03 at the 6-year horizon, implying that at the 4-year horizon, a one percent increase in marginal retention causes a 1 percentage point decrease in the probability of a worker getting out of the labor force. The relatively larger effects at the shorter horizons when compared to the elasticity of wages is consistent with the idea that, after a tax increase, first some of the workers drop out of the labor market, and then, as the employers learn that the pool of the remaining workers is more productive, wages gradually increase as time passes.

Our estimates are in line with a literature that has found that the implied tax rates on labor income from public pension rules have large disincentive effects on work (Gruber and Wise, 1998; Coile and Gruber, 2007). On the other hand, they are higher than the substitution elasticities inferred from variation in Social Security benefits from the 1977 Social Security Act that created the so called “Notch generation” Gelber et al. (2016). While our confidence intervals are relatively wide, the difference can also be explained by the fact that the substitution incentives of their reform, and of Social Security benefits more generally, are relatively more opaque (Blinder et al., 1980) than the substitution incentives from income tax reforms.

Taking the ratio of the estimated coefficients \( \frac{e}{\eta_p} \) (while multiplying \( \eta_p \) by one hundred so numerator and denominator are in the right units) we obtain estimates for the magnitude of the labor market informational externality \((1 - \chi)\). These results are presented in Tables 6 and 7. For the main specification, which includes the full set of controls and the 10-piece spline on hourly wages in both the participation and wages regressions, and looks at the effects over a 4-year horizon, the estimated negative informational externality is around 0.16. In other words, workers are paid around 16% more than their marginal productivity for the last unit of labor they supply.

5.2.2 Cognitive Measures: Inspecting the Mechanism

There could be other stories that explain why wages increase and participation falls when there are tax increases. Most notably, firms’ labor demand may be partially elastic. In order to provide further evidence that the pool of workers is playing a role in salary changes, we look at cognitive measures collected by the Health and Retirement Study. The average cognitive measure of the individuals working is lower after a tax increase, even when that measure is taken before the tax change.

Mental status scores in the Rand harmonized longitudinal files from the Health and Retirement Survey are computed as the sum of vocabulary, naming, and counting scores
from the HRS. Those scores are the sum of correct answers from questions ranging from 
“Who are the current president and vice-president of the United States?” to “How much is 
100 minus 7? How much is that minus 7? [...].” The detailed construction of this variable is 
presented in Appendix B.3. Those measures can be seen as a proxy for ability, similarly to 
how Armed Forces Qualification Test (AFQT) scores in the National Longitudinal Survey of 
Youth is often used as a measure of ability (Farber and Gibbons, 1996; Altonji and Pierret, 
2001; Lange, 2007; Craig, 2020). While it has the disadvantage of being less detailed, it has 
the advantage of being assessed repeatedly for each respondent, at every survey year, and for 
that reason may be a more accurate measure of ability if ability is not constant but evolves 
dynamically over time.

Looking at the HRS total mental status scores as a proxy of ability, we regress those 
scores as measured two years before the baseline year, on changes in marginal retention 
and a set of control variables ($X_{it}$, including year fixed effects, marital status and flexible 
functions of hourly wages), restricting the sample to those who are working in the baseline 
year and four years in the future. The coefficient $\eta_{ir}^m$ on regression 8 translates how different 
are the average mental status scores of those who are working after changes in marginal 
retention.

$$scores_{it} = \eta_{ir}^m \Delta \log r_{it} + \gamma'X_{it} + u_{it} \quad (8)$$

The results for different sets of controls are presented in Table 10 in the Appendix Section 
B.4. Fixing the set of controls, the results for different time horizons are presented in Figure 
11 in the Appendix Section B.5. For a horizon of four years after the baseline year, under 
the most stringent specification, a one percent increase in marginal retention between the 
baseline year a two years ahead a change in the composition of the pool of workers such that 
average mental status scores (as measured before the tax change) decrease by 1.3 points (out 
of 15), with a standard deviation of 0.5, conditional on hourly salaries. This effect is in line 
with the mechanism emphasized in this paper, where salaries change as a response to changes 
in the productivity composition of the workers who are are willing to supply the deliverables 
given the current incentives, and where salaries and taxes work as screening devices both for 
the government and for the firms.

### 5.2.3 Heterogeneity Across Income Levels

The argument made on Section 5.2 is built on observing salaries and cognitive measures 
for the workers who have the same equivalent resumes, and thus would face the same remu-
neration for their next unit of labor they would supply. However, the elasticities of wages 
and the semi-elasticities of participation may be heterogeneous across different resumes and
labor contracts.

To address that heterogeneity, we estimate equations 6 and 7 locally as a function of hourly salaries,\footnote{Even within remuneration levels, there could be heterogeneity in elasticities and labor market wedges, in which case the relevant optimal taxation formulas also call for the estimation of correlation among those, as in the tax formula 10. For evidence of further heterogeneity across education groups and occupations, see the Appendix Section B.4.} using local polynomial methods. That is, for different dependent variables \( dep_{it} \), regression equations as in 9, where \( y \) denotes an hourly wage level. In those regressions, observations are weighted by their distance from the hourly wages level \( y \) where the equation is being evaluated using the Epanechnikov kernel, and an optimal bandwidth selected using a leave-one-out cross validation procedure. An additional cross-term \( \Delta \log r_{it}(y_{it} - y) \) is included to improve on the bias-variance tradeoffs, as explained in more detail in Fan and Gijbels (1996). Optimal bandwidths are selected with the leave-one-out cross validation procedure proposed by Racine (1993). Bootstrap confidence intervals are generated using the basic bootstrap method described in Chapter 5 of Davison and Hinkley (1997).

\[
dep_{it} = \epsilon_r(y)\Delta \log r_{it} + \beta(y)\Delta \log r_{it}(y_{it} - y) + \gamma(y)'X_{it} + u_{it} \tag{9}
\]

The local results are presented in Figures 12 to 14 in the Appendix Section B.5. They show that elasticities of wages are higher in magnitude for high-earners, which also have lower participation semi-elasticities. Moreover, semi-elasticities of mental status scores are also higher in magnitude for high-earners. This is in line with the idea that informational imperfections are a larger issue for high-earning occupations and jobs. The point estimate for the elasticities of wages suggest that at the top of the distribution of hourly salaries the labor wedges could be very high, but also are imprecisely estimated. The lower bound on the confidence interval at the 90th percentile would rule out a value lower the 0.5, implying that those workers could be paid more than twice their marginal productivities. However, these values come from a combination of high elasticities of hourly wages and low participation elasticities, which approach zero, raising concerns about the validity of the bootstrap confidence intervals. The general message however, is that we should expect the labor wedge \( \chi \) to be decreasing in income, and it is reasonable to expect values ranging from 0.9 to less than 0.5, where 0.9 is the estimated \( \chi \) for the upper third of the income distribution.

5.3 Comparison to Existing Evidence

In this section, we compare our estimates to the available evidence on the time patterns of salaries and signaling on labor markets. We show that a back-of-the envelope calculation using the existing evidence from the literature would result in a similar magnitude for the
Pigouvian component of taxes we found in the previous section.

There is documented evidence that workers experience large growth rates of salaries as a function of experience. In fact, Guvenen et al. (2021) have documented that the top 1% earners have a very steep growth rate of salaries, of around 2700% over a 30-year period, or 11.3% per year (and approximately 3% per year on average across workers). While part of this pattern may be thought 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. In fact, Guvenen et al. (2021) argue that no empirically plausible model of stochastic productivities could explain the large growth rates of salaries observed at the top. Moreover, there is evidence that signaling and learning are important to explain the dynamics of salaries and tenure in some occupations. For example, Cella et al. (2017) have shown that the relationship between the volatility of stock returns and the tenure of CEOs of large US firms is consistent with the idea that the market gradually learns about the CEO ability throughout the years of the CEO tenure.

For the purpose of this paper, the key question is what share of the growth rate of salaries is due to signaling and how can it be translated into an estimate of the corrective component of taxes. However, 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. Out of the return to schooling, recent work has concluded that on average 30% can be attributed to signaling and 70% to human capital accumulation (Aryal et al., 2019). Combining the growth rate of salaries for top earners from Guvenen et al. (2021) with the signaling fraction of the return to schooling from Aryal et al. (2019), we can guess that the return to experience due to signaling, for top earners, may be of the order of 3.4% (30% of 11.3%) per year at 10 years of experience. To translate those numbers to a magnitude for the corrective component of taxes, using the free entry condition 3, we can show that 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 \(\gamma = \frac{\alpha_s (1-\chi)}{1 + \alpha_s (1-\chi)}\). This implies that, for the values above, the Pigouvian component of taxes at the top that could be as high as 25%, and on average around 6%.
Richer Type Space, Signal Structure and On-the-job Learning

To add realism to an otherwise stylized model, several extensions to the basic model are presented. These extensions include human capital accumulation, richer heterogeneity in elasticities, and richer signal structures. Key insights from the generalized optimal taxation formulas and their empirical implications will hold with some caveats in those extensions.

The first extension allows human capital accumulation in the form of learning-by-doing as in Arrow (1962), and relaxes the assumption that workers have constant productivities over their lifetimes, which is an evidently implausible assumption. It is reasonable to expect that at least some part of the return to experience observed in the data is due to increases in productivity due to on-the-job learning, or training efforts. However, the extended model shows that this assumption is to some extent innocuous. Although it complicates the relationship between the observed return to experience, the degree of information asymmetry in the market and the rate of human capital accumulation, the same optimal tax formula applies when on-the-job learning is costless.

The second extension allows for richer heterogeneity in elasticities. This extension is motivated by the empirical evidence that there is substantial heterogeneity in how people respond to taxes, even within tax brackets (Eissa and Liebman, 1996; Gruber and Saez, 2002; Blau and Kahn, 2007; Vere, 2011; Sturm and Sztutman, 2021). The key modification that multidimensional types introduce is that now, at a given income level, it matters not only how much a worker produces per unit of pretax income, but how these are correlated with the elasticities. Intuitively, an increase in post-tax salaries at a given income will affect people of different elasticities differently, and production will increase proportionally to the product of elasticities and productivities, and therefore will increase more if the elasticities and the unobserved productivities are positively correlated.

The third extension allows firms to see additional signals the government does not see. For example, from the point of view of the employers it may be clear that some workers are on different career tracks, and that those workers can hardly change that. But for the government, it may be hard to distinguish them, or it may be hard to codify those distinctions into the tax system in a way that cannot be manipulated. In this case, optimal taxes are described by a weighted version of the basic taxation formula, where the weights are given by the sensitivity of the different post-tax retention functions to changes in marginal taxes. This modification of the optimal tax formulas follows from the fact that there are multiple career tracks, there are multiple pretax salary functions, while there is a single nonlinear taxation instrument the government can use. Considering a variation in the income retention schedule
and tracking how this variation affects the post-tax salaries of different careers in response to it results in the weighted version of the basic optimal tax formula. The effects of changes in marginal taxes on post-tax salaries are attenuated by changes in pre-tax salaries whenever there are information asymmetries. For this reason taking into account these different career tracks may attenuate the magnitude of the corrective component of taxes.

The fourth extension, similarly considers the possibility that other functions of the detailed timing of the completion of tasks may be observed by firms, while the government could see the history of earnings. This possibility makes the signaling return to experience more involved, and introduces the possibility that without taxes there may be intertemporal distortions. Under common preferences over the timing of labor and consumption decisions, the optimal tax system can be written in a way where taxes that depend on the history of earnings would correct for this distortion, and on top of these taxes, optimal lifetime income taxes that are described by the same optimal lifetime income taxation formulas.

### 6.1 On-the-job Learning

The assumption that workers have a constant productivity over their lifetimes is, of course, extreme. It is reasonable to expect that at least some part of the return to experience observed in the data is due to increases in productivity due to on-the-job learning, or training efforts. A simple way to accommodate these concerns is to allow for productivities to depend not only on the types of workers but also on experience itself, that is \( v = v(\theta, h) \). Indeed, one of the key reasons why employers may focus on experience as a signal for the productivity of workers is exactly because accumulating experience may directly increase the productivity of workers.

Turning back to the example from Section 3.1, a simple way to enrich the setup is to assume that that productivities increase proportionally with experience, that is, productivities are \( v(\theta, h) = \hat{v}(\theta)h^\beta \). Then, because wages are the expectation of productivities conditional on \( h \), \( w(h) = \mathbb{E}[v(\hat{\theta}, h)|\hat{\theta} \geq \theta(h), h] = h^\beta \mathbb{E}[\hat{v}(\hat{\theta})|\hat{\theta} \geq \theta(h)] \), and the return to experience would have two components, a signaling and human capital accumulation component. Salaries still satisfy a log-linear relationship, with a coefficient that is the sum of the signaling \( \delta \) and the human capital component \( \beta \):

\[
\log(w) = \left( \frac{\delta}{1 - \delta + \epsilon} + \beta \right) \cdot \log(h) + \left( \frac{(1 - \delta)(1 + \epsilon)}{(1 - \delta)(1 + \epsilon) - \delta \epsilon} \right) \cdot \log\left( \frac{\alpha}{\alpha - \delta} \right)
\]

More generally, while human capital accumulation very much changes the meaning of the return to experience, the next proposition, focusing on the simpler case where the resume is defined as the cumulative discounted sum of deliverables, shows that the necessary condition
for the optimality of taxes from 4 is unchanged.

**Proposition 7.** Suppose productivity depends on experience and the unobserved types of workers. Then we can write the planner’s problem as:

$$\max_{\tilde{R}(h)} \mathbb{E}[W(V(\tilde{R}; \theta)) | h] \text{ s.t. } \mathbb{E}[\int_0^{h(\tilde{R}, \theta)} (v(\theta, \tilde{h}) - \tilde{r}_h(\tilde{h})) dh - I] \geq 0$$

And a necessary condition for a tax schedule to be optimal is given by the following formula (which is analogous to 4):

$$\left(\frac{\chi(y) - r(y)}{r(y)}\right) \epsilon_t(y) g(y) y = \int_y^\infty \left(1 - \lambda(\tilde{y})\right) g(\tilde{y}) d\tilde{y} + \int_y^\infty \left(\frac{\chi(\tilde{y}) - r(\tilde{y})}{r(\tilde{y})}\right) \eta_I(\tilde{y}) g(\tilde{y}) d\tilde{y},$$

where $\chi(y) = \frac{v(y)}{g(h(y))}$, and $v(y) = v(\theta(y), h(\theta(y)))$ is the productivity at retirement of the worker with lifetime income $y$. While the formula looks exactly the same, there is a subtle but important distinction. We cannot use estimates of productivity of workers far from retirement to infer their productivity at retirement, as these two quantities can be significantly different. Second, we cannot infer the degree of informational asymmetry by only looking at the return to experience. For these reasons, the empirical strategies aimed at estimating $\chi(y)$ applied in the previous section are designed to be robust to these concerns, and to aim precisely at disentangling the return to experience coming from human capital accumulation from the return to experience coming from employers learning about productivities through job histories.

### 6.2 Heterogeneity in Elasticities

The main version of the model presented features a single-dimension of heterogeneity. Types of different productivities and willingness to provide the deliverables sort themselves into different lifetime income levels, and within lifetime income levels there is no heterogeneity.\(^{39}\) However, it is reasonable to expect that productivities, and elasticities might be heterogenous within lifetime income levels. Although the standard Mirrleesian first order condition is basically unchanged when agents have heterogeneous elasticities (Scheuer and Werning, 2016; Jacquet and Lehmann, 2021; Bierbrauer et al., 2020; Sturm and Sztutman, 2021), when there is imperfect information this heterogeneity creates a subtle interaction between the Pigouvian and Mirrleesian component of taxes in a way that is reminiscent

\(^{39}\)Indeed, Sturm and Sztutman (2021), among others, have shown that there is substantial heterogeneity in elasticities within (annual) income levels.
of Diamond (1973). Intuitively, when taxes increase at a given bracket, different workers respond differently, while simultaneously facing different labor wedges. Thus, how the correlation of elasticities and labor wedges matter for the total effect of the imperfect information externality, or the total impact on the resource constraint of the economy. This is formally presented in the following proposition, where a more general version of the tax formula 4 is presented, focusing on the simpler case where the resume is defined as the cumulative discounted sum of deliverables.

**Proposition 8.** If a tax schedule is optimal then it needs to satisfy the following relationship:

\[
E\left[\left(\frac{\chi(y) - r(y)}{r(y)}\right)e_r(y)\right]g(y)y = \int_y^\infty g(\tilde{y}) \left(1 - \frac{E[\lambda(\tilde{y})]}{E[\epsilon_r(y)]}\right) d\tilde{y} + \int_y^\infty E\left[\left(\frac{\chi(\tilde{y}) - r(\tilde{y})}{r(\tilde{y})}\right)\eta_r(\tilde{y})\right]g(\tilde{y}) d\tilde{y},
\]

(10)

where \(\chi(y) \equiv v(y, \theta)/y'(h(y))\), that is, how much more product is generated per unit of wages for someone who is currently earning \(y\), which conditional on the income level can still depend on the type \(\theta\).

**Proof.** See Appendix Section A.13.

If there are no income effects, we can write: \(r(y) = r_m(y) \cdot r_d(y)\), where \(r_d(y) = \frac{E[\chi(y)e_r(y)]}{E[\epsilon_r(y)]}\), and

\[
\left(1 - \frac{r_m(y)}{r_m(y)}\right)E[\epsilon_r(y)]g(y)y = \int_y^\infty g(\tilde{y}) \left(1 - \frac{E[\lambda(\tilde{y})]}{E[\epsilon_r(y)]}\right) d\tilde{y}.
\]

In this case, the formula looks very similar to Diamond (1973) equation 10 in the context of a model of Pigouvian taxation with limited instruments, linear utilities in income and separable in externalities. As in that paper, the elasticity weights how much the externality matters. However, here there are no further multiplier effects: what we call externalities in this model – the fact that workers appropriate more than their marginal products when they work more – does not impact the labor supply of others except to the extent that salaries change.

### 6.3 Richer Signal Structure, Exogenous Signals

The baseline version of the model features a very simple signal structure. This simplicity allows us to clearly understand how incentives and the distribution of income interact when there are career concerns. However, it leaves out important elements of real labor markets, such as richer signals that firms can extract from workers. For example, from the point-of-view of firms it may be very clear that some workers are on different career tracks, and that those workers can hardly change that. For the government, it may be harder to distinguish these workers, and even harder to create taxes that are specific to each career track. This
possibility can be accommodated by introducing exogenous signals that the only firms and not the government sees.

If firms see additional exogenous signals the government does not see, then we cannot apply proposition 2. The reason behind it is that there are multiple pretax salary functions, a different career path for each signal realization. However, we can still consider a variation in the income retention schedule and track how this variation affects the salaries of different careers change in response to it. The following proposition uses this idea to derive comparable tax formulas in this more complex environment, but focusing on the simpler case where the resumes are defined as the cumulative discounted sum of deliverables.

**Proposition 9.** If a tax schedule is optimal then it needs to satisfy the following relationship:

\[
E_z \left[ \int_y^\infty \left( \frac{\chi(y, z) - r_y}{r_y} \right) c^e_{r_y}(y, z) g(y|z) y \, dr_h(y; z) \, dy \right] = E_z \left[ \int_y^\infty \left( \frac{\chi(y, z) - r_y}{r_y} \right) g(y|z) \, dr_h(y; z) \, dy \right] 
+ E_z \left[ \int_y^\infty \left( \frac{\chi(y, z) - r_y}{r_y} \right) \eta^e(y) g(y|z) \, dr_h(y; z) \, dy \right]
\]

where \( \chi(y, z) = \frac{v(h(y, z); z)}{y'(h(y, z); z)} \), and \( \frac{dr_h(y; z)}{dr_y} \) is the response of post-tax salaries of someone who initially earn income \( y \), and who gets the signals \( z \).

**Proof.** See Appendix Section A.14. \( \square \)

In this case the optimal policy described by a weighted version of the standard first order condition, where weights are given by how much post tax wages change when income taxes change at different career paths. Taking into account this heterogeneity across career tracks may attenuate the size of the corrective component of taxes. For example, if there are no income effects, then \( \frac{dr_h(y; z)}{dr_y} = 0 \) for \( \tilde{y} \neq y \), and the formula above reduces to:

\[
E_z \left[ \frac{\chi(y, z) - r_y}{r_y} c^e_{r_y}(y, z) g(y|z) y \, dr_h(y; z) \, dy \right] = E_z \left[ \frac{dr_h(y; z)}{dr_y} \int_y^\infty \left( 1 - \lambda(\tilde{y}; z) \right) g(\tilde{y}|z) \, d\tilde{y} \right]
\]

For any signal realization \( z \), the weights \( \frac{dr_h(y; z)}{dr_y} \) are equal to one if there is no information asymmetry and are less than one if there is any informational asymmetry.\(^{40}\) In this sense, relative to the basic Equation 4, holding fixed the other sufficient statistics, taking into account this source of heterogeneity would call for relatively lower marginal taxes, attenuating the intensity with which informational asymmetries push towards higher marginal taxes.

\(^{40}\)For more details on how changes in salaries and the degree of informational asymmetry \( \chi \) are related see section 5.2.
7 Conclusion

While incomplete information is a key feature of labor markets, standard benchmark models of taxation often ignore it. 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 the impact of these decisions today on future wages. In this model, the interest from firms on learning the productivity of workers arises even when firms are allowed to pay-for-performance and both workers and firms are risk neutral. Moreover, we incorporated optimal taxation in this model, deriving generalized Mirrleesian formulas that apply not only to this particular model of imperfect information in the labor market, but more generally to models with labor market frictions, as long as the mechanism that explains how firms set wages satisfies some simple conditions. Furthermore, the main insights from the optimal taxation formulas hold under several extensions of the basic model, including richer signal structures, human capital accumulation, and multidimensional heterogeneity. These generalized formulas, of independent interest, can be applied to other setups, such as health insurance and financial markets, where taxes may need to play a dual role: correct for informational frictions or other sources of externalities, and redistribute between different types of workers.

These formulas decompose optimal taxes into two components: a redistributive component, and a corrective component. 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 second, especially in the context of dynamic imperfect information in labor markets. Using data from the Health and Retirement Study survey this paper has shown that for an average worker, the corrective component of taxes is of the order of 5%, while for high earners it ranges from 10% to as high as 60%. This result has implications the redistributive effects of the tax system, and is consistent with the view that the current tax system may be less redistributive than it would thought, if imperfect information was not taken into account.
References


55
A Appendix

A.1 Salaries with Overlapping Generations (Proof of Lemma 1)

Lemma. Given the overlapping generations structure of the model, equation 2 is equivalent to:

\[ w(h(s)) = \mathbb{E}[v(\theta)|h(\theta) \geq h(s)] \]  \hspace{1cm} (11)

Proof. First notice that for every type who supplies \( h(\theta) > h(s) \) over their lifetime, there is someone (potentially from a different cohort) who now has the experience \( h(s) \). Thus \( \mathbb{E}[v(\theta)|h(\theta, s) = h(s)] = \mathbb{E}[v(\theta)|h(\theta) \geq h(s)] \). Now suppose a firm sets salaries that are not equal to the average productivity, so that equation above is violated for some \( h(s) \). Then, a firm that offers a contract conditional on the experience level \( h(s) \) is either making losses or positive profits. In the first case, the firm would be better off by not offering the contract and in the second case a firm could enter the market offering an infinitesimally lower price and making strictly positive profits. \( \square \)

A.2 Algebra Behind Example in Section 3.1

Salaries are given by the expectation of productivity of those who provide at least \( h \). In terms of types salaries are given by \( w(h(\theta)) = \mathbb{E}[v(\tilde{\theta})|\tilde{\theta} \geq \theta] = \frac{\alpha}{\alpha - \delta} \theta^\delta \). The first order conditions of the worker imply that:

\[ w(h) = h^{1/\epsilon} b(\theta)^{-(1+1/\epsilon)} \]

Guess and verifying that the wage is a power function, i.e., \( w(h) = kh^\gamma \), for some \( k \) and \( \gamma \). We conclude that:

\[ w(h) = kh^\gamma = \frac{\alpha}{\alpha - \delta} k^{-\frac{\delta}{1+\epsilon}} h^{-\frac{\eta}{1+\eta}} \]

And therefore

\[ k = \left( \frac{\alpha}{\alpha - \delta} \right)^{\frac{1}{1+\delta}(1+\epsilon) \gamma} \]

\[ \gamma = \frac{\delta}{1 - \delta + \epsilon} \]
A.3 Equilibrium Existence Without Taxes

Proposition 10. Assuming $v(\theta)$ smoothly increasing in types, MRS smoothly decreasing in types, consumption and labor, and a continuous distribution of types, there exists an equilibrium.

Proof. Consider the direct mechanism that offers the allocation $c(\theta), h(\theta)$, where $c(\theta)$ and $h(\theta)$ is the solution to the following system of equations.

$$\frac{c'(\theta)}{h'(\theta)} - \text{MRS}(c(\theta), h(\theta), \theta) = 0$$

$$\mathbb{E}[v(\tilde{\theta})|\tilde{\theta} \ge \theta] = \frac{c'(\theta)}{h'(\theta)}$$

$$\int (c(\theta) - v(\theta)h(\theta))f(\theta)d\theta = 0$$

This allocation satisfies IC, feasibility. The first is equation is the local IC. $c(\theta)$ and $h(\theta)$ are increasing in $\theta$ because $\mathbb{E}[v(\tilde{\theta})|\tilde{\theta} \ge \theta]$ is increasing in $\theta$ and MRS is decreasing in $\theta$, and increasing in $c$ and $h$. Because of single-crossing the local IC and monotonicity together imply that incentive constraints are globally satisfied. The third equation is the feasibility constraint. In this case, the feasibility constraint is automatically satisfied given the first two equations, as workers get paid what they produce in expectation, firms are offering salaries $\mathbb{E}[v(\tilde{\theta})|\tilde{\theta} \ge \theta]$ and making zero profits. Notice as well that $\int_0^\theta v(\tilde{\theta})h'(\tilde{\theta})d\tilde{\theta} < v(\theta)\int_0^\theta h'(\tilde{\theta})d\tilde{\theta} \le v(\theta)h(\theta)$, thus $c(\theta) > 0$, that is, higher types workers in this allocation subsidize lower type workers and consumption is positive everywhere. The existence of a solution is guaranteed by the Picard-Lindelöf theorem, given the regularity assumptions on MRS, $v$ and the distribution of types.

A.4 Atkinson-Stiglitz with “Double Adverse Selection” (Proof of Proposition 1)

Proof. Any tax system generates a common budget set $B$ that determines which pairs $(C, H)$ are feasible. As a first step in the proof, we are going to show that we can replicate this budget set for the workers and save resources, without imposing taxes on timed consumption or labor flows. To do that, we generate a new tax system where, for each pair $(C, H)$ the new pre tax income is $e(H)$ and post tax income is $e(C)$, where these are defined below:

$$e(C) = \min_{\tilde{c}(\cdot)} \int q(t)\cdot \tilde{c}(t)dt \text{ st. } C(\tilde{c}(\cdot)) \ge C$$

$$e(H) = \max_{\tilde{h}(\cdot)} \int q(t)\cdot w(h(\cdot))\cdot \tilde{h}(t)dt \text{ st. } H(\tilde{h}(\cdot)) \le H$$

Under this new tax system, the worker problem can be written in three parts:
\[
\max_{C, H} U(C, H, \theta) \text{ st. } (C, H) \in B \text{ and } e(C) = \min_{\tilde{c}(\cdot)} \int q(t) \cdot \tilde{c}(t) dt \text{ st. } C(\tilde{c}(\cdot)) \geq C \text{ and } e(H) = \max_{\tilde{h}(\cdot)} \int q(t) \cdot w(h(t)) \cdot \tilde{h}(t) dt \text{ st. } H(\tilde{h}(\cdot)) \leq H
\]

Notice that \(e(H)\) is the maximum pre-tax income that can be generated by generating at most the disutility \(H\). Moreover, it depends only on lifetime labor supply and not on the timing of these labor supply decisions, because \(\int q(t) \cdot w(h(t)) \cdot \tilde{h}(t) dt = W(h)\). Thus, it is the maximum lifetime labor that generates at most the disutility \(H\). Because production depends only on lifetime labor supply (and is increasing in lifetime labor supply), it is also the maximum production that generates at most the disutility \(H\).

Further, because \(e(C)\) is the smallest amount of resources that achieves the subutility level \(C\), and \(e(H)\) is the maximum production that can be generated by generating at most the disutility \(H\), whenever choices change, there are more resources than required to obtain the same allocation. These extra resources can be used to increase all the consumption possibilities \((C + \Delta(H), H)\) by some small amount \(\Delta(H)\), chosen in such way that everyone still prefers their originally labor aggregate choice \(H\).

We have assumed that the discount rates \(q(\cdot)\) were exogenously given, and in this case we have shown that there is a lifetime income taxation system that is weakly better than any other tax system, as claimed in the Proposition. An analogous argument extends the result to arbitrary endogenous discount rates \(q(\cdot)\), that is, discount rates that arise from competitive firms doing the intertemporal allocation of resources. Indeed, the result is implied by a production efficiency argument as in (Diamond and Mirrlees, 1971). Moreover, we assume throughout the paper that the economy is dynamically efficient in the sense that the present value of the output exists and it is finite.

\[\square\]

### A.5 Invertibility Condition (Proof of Lemma 2)

**Lemma.** Without loss, we can solve directly for \(\tilde{R}(h) = R(y(h))\), and then find \(y(h)\), and \(R(y)\). That is, the planner can solve the simpler problem:

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

**Proof.** Because \(y'(h)\) is a well-defined function of the allocation and is always positive, \(y(h)\) always exists and it is strictly increasing. Thus, there exists an inverse function \(h^{-1}(y)\). Therefore we can define \(R(y)\) so that \(R(y(h)) = \tilde{R}(h)\). Thus, we found the income tax schedule and equilibrium salaries that prevail in the economy where the planner solved directly for the retention function \(\tilde{R}(h)\). \[\square\]
A.6 Optimal Taxes with Single Dimensional Heterogeneity (Proof of Proposition 2)

Proof. The optimal tax schedule solves the following problem:

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

where $E[v(\theta)|MRS^{\theta}_{c,h} \leq \tilde{R}'(h)] = y'(h)$

$$\tilde{R}(h) = R(y(h))$$

Considering a small variation on marginal retention rates as a function of effort at a given level of effort:

$$E\left[ \mu \frac{dV(\tilde{r}, I, \theta)}{d\tilde{r}} \right] 1(h(\theta) \geq \theta) = -\mu E\left[ (v(\theta) - \tilde{r}(h(\theta))) \frac{dh^{c}(\theta)}{d\tilde{r}} 1(h(\theta) = h) - (v(\theta) - \tilde{r}(h(\theta))) \frac{dh(\theta)}{d\tilde{r}} 1(h(\theta) \geq h) - 1(h(\theta) \geq h) \right]$$

$$\left( v(h) - \tilde{r}(h) \right) f(h) \frac{dh^{c}(\theta)}{d\tilde{r}} = \int_{h}^{\infty} f(\tilde{h}) \left( 1 - \lambda(\tilde{h}) \right) d\tilde{h} + \int_{h}^{\infty} \left( v(\tilde{h}) - \tilde{r}(\tilde{h}) \right) f(\tilde{h}) \frac{dh}{d\tilde{h}} \tilde{R}(\tilde{h}) d\tilde{h}$$

Now two steps: writing it in terms of elasticities and converting to a formula in terms of earnings. First:

$$\left( \frac{v(h) - \tilde{r}(h)}{\tilde{r}(h)} \right) f(h) h \frac{dh^{c}(\theta)}{d\tilde{r}} \frac{\tilde{r}(h)}{h} = \int_{h}^{\infty} f(\tilde{h}) \left( 1 - \lambda(\tilde{h}) \right) d\tilde{h} + \int_{h}^{\infty} \left( \frac{v(\tilde{h}) - \tilde{r}(\tilde{h})}{R(\tilde{h})} \right) f(\tilde{h}) \frac{dh}{d\tilde{h}} \tilde{R}(\tilde{h}) d\tilde{h}$$

$$\left( \frac{v(h) - \tilde{r}(h)}{\tilde{r}(h)} \right) f(h) h \epsilon^{\tilde{r}}(h) = \int_{h}^{\infty} f(\tilde{h}) \left( 1 - \lambda(\tilde{h}) \right) d\tilde{h} + \int_{h}^{\infty} \left( \frac{v(\tilde{h}) - \tilde{r}(\tilde{h})}{R(\tilde{h})} \right) f(\tilde{h}) \eta^{\tilde{r}}(\tilde{h}) d\tilde{h}$$

Then, using that:

$$\tilde{r}(h) = r(y(h))y'(h)$$

$$\frac{dy(h)}{dr} = y'(h) \frac{dh}{d\tilde{r}} \frac{d\tilde{r}}{dr} = y'(h) \frac{dh}{d\tilde{r}} \text{ (micro elasticities)}$$

$$f(h) = g(y(h))y'(h)$$

$$\frac{dy(h)}{dI} = y'(h) \frac{dh}{dI}$$

$$dh = \frac{1}{y'(h)} dy$$
\[
\left( \frac{v(h)/y(h) - r(g(h))}{r(y(h))} \right) g(y(h))y(h)\epsilon_r(y(h)) = \int_{h}^{\infty} f(\tilde{h}) \left( 1 - \lambda(\tilde{h}) \right) d\tilde{h} \\
+ \int_{h}^{\infty} \left( \frac{v(\tilde{h})/y'(\tilde{h}) - r(y(\tilde{h}))}{R(y(\tilde{h}))} \right) f(\tilde{h})\eta_I(y(\tilde{h}))d\tilde{h}
\]

\[
\left( \frac{v(y)/y'(h(y)) - r(y)}{r(y)} \right) g(y)y\epsilon_r(y) = \int_{y}^{\infty} g(\tilde{y}) \left( 1 - \lambda(\tilde{y}) \right) d\tilde{y} + \int_{y}^{\infty} \left( \frac{v(\tilde{y})/y'(h(\tilde{y})) - r(\tilde{y})}{R(\tilde{y})} \right) g(\tilde{y})\eta_I(\tilde{y})d\tilde{y}
\]

### A.7 Elasticities

The optimal tax formulas presented in Section 4 feature compensated and income elasticities. In the formulas, because we are keeping the pretax salaries fixed, these elasticities can be called “micro elasticities”. A number of empirical studies use aggregate variation in taxes (such as state-tax variations, or kinks in the tax schedule) to estimate elasticities of taxable income. These statistics are better described as “macro elasticities” – how income changes when the tax schedule changes for everyone, potentially affecting pretax salaries. That is, these elasticities are inferred from the observation of how incomes from people living in different states react to state tax reforms, or how income in different years reacted differently to federal tax reforms (Gruber and Saez, 2002). When measuring the change in individual taxable income, they conflate the change in behavior that responds directly to the changes in marginal taxes and the change in behavior that responds to changes in wages that are induced by these economy-wide tax reforms. To extract the micro elasticities from the estimated macro elasticities it is necessary to rescale them up to account for the endogenous changes in wages due to imperfect information in labor markets. Increases in marginal retention induce the marginal types – who are the less productive types – to work more, and therefore reduce pretax salaries, making the effective change in post-tax wages smaller. Thus, the estimated elasticities of taxable income are lower in magnitude than the micro elasticities that keep the pretax salaries fixed. Proposition 11 relates steady state compensated “micro elasticities” to “macro elasticities”, in the case where resumes are defined as the cumulative discounted sum of deliverables a worker has produced.

**Proposition 11.** Compensated “micro elasticities” and “macro elasticities” are related by the following formula:

\[
\epsilon^y_{r,y,m}(y) = \frac{\epsilon^y_{r,y,M}(y)}{1 - \alpha(y)\epsilon^y_{r,y,M}(y) \left( 1 - \chi(y) \right)}
\]

“Macro elasticities” are lower than “micro elasticities”. A locally flat increase in marginal retention at pretax income level \( y \) makes workers who were just indifferent at that region...
increase their labor supply, lowering the average productivity of the workers at \( y \), and therefore lowering pretax wages, and attenuating the original increase in labor supply. How much salaries decrease is proportional to the density of people at \( y \) relative to the mass of people above \( y \) (from which the shape parameter \( \alpha(y) = \frac{g(y)y}{1-G(Y)} \) shows up in the formula), to how much people are changing their income (\( \epsilon_{y,M}(y) \)), and to how far their productivities are from their salaries (hence \( 1 - \chi(y) \) in the formula). The “micro elasticity” can be thought of as renormalizing the original elasticity by the effective change in retention, coming both from the mechanical change induced by the reform and from the endogenous change in salaries.

These are “own elasticities”: they tell how income changes for someone who initially is earning \( y \) as a response to a change in marginal rates at the same income level. Changes in marginal retention in other income levels affect salaries in multiple other income levels: the pool of workers from each income level is shifted, generating further compensated and uncompensated changes. For this reason, the expression for income elasticities is also more involved. A change in the intercept of the tax schedule generates not only further income effects but also further compensated effects. Proposition 12 relates the change in pretax salaries to changes in income.

**Proposition 12.** Income “macro” elasticities and changes in pretax salaries and income “micro” elasticities are related by the following formula:

\[
\epsilon_{y}^{(h(y))} = -\alpha(y)\epsilon_{y,M}(y)\left(1 - \chi(y)\right)
\]

\[
\eta(y) = \frac{\epsilon_{y,M}(y)\left(1 + \epsilon_{y,M}^{(h)}(y)\alpha(y)\left(1 - \chi(y)\right)\right)}{1 - \int_{0}^{y} \alpha(\tilde{y})\epsilon_{y,M}^{(h)}(\tilde{y})\left(1 - \chi(\tilde{y})\right)\frac{y'(\tilde{y})}{R(\tilde{y})}d\tilde{y}}
\]

There are two effects playing a role in the relationship between income “micro” and “macro” elasticities. First, assuming that income elasticities are negative, an income transfer to all workers induces them to work less, and those who work less are the least productive, pushing towards higher salaries. Now because salaries increase, a compensated effect is increasing the labor supply of workers, making the income “macro” elasticity higher (or smaller in absolute value) than the income “micro” elasticity. Second, because salaries are increasing not only at a given income level \( y \), but everywhere below (and above) it, the income transfer is effectively higher than what the “macro” elasticity accounts for. Thus, this is a force making the “micro” elasticity lower (higher in absolute value) than the “macro” elasticity. Thus, whether the “macro” income elasticity or the “micro” is larger depends on which effect dominates the other.

It should be noticed that these elasticities and the relationship between them hold in the
steady state. Moreover, they are long-term elasticities – they look at how lifetime income changes as a response to tax reforms. As long-term elasticities, the empirical evidence on the magnitude of these is relatively scarcer compared to elasticities over shorter horizons.

A.7.1 Proofs of Propositions 11 and 12

Proposition: compensated “micro elasticities” and “macro elasticities” are related by the following formula:

\[
\epsilon_{r_y}^{y,m} = \frac{\epsilon_{r_y}^{y,M}}{1 - \alpha(y)\epsilon_{r_y}^{y,M}(1 - \chi(y))}
\]

Proof. Because

\[
\frac{d\log r_y}{d\log y} = \frac{d\log y'}{d\log y} + 1
\]

\[
\epsilon_{r_y}^{y,M} = \epsilon_{r_y}^{y,m}(1 + \epsilon_{r_y}^{y,(h(y))})
\]

and thus

\[
\epsilon_{r_y}^{y,(h(y))} = -\alpha(y)\epsilon_{r_y}^{y,m}(1 + \epsilon_{r_y}^{y,(h(y))})(1 - \chi(y))
\]

\[
\Rightarrow \epsilon_{r_y}^{y,(h(y))} = \frac{-\alpha(y)\epsilon_{r_y}^{y,m}(1 - \chi(y))}{1 + \alpha(y)\epsilon_{r_y}^{y,m}(1 - \chi(y))}
\]

\[
\epsilon_{r_y}^{y,m} = \frac{\epsilon_{r_y}^{y,M}}{1 + \epsilon_{r_y}^{y,(h(y))}} = \epsilon_{r_y}^{y,M} \left( 1 + \alpha(y)\epsilon_{r_y}^{y,m}(1 - \chi(y)) \right)
\]

\[
\Rightarrow \epsilon_{r_y}^{y,m} = \frac{\epsilon_{r_y}^{y,M}}{1 - \alpha(y)\epsilon_{r_y}^{y,M}(1 - \chi(y))}
\]

Proposition. Income “macro” elasticities and changes in pretax salaries and income “micro” elasticities are related by the following formula:

\[
\epsilon_{r_y}^{y,(h(y))} = -\alpha(y)\epsilon_{r_y}^{y,M}(y)(1 - \chi(y))
\]

\[
\eta(y) = \frac{\epsilon_{r_y}^{y,M}(y)\left( 1 + \epsilon_{r_y}^{y,m}(y)\alpha(y)(1 - \chi(y)) \right)}{1 - \int_0^y \alpha(\tilde{y})\epsilon_{r_y}^{y,M}(\tilde{y})(1 - \chi(\tilde{y}))\frac{\epsilon^{y,(h(\tilde{y}))}}{R(\tilde{y})}d\tilde{y}}
\]

63
Proof. Because
\begin{align*}
dz &= \frac{dz}{dr_h} \cdot dr_h + \frac{dz}{df} \cdot df \\
dz I_o &= \frac{dz}{dr_h} \cdot r_h \cdot I_o \cdot \frac{dr_h}{df} + \frac{dz}{df} \cdot I_o \cdot \frac{df}{dr_h}
\end{align*}

\[ e_{I}^{y,m} = e_{r_h}^{y,m} e_{I}^{y} + \eta^{y} \frac{df}{df} \]

\[ \frac{df}{df} = 1 + \int_{0}^{y} \frac{dr_h}{df} dy = 1 + \int_{0}^{y} \frac{I_o}{y'(h(\tilde{y}))} \frac{dy'(h(\tilde{y}))}{df} d\tilde{y} = 1 + \int_{0}^{y} \frac{\epsilon_{r_h}^{y,m} e_{I}^{y}}{df} dy \\
\]

\[ \epsilon_{I}^{y,M} = -\epsilon_{r_h}^{y,m} \alpha(y) e_{I}^{y,M}(y) \left( 1 - \chi(y) \right) + \eta(y) \left( 1 - \int_{0}^{y} \alpha(\tilde{y}) e_{I}^{y,M}(\tilde{y}) \left( 1 - \chi(\tilde{y}) \right) \frac{y'(h(\tilde{y}))}{R(\tilde{y})} d\tilde{y} \right) \]

\[ \eta(y) = \frac{\epsilon_{I}^{y,M}(y) \left( 1 + e_{r_h}^{y,m}(y) \alpha(y) \left( 1 - \chi(y) \right) \right)}{1 - \int_{0}^{y} \alpha(\tilde{y}) e_{I}^{y,M}(\tilde{y}) \left( 1 - \chi(\tilde{y}) \right) \frac{y'(h(\tilde{y}))}{R(\tilde{y})} d\tilde{y}} \]

A.8 First Best with Pigouvian Taxes

Proposition 13. Assuming \( v(\theta) \) smoothly increasing in types, MRS smoothly decreasing in types, and a continuous distribution of types, a first best allocation can be achieved, with \( r(y) = \chi(y) \).

Proof. Consider the direct mechanism that offers the allocation \( c(\theta), h(\theta) \), where \( c(\theta) \) and \( h(\theta) \) is the solution to the following system of equations.

\[ \frac{c'(\theta)}{h'(\theta)} - MRS(c(\theta), h(\theta), \theta) = 0 \]

\[ v(\theta) = \frac{c'(\theta)}{h'(\theta)} \]

\[ \int (c(\theta) - v(\theta) h(\theta))) f(\theta) d\theta = 0 \]

This allocation satisfies IC, feasibility. The first is equation is the local IC. \( c(\theta) \) and \( h(\theta) \) are increasing in \( \theta \) because \( v(\theta) \) is increasing in \( \theta \) and MRS is decreasing in \( \theta \), and increasing in \( c \) and \( h \). Because of single-crossing the local IC together with monotonicity implies global IC's are satisfied. The third equation is the feasibility constraint. Notice as well that \( \int_{0}^{\theta} v(\tilde{\theta}) h'(\tilde{\theta}) d\tilde{\theta} < v(\theta) \int_{0}^{\theta} h'(\tilde{\theta}) d\tilde{\theta} \leq v(\theta) h(\theta) \), thus \( c(\theta) > 0 \), that is, higher
types workers in this allocation subsidize lower type workers and we do not need to worry about consumption being negative. Further, notice that in this allocation \( r(y) = \chi(y) \). The existence of a solution to these differential equations is guaranteed by the Picard-Lindelöf theorem, given the regularity assumptions on \( MRS, v \) and the distribution of types.

Finally, we show that it is a first best allocation: it corresponds to an allocation where workers face linear budgets and get paid their marginal products while receiving transfers \( I(\theta) = c(\theta) - w(\theta)h(\theta) \).

### A.9 Welfare (Proof of Proposition 5)

**Proposition.** If the original tax schedule is optimal and the planner has strongly enough redistributive preferences, and labor is a normal good, then decreasing information asymmetries in labor markets decreases welfare.

**Proof.** We can set the planner’s problem as maximizing a welfare function of workers utility, subject to incentive compatibility constraints and a feasibility constraint.

\[
\max_{l(\theta), c(\theta)} \int W(u(c(\theta), l(\theta), \theta))f(\theta)d\theta \\
\text{s.t. } u(c(\theta), l(\theta), \theta) \geq u(c(\theta'), l(\theta') \frac{b(\theta')}{b(\theta)}, \theta) \forall \theta, \theta' \text{ [IC's]} \\
\int (c(\theta) - v(\theta)b(\theta)l(\theta)f(\theta)d\theta \leq 0 \text{ [Feasibility]}
\]

The incentive compatibility constraints translate the idea that under the results from Lemma 2 (or the more general Lemma 4) the planner can solve for the allocation in terms of consumption and deliverables \( h(\theta) \) subject to workers not being willing to misreport their types and trade their allocation \( (c(\theta), h(\theta)) \) to another allocation \( (c(\theta'), h(\theta')) \), and that the amount of deliverables is equal to the product of effort \( l(\theta) \) and observable component of productivities \( b(\theta) \), i.e. \( l(\theta) \cdot b(\theta) = h(\theta) \).

In our context the results in Seade (1982); Werning (2000), imply that out of the set of IC’s, only the local downward are binding, provided that \( W'(U)U_C \) is positive and decreasing, and that leisure is a normal good (they also further imply \( v - r \) is positive,. For more details, see Appendix Section A.10 and the argument in Werning (2000)). Now, notice that those downward incentive compatibility constraints become tighter whenever informational asymmetries decrease, while the feasibility constraint is unchanged. Thus, welfare decreases. The result does not depend on the particulars of the information structure and besides holding in the MWS case (as in Stantcheva (2014)), and in other models of competition with imperfect information such as the competitive screening model in Azevedo and Gottlieb
A.10 Formula in Terms of Types (Proof of Proposition 6)

**Proposition.** Optimal taxes as a function of types $\theta$ must satisfy the following equations:

$$r_y(\theta) = r_m(\theta) \cdot \chi(\theta)$$

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

**Proof.** Let’s set up the planners problem as:

$$\max_{u,h} \int \lambda(\theta) u(\theta) f(\theta) d\theta$$

s.t. $u'(\theta) = U_\theta(e(u(\theta), h(\theta), \theta), h(\theta), \theta) - v(\theta) h(\theta)) f(\theta) d\theta \leq 0$

Lagrangian

$$\max_{u,h} \int \lambda u f + \mu(u' - U_\theta) - \kappa(e - vh) f d\theta$$

Integrate by parts

$$\max_{u,h} \int \lambda u f - \mu' u - \mu U_\theta - \kappa(e - vh) f d\theta + u \mu' \frac{\bar{\theta}}{\theta}$$

FOC’s for $u(\theta)$, and $h(\theta)$

$$\lambda f - \mu' - \mu U_{\theta,c} e_u - \kappa e_u f = 0$$

$$\mu(U_{\theta,c} e_h + U_{\theta,h}) - \kappa e_h f + k v f = 0$$

Replacing $e_u = U_c^{-1}$ on the first equation.

$$\lambda U_c f - \mu' U_c - \mu U_{\theta,c} f \kappa = 0$$
Define $\hat{\mu} = \mu_c / \kappa$ and $\hat{\lambda} = \lambda U_c / \kappa$, plus some rearrangement (as in Scheuer and Werning (2017))

$$\hat{\lambda} f - \hat{\mu}' - \hat{\mu} MRS_c h' = f$$

Rearrange second line and take a derivative

$$\hat{\mu} U_{\theta,c} MRS + U_{\theta,h} U_c MRS = f\frac{v - MRS}{MRS}$$

$$-\hat{\mu} \frac{\partial \log MRS}{\partial \theta} = f\frac{v - MRS}{MRS}$$

We arrived at two equations:

$$\hat{\lambda} f - \hat{\mu}' - \hat{\mu} MRS_c h' = f$$

$$-\hat{\mu} \frac{\partial \log MRS}{\partial \theta} = f\frac{v - r}{r}$$

Differentiating the second equation:

$$-\hat{\mu}' = f' \frac{v - r}{r} + f \frac{d}{d \theta} \left( \frac{v - r}{r} \right) - \frac{\partial^2 \log MRS}{\partial \theta^2} f \frac{v - r}{r}$$

$$-\hat{\mu} = f \frac{v - r}{\partial \log MRS}$$

Plugging back in the first equation:

$$\hat{\lambda} f + f' \frac{v - r}{r} + f \frac{d}{d \theta} \left( \frac{v - r}{r} \right) - \frac{\partial^2 \log MRS}{\partial \theta^2} f \frac{v - r}{r} + \frac{\partial^2 \log MRS}{\partial \theta^2} MRS_c h' = f$$

$$\hat{\lambda} - 1 + f'/f \frac{v - r}{r} + \frac{d}{d \theta} \left( \frac{v - r}{r} \right) - \frac{\partial^2 \log MRS}{\partial \theta^2} \frac{v - r}{r} + \frac{v - r}{\partial \log MRS} MRS_c h' = 0$$

$$\frac{v - r}{r} \left( \frac{\partial \log MRS}{\partial \theta} \right)^{-1} \left( \frac{f'}{f} - \frac{\partial^2 \log MRS}{\partial \theta^2} + MRS_c h' + \frac{d \log}{d \theta} \left( \frac{v - r}{r} \right) \right) = 1 - \hat{\lambda}$$
\[
\frac{v - r}{r} \left( f' \left( \frac{\partial \log MRS}{\partial \theta} \right) \right)^{-1} - f \left( \frac{\partial^2 \log MRS}{\partial \theta^2} \right) + \\
f \left( \frac{\partial \log MRS}{\partial \theta} \right)^{-1} \frac{d}{d\theta} \left( \frac{v - r}{r} \right) = (1 - \hat{\lambda})f - \left( \frac{v - r}{r} \right) f \left( \frac{\partial \log MRS}{\partial \theta} \right)^{-1} \text{MRS} c h'
\]

Integrate both sides with respect to \( \theta \)

\[
\frac{v - r}{r} f \left( \frac{\partial \log MRS}{\partial \theta} \right)^{-1} \bigg|_\theta^\infty = \int_\theta^\infty (1 - \hat{\lambda})f d\theta - \int_\theta^\infty \left( \frac{v - r}{r} \right) \text{MRS} c h' \frac{d}{d\theta} f d\theta
\]

Notice that: \( \eta(\theta) = \frac{d}{d\theta} \frac{\text{MRS} c h'}{\partial \log MRS} \),

Using that \( \lim_{\theta \to \infty} \frac{v - r}{r} f \left( \frac{\partial \log MRS}{\partial \theta} \right)^{-1} = 0 \):

\[
\frac{v - r}{r} f \left( - \frac{\partial \log MRS}{\partial \theta} \right)^{-1} = \int_\theta^\infty (1 - \hat{\lambda})f d\theta + \int_\theta^\infty \left( \frac{v - r}{r} \right) \eta d\theta
\]

Finally, notice that \( \frac{v - r}{r} = \frac{\chi - ry}{ry} \)

\[
\frac{\chi - ry}{ry} f \left( - \frac{\partial \log MRS}{\partial \theta} \right)^{-1} = \int_\theta^\infty (1 - \hat{\lambda})f d\theta + \int_\theta^\infty \left( \frac{\chi - ry}{ry} \right) \eta d\theta
\]

A.11 Conditioning on the “Strength of the Resume” (Proof of Proposition 3)

In this section, we assume that signals take the form of \( h_\phi(\tilde{h}(\tilde{a})_0, a) = \int_0^a \phi(\tilde{a}, a)\tilde{h}(\tilde{a})d\tilde{a} \), with \( \phi(\tilde{a}, a) > 0 \), continuous in \( a \) and \( \tilde{a} \).

Using the logic from Proposition 2 and the assumption of homogeneous preferences over the timing of labor supply and consumption as in Proposition 1, we will show that optimal taxation formulas will still take a simple structure.

We first set aside the issue of implementation, and assume and characterize the set of optimal incentive compatible allocations. Those allocations need to satisfy two properties, as stated in Lemma 3.

Lemma 3. Any optimal and incentive compatible allocation satisfies two properties:

Efficient timing: For any \( H(\theta), \tilde{h}(:,:,\theta) = \arg\max \int_0^1 q(a)\tilde{h}(a)da \) st. \( H(\tilde{h}(:,:,\theta)) = H \)
Lifetime optimality:

\[
\left(\frac{F_H(\theta, H) - \bar{r}(H)}{\bar{r}(H)}\right)f(H)H\epsilon(H) = \int_H^\infty f(\bar{H})\left(1 - \lambda(\bar{H})\right)d\bar{H} + \int_H^\infty \frac{\left(F_H(\theta, H) - \bar{r}(H)\right)}{\bar{r}(H)}f(\bar{H})\eta(\bar{H})d\bar{H},
\]

where \(F(\theta, H) = \max_v v(\theta) \int q(a)\bar{h}(a)da\) s.t. \(H(\bar{h}(\cdot)) = H\), and \(\bar{r}(H) = \bar{R}'(H)\).

Proof. Efficient timing is an implication of production efficiency theorem of Diamond and Mirrlees (1971). Lifetime optimality follows from standard variational argument over a retention schedule \(R(H)\), which maps choices of \(H\) to an assigned consumption \(C = R(H)\), where we assume that no bunching takes place at the optimal assignment.

Lemma 4. Assuming \(\frac{d^2H(\bar{h}(\cdot)\beta)}{dh(\alpha)dh(\alpha')} < 0\), \(MRS(C, H, \theta)\) decreasing in \(\theta\), \(I(\bar{h}(\cdot)\beta, a) = \int_0^a \phi(a, \bar{a})\bar{h}(\bar{a})\), salaries are increasing in \(I\).

Proof. Because \(MRS(C, H, \theta)\) is decreasing in \(\theta\), for any \(R(H)\) strictly increasing – which is a property of the optimal allocation described in Lemma 3 – higher types \(\theta\) pick higher levels of \(H\). Since \(\frac{d^2H(\bar{h}(\cdot)\beta)}{dh(\alpha)dh(\alpha')} < 0\), higher levels of \(H\) are followed by higher levels of each \(\bar{h}(a)\). Therefore, conditional on age, higher types will pick strictly higher indexes, and thus for any \(I > I'\) and \(a > 0\), \(\mathbb{E}[v(\theta)|I, a] > \mathbb{E}[v(\theta)|I', a]\). Therefore, \(\mathbb{E}[v(\theta)|I] > \mathbb{E}[v(\theta)|I']\), and salaries are increasing in \(I\).

Lemma 5. No two different continuous sequences \(\bar{h}(a)\) and \(\bar{h}'(a)\) map into the same \(\bar{y}(a)\).

Proof. Remember that \(\bar{y}(\bar{h}(a), \bar{h}(\alpha)\beta) = \bar{h}(a) \cdot w(\bar{h}(\alpha)\beta) = \bar{h}(a) \cdot \mathbb{E}[v(\theta)|I(\bar{h}(\alpha)\beta)]\). Consider the first non zero measure interval where \(\bar{h}(a)_0\) and \(\bar{h}'(a)_0\) differ, and without loss consider a ball \((a, \bar{a})\) where \(\bar{h}(a) > \bar{h}'(a)\). If salaries were the same for both sequences, \(\bar{y}(\bar{h}(\alpha)\beta) > \bar{y}(\bar{h}'(\alpha)\beta)\), and the proof would be complete. However, by Lemma 4 salaries are increasing as a function of \(\bar{h}(a)\), thus indeed we have that \(\bar{y}(\bar{h}(\alpha)\beta) > \bar{y}(\bar{h}'(\alpha)\beta)\).

Lemma 5 allows us to decentralize any allocation with the properties from Lemma 3.

Proposition. If \(R(\bar{y}(\cdot)\beta)\) is optimal, then, there exists \(R_m, R_p\) with \(R(\bar{y}(\cdot)\beta) = R_m(R_p(\bar{y}(\cdot)\beta))\). Such that \(R_m\) and \(R_p\) satisfy the following conditions:

1. Intertemporal, Pigouvian: for any \(\bar{a}, \alpha\), and \(H(\bar{h}(\alpha)\beta) = H\), switching the timing of labor supply decisions and holding fixed \(H^*\), should leave lifetime earnings unaffected:

\[
\int_0^1 \frac{dR_p}{d\bar{y}(\bar{h}(\alpha)\beta)} \frac{\bar{d}h(\bar{a})q(\bar{a})}{dH} q(a)da = \int_0^1 \frac{dR_p}{d\bar{y}(\bar{h}(\alpha)\beta)} \frac{\bar{d}h(\bar{a})q(\bar{a})}{dH} q(a)da
\]
Notice that this only describes what happens when labor supply changes across time, holding $H$ fixed. So leaves both retention as a function of the index $H$, $\bar{R}_p(H)$, and retention as a function of the timing of earning $R_p(\tilde{y}(\cdot)_0^1)$, partially defined.

2. Lifetime, Pigouvian: increasing $H^*$ should increase lifetime earnings proportionally to the increase in output:

$$\bar{R}(H) = R_p(\tilde{y}(\bar{h}(\cdot)^0_0; H^*_0))$$

$$\bar{R}_p'(H) = \int_0^1 \frac{dR_p(\tilde{y}(\bar{h}(\cdot)^0_0; H^*_0))}{d\tilde{y}(\bar{h}(\cdot)^0_0; H^*_0)} \frac{d\tilde{y}(\bar{h}(\cdot)^0_0; H^*_0)}{d\bar{h}(\cdot)} \frac{d\bar{h}(\cdot)}{dH} da = v(H)$$

3. Lifetime, redistributive: Define the retention that workers face as $R_m(R_p(\tilde{y}(\cdot)))$, and $r_m = R_m'(R_p)$. After correcting for distortions, then $R_m$ should satisfy standard Mirrleesian formulas:

$$\left(1 - \frac{r_m(R_p)}{r_m(R_p)}\right)g(R_p)R_p^{e^\xi}(R_p) = \int_{R_p}^{\infty} g(R_p) \left(1 - \lambda(R_p)\right) d\bar{R}_p + \int_{R_p}^{\infty} \left(1 - \frac{r_m(R_p)}{r_m(R_p)}\right)g(R_p)\eta_{\xi}(R_p) d\bar{R}_p,$$

Proposition A.11 states that the tax system should be such that i) history dependent taxes ($R_p$) should be used to correct for labor wedges, and ii) after correcting for these distortions, lifetime income redistributive taxes should be imposed on top these taxes, according to standard redistributive formulas.

**Remark.** We can define $R_p$ to be such that:

$$\frac{dR_p(\tilde{y}(\cdot)_0^1)}{d\tilde{y}(\cdot)} = \int_\alpha^1 \frac{v(H)}{q(\cdot)w(\bar{h}(\cdot)_0^1)} \frac{v(H)}{\int_\alpha^1 \frac{d\tilde{y}(\cdot)^0_0}{dh(\cdot)} da}$$

where $v(H)$ is the marginal productivity of the type that supplies the level $H$ of labor, and where for ease of notation the dependence on $\tilde{y}(\cdot)_0^1$ is omitted. That is, the formula should be read as a function of earnings flows $\tilde{y}(\cdot)$, through the inverse operator $\bar{h}(\tilde{y}(\cdot)_0^1)$.

**Proof.** Notice it satisfies the conditions both the lifetime, and intertemporal Pigouvian properties stated in the previous Proposition. Now, notice that those conditions pin down the slope of $R_p$ with respect to any direction. Thus if they satisfy these two properties, they should also satisfy the property stated in the Remark above. \qed

A.12 Conditioning on the Full History of Completion of Deliverables (Proof of Proposition 4)

We assume in this section that preferences now take the form $U(C, \tilde{h}(\cdot), \theta)$ and that $\theta$ is high-dimensional, in the sense that, given a retention function $R(\tilde{h}(\cdot))$, strictly increasing
in $\tilde{h}(a)$ for any $a$, for any $\tilde{h}(\cdot)_0^1$, there exists a type $\theta$, for which supplying $\tilde{h}(\cdot)_0^1$ is optimal. We restrict household choices to the set of continuous $\tilde{h}(\cdot)$. We retain the assumption more productive types are more willing to provide the deliverables which in this case can be stated as: if $v(\theta) > v(\theta')$ then for any $\tilde{h}(\cdot)_0^1$, $C$, $MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot), \theta) < MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot), \theta')$, where $MRS_{C,\tilde{h}(a)}(C, \tilde{h}(\cdot)_0^1, \theta) = -\frac{U_{\tilde{h}(a)(\cdot)\tilde{h}(\cdot)}(C, \tilde{h}(\cdot), \theta)}{U_{C}(C, \tilde{h}(\cdot), \theta)}$.

**Lemma 6.** If the planner could choose the allocation, while being restricted to set the of incentive compatible allocations, any optimal allocation would lie at the frontier of production possibilities set.

**Proof.** Follows from analogous arguments from the production efficiency theorem of Diamond and Mirrlees (1971). That is, consider the problem

$$\max_{C,\tilde{h}(\cdot)} \mathbb{E}[W(U(C_\theta, \tilde{h}_\theta(\cdot), \theta))] \text{ s.t. } \mathbb{E}[v(\theta) \int_0^1 q(a)\tilde{h}_\theta(a)da - C_\theta] \geq 0$$

$$U(C_\theta, \tilde{h}_\theta(\cdot), \theta) \geq U(C_{\theta'}, \tilde{h}_{\theta'}(\cdot), \theta') \forall \theta, \theta'$$

By the taxation principle, we can incorporate the incentive compatibility constraints into the indirect utility function of the workers and solve the equivalent problem:

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

(12)

Now, suppose production takes place at the interior of the production possibility frontier. Then can increase $R(\tilde{h}_\theta(\cdot))$ uniformly (as lowering the price of the consumption good). Because the indirect utility function is increasing in $R$, everyone would be better off, and welfare would be higher. By the assumption that labor supply decisions are continuous in a uniform increase in $R$, there is a small enough increase in $R(\tilde{h}_\theta(\cdot))$ that does keeps the allocation inside the production possibilities set.

The second result is that the planner can use Pigouvian taxes to achieve the frontier of the production efficient set of allocations, because sequences of $\tilde{h}(\cdot)_0^1$ will map to sequences of $\tilde{y}(\cdot)_0^1$ one-to-one, as in Lemma 2 and Lemma 5.

**Lemma 7.** There is a positive return to experience. That is, in any optimal allocation, salaries $w(\tilde{h}(\cdot)_0^1) = \mathbb{E}[v(\theta)|\tilde{h}(\cdot)_0^1]$ are increasing in labor supply choices $\tilde{h}(\tilde{a})$, where $\tilde{a} \leq a$.

**Proof.** This is an immediate consequence of the assumption that the more productive types are more willing to provide the deliverables, and that the type space is rich enough so that for any path $\tilde{h}(\cdot)$, expectations are well-defined.
This Lemma, analogously to 4, establishes that there is positive return to experience.

**Lemma 8.** The planner can infer labor supply choices from earnings. That is, no two continuous $\tilde{h}(\cdot)\|_0$ map to the same $\tilde{y}(\cdot)\|_0$.

**Proof.** Remember that $\tilde{y}(\tilde{h}(\tilde{a})_{\|}) = \tilde{h}(\tilde{a}) \cdot E[v(\theta)|\tilde{h}(\tilde{a})_{\|}]$. Consider the first non zero measure interval where $\tilde{h}(\tilde{a})_{\|}$ and $\tilde{h}'(\tilde{a})_{\|}$ differ, and without loss consider a ball $(a, \bar{a})$ where $\tilde{h}(a) > \tilde{h}'(a)$. If salaries were the same for both sequences, $\tilde{y}(\tilde{h}(\tilde{a})_{\|}) > \tilde{y}(\tilde{h}'(\tilde{a})_{\|})$, and the proof would be complete. However, by the previous Lemma salaries are increasing as a function of $\tilde{h}(a)$, thus indeed we have that $\tilde{y}(\tilde{h}(\tilde{a})_{\|}) > \tilde{y}(\tilde{h}'(\tilde{a})_{\|})$.

Those results imply Pigouvian taxes should play an important role, as stated in the following Proposition.

**Proposition.** The planner can guarantee that the allocation would lie at the frontier of the production possibilities set by using Pigouvian taxes.

**Proof.** Consider the problem 12. This formulation can be thought of as solving for the redistributive wage schedule after Pigouvian taxes have been imposed, so that pre-tax salaries of the workers would have been equal to their productivities (or the average productivity of the workers with the same labor supply history $\tilde{h}(\cdot)$, if at the solution there are multiple types sharing the same history). The solution of this problem results in a wage schedule, $R(\tilde{h}(\cdot))$, which is as a function of labor supply decisions. This wage schedule, by the previous Lemma, can be written as a function of the history of earnings $\tilde{y}(\cdot)$, $R(\tilde{y}(\cdot))$, so it can be implemented with history dependent earnings taxes.

These Pigouvian taxes take the same general form as in the previous section. Thus, although this economy may look quite complicated, the same principles of tax design can be applied. There is a caveat though. Because we have unrestricted preferences, and multiple goods, now the design of optimal redistributive taxes, after correcting for the Pigouvian distortions is more complicated, and without further normative assumptions, we cannot point to lifetime income taxation as the preferred form of redistribution.

### A.13 Heterogeneity in Elasticities (Proof of Proposition 8)

**Proposition.** If a tax schedule is optimal then it needs to satisfy the following relationship:

$$\mathbb{E}\left[\left(\frac{\chi(y) - r(y)}{r(y)}\right)\varepsilon_r^c(y)\right]g(y)y = \int_y^\infty g(\tilde{y})\left(1-\mathbb{E}[\lambda(\tilde{y})]\right)d\tilde{y} + \int_y^\infty \mathbb{E}\left[\left(\frac{\chi(\tilde{y}) - r(\tilde{y})}{r(\tilde{y})}\right)\eta_1(\tilde{y})\right]g(\tilde{y})d\tilde{y}$$
where $\chi(y) \equiv v(y, \theta)/y'(h(y))$

**Proof.** The proof is analogous to the single dimensional heterogeneity case, except that the expectation on the outside cannot be dropped.

A.14 Firms See Additional Signals (Proof of Proposition 9)

**Proof.**

\[
\max_{\hat{R}} E[\lambda(\theta)V(\hat{R}, \theta)] \text{ s.t. } E[v(\theta)(h(\theta)) - \hat{R}(h(\theta))] \geq 0
\]

where $E[v(\theta)|MRS_{c,h}^\theta \leq \hat{R}'(h)] = y'(h)$

\[
\hat{R}(h) = R(y(h))
\]

Considering a small variation on marginal retention rates as a function of earnings, and noticing that this variation translates into an $\frac{dv}{dy}$ variation in the marginal retention as a function of effort:

\[
E \left[ \lambda(\theta) \frac{dV(r, \theta)}{dr_h} \frac{dy}{dy} \right] = -\mu E \left[ \int_0^{\hat{R}} \lambda(\theta) \frac{dr_h(\theta; z)}{dy} \frac{dy}{dy} \right]
\]

\[
-\mu E \left[ \int_0^{\hat{R}} v(\theta) \frac{dh(\theta)}{dr_h(\theta; z)} \frac{dy}{dy} + \frac{dY(\theta)}{dr_h(\theta; z)} \frac{dy}{dy} \right]
\]

\[
-1(y(h(\theta), z(\theta)) \geq y) \frac{dr_h(\theta; z)}{dy}
\]

\[
E \left[ \int_0^{\hat{R}} \lambda(\theta) \frac{dV(r, \theta)}{dr_h} \frac{dy}{dy} \right] = -E \left[ \int_0^{\hat{R}} \left( v(\theta) - r_h(h(\theta), z(\theta)) \right) \left( \frac{dh(\theta)}{dr_h(\theta; z)} \frac{dy}{dy} \right) \frac{dy}{dy} \right]
\]

\[
-1(y(h(\theta), z(\theta)) \geq y) \frac{dr_h(\theta; z)}{dy}
\]

\[
E \left[ \int_0^{\hat{R}} \frac{f(h; z)}{dy} \frac{dy}{dy} \right]
\]

\[
-\mu \int_0^{\hat{R}} \left( v(\theta) - r_h(h(\theta), z(\theta)) \right) \left( \frac{dh(\theta)}{dr_h(\theta; z)} \frac{dy}{dy} \right) \frac{dy}{dy} \right]
\]

\[
-1(y(h(\theta), z(\theta)) \geq y) \frac{dr_h(\theta; z)}{dy}
\]

73
\[ E_z \left[ \int_y^\infty \int_{h(y,z)}^\infty \lambda(\hat{h}, z) f(\hat{h}|z) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{h}d\hat{g} \right] = \]

\[ -E_z \left[ \int_y^\infty \int_{h(y,z)}^\infty (v(\theta) - r_h(h(\theta), z(\theta))) \frac{dh^n(\theta)}{dh_h(h(\theta), z)} f(h(y,z) | z) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{g} \right] \]

\[ +E_z \left[ \int_y^\infty \int_{h(y,z)}^\infty (v(\theta) - r_h(h(\theta), z(\theta))) \frac{dh^n(\theta)}{dh_h(h(\theta), z)} f(h(y,z) | z) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{g} \right] \]

\[ E_z \left[ \int_y^\infty \left( v(\theta) - r_h(h(\theta), z(\theta)) \right) \frac{dh^n(\theta)}{dh_h(h(\theta), z)} f(h(y,z) | z) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{g} \right] = \]

\[ E_z \left[ \int_y^\infty \int_{h(y,z)}^\infty \left( 1 - \lambda(\hat{h}, z) \right) f(\hat{h}|z) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{h}d\hat{g} \right] \]

\[ +E_z \left[ \int_y^\infty \int_{h(y,z)}^\infty \left( v(\hat{h}; z) - r_h(h(\theta), z(\theta)) \right) \frac{dr_{h(\hat{h}, z)}}{dr_y} \, d\hat{h}d\hat{g} \right] \]

\[ E_z \left[ \int_y^\infty \left( \lambda(\hat{g}, z) - r_\theta \right) \frac{dr_{r_\theta(h,g)}}{dr_y} \, d\hat{g} \right] = E_z \left[ \int_y^\infty \int_{\hat{g}}^g \left( 1 - \lambda(\hat{g}, z) \right) p(\hat{g}|z) \frac{dr_{h(\hat{g}, z)}}{dr_y} \, d\hat{g}d\hat{f} \right] \]

\[ +E_z \left[ \int_y^\infty \int_{\hat{g}}^g \left( \lambda(\hat{g}, z) - r_\theta \right) \frac{dr_{r_\theta(h,g)}}{dr_y} \, d\hat{g}d\hat{f} \right] \]

\[ \square \]

74
B Empirical Appendix

B.1 Computing Simulated Marginal Rates and Their Changes

Whenever possible, to match available HRS variables to TAXSIM32 variables, we follow the same treatment of input variables to TAXSIM as outlined in RAND’s 2014 HRS tax calculations (Pantoja et al., 2018). Discrepancies are recorded in footnotes under the corresponding HRS variables.

<table>
<thead>
<tr>
<th>HRS Variable(s) Used</th>
<th>taxsim32 variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>RwIWENDY</td>
</tr>
<tr>
<td>3</td>
<td>RwSTATE</td>
</tr>
<tr>
<td>4</td>
<td>RwMSTAT</td>
</tr>
<tr>
<td>5</td>
<td>RwAGEY_B</td>
</tr>
<tr>
<td>6</td>
<td>SwAGEY_B</td>
</tr>
<tr>
<td>7</td>
<td>depx</td>
</tr>
<tr>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>RwIEARN</td>
</tr>
<tr>
<td>12</td>
<td>SwIEARN</td>
</tr>
<tr>
<td>13</td>
<td>HwIDIVIN</td>
</tr>
<tr>
<td>14</td>
<td>N/A</td>
</tr>
<tr>
<td>15</td>
<td>N/A</td>
</tr>
<tr>
<td>16</td>
<td>N/A</td>
</tr>
</tbody>
</table>

41 Same as RAND: subtract one year from the interview end year to get the calendar year the income is reported for.
42 Same as RAND: married and married with “spouse absent” as defined in the HRS survey are both recorded as jointly married; observations are recorded as single otherwise.
43 For 1992 and 1992, we record dependents as max(children ever, non-resident children). No information about dependents was found in the 1995 FAT File, so we extend values of the previous wave for all interviewees. For 1996, we use the resident children variable instead as the dependents variable is not available. For all other waves, we use the number of dependents variable.
44 Only available starting at wave 3.
| 17 | $^{45}$HwRNTIN, HwIOTH1, HwIOTH2, HwITRSIN, HwIBNDIN, HwIBUSIN, HwICHKIN, HwIDIN, HwIALMNY, HwICDIN, HwICHKIN, H1ISAV1, H2ISAV2, HwIBUSIN, HwILUYR1-HwILUYR3 | otherprop |
| 18 | N/A | nonprop |
| 19 | RwIPENA, SwIPENA | pensions |
| 20 | $^{46}$RwISSDI, RwISSI, SwISSDI, SwISSI, Medicare part b) coverage variable from FAT Files | gssi |
| 21 | RwIUNEM, SwIUNEM | ui |
| 22 | HwISSI, HwIFOOD, HwIWELF, RwIWCMP, RwIVET | transfers |
| 23 | $^{47}$Dollar amount of rent paid variable from FAT Files | rent paid |
| 24 | $^{48}$Dollar amount of real estate tax paid variable from FAT Files | proptax |
| 25 | $^{49}$RwOOPMD, SwOOPMD | otheritem |
| 26 | N/A | childcare |
| 27 | $^{50}$HwAMORT, Dollar amount of donations variable from FAT Files, RwOOPMD, SwOOPMD | mortgage |
| 28 | N/A | scorp |
| 29 | N/A | pbusinc |

$^{45}$Same as RAND. H1CHKIN and H2CHKIN are not available, so we use H1ISAV1 and H2ISAV2 instead. Unlike HwCHKIN variables, which document both savings and checking interest income, H1ISAV1 and H2ISAV2 only record savings interest.

$^{46}$Same as RAND except that Medicare part d) premiums are not added back (HRS only started recording it in 2006) and additional premiums deducted for higher income individuals, which started in 2007, are not added back. Medicare part b) coverage variable is not available for 1994.

$^{47}$Values exceeding 999996 are reset to zero. These represent “don’t know”/”refused to answer” responses to the survey question instead of an actual dollar amount.

$^{48}$Values exceeding 999994 are reset to zero. These represent “don’t know”/”refused to answer” responses to the survey question instead of an actual dollar amount.

$^{49}$The medical expenditure variable is not available for wave 1. Except for wave 2 variable of medical expenditure, which records only one year of medical expense, we divide the observation by 2 to reflect that the medical expenditure is over a two-year period. We follow all of RAND’s procedure; we interpret worker AGI as the sum of the AGIs of the primary respondent and spouse.

$^{50}$Like RAND, we use the 30-year fixed rate mortgages (FRMs) multiplied by HwAMORT to calculate mortgage interest. However, we use FRMs published by the St. Louis Fed instead and average weekly values for the whole year to find the interest rate for the corresponding year. For the dollar amount of donations variable, values exceeding 999995 are reset to zero. These represent “don’t know”/”refused to answer” responses to the survey question instead of an actual dollar amount. Again, in calculating medical expenditure that is not a preference for AMT, the medical expenditure variable is not available for wave 1. Except for wave 2 variable of medical expenditure, which records only one year of medical expense, we divide the observation by 2 to reflect that the medical expenditure is over a two-year period. We follow all of RAND’s procedure, which now uses slightly different equations than the ones in computing for itemized deductions that are a preference for AMT; we interpret worker AGI as the sum of the AGIs of the primary respondent and spouse.
B.2 State Variation in Taxes

Figures 2 to 7 illustrate in which states have had the largest and most frequent changes in real marginal rates, for both single and joint tax returns. To observe the changes in state income tax codes, we obtain income percentile cutoffs by converting all nominal incomes reported in the HRS dataset to 2021 dollars, using the PCE index. Then, we construct a pseudo dataset with these cutoffs in all 50 states plus Washington DC from 1992 to 2018. We use the NBER tax simulator (taxsim32) to simulate marginal income tax rates for these constructed individuals. Finally, we increment year by one, inflate accordingly using the PCE index, obtain a new set of marginal income tax rates, and take the difference between the two rates to find the policy change at each income level.

B.3 Mental Status Scores

The mental status summary sums the scores for serial 7’s (RwSER7, 0-5), backwards counting from 20 (RwBWC2, 0-2), and object (RwCACT, RwSCIS; 0-2 total), date (RwDY, RwMO, RwYR, RwDW; 0-4 total), and President/Vice-President (RwPRES, RwVP; 0-2 total) naming tasks. The resulting range is 0-15. Since these items were not included in Waves 1 and 2H, there is no mental summary score for these waves, and the Wave 2A summary is called R2AMSTOT to indicate that it is limited to the AHEAD cohort in Wave 2.

Those questions are presented in the table below.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Content</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RwCACT</td>
<td>&quot;What do you call the kind of prickly plant that grows in the desert?&quot;</td>
<td>0-1</td>
</tr>
<tr>
<td>RwSCIS</td>
<td>&quot;What do you usually use to cut paper?&quot;</td>
<td>0-1</td>
</tr>
<tr>
<td>RwSCIS and RwVP</td>
<td>whether the Respondent was able to correctly name the current president and vice-president of the United States, respectively.</td>
<td>0-2</td>
</tr>
<tr>
<td>RwSER7</td>
<td>Number of correct subtractions in the serial 7s test. This test asks the individual to subtract 7 from the prior number, beginning with 100 for five trials. Correct subtractions are based on the prior number given, so that even if one subtraction is incorrect subsequent trials are evaluated on the given (perhaps wrong) answer.</td>
<td>0-5</td>
</tr>
<tr>
<td>RwBWC20 and RwBWC86</td>
<td>whether the Respondent was able to successfully count backwards for 10 continuous numbers from 20 and 86, respectively. Two points are given if successful on the first try, one if successful on the second, and zero if not successful on either try.</td>
<td>0-2</td>
</tr>
<tr>
<td>RwDY, RwMO, RwYR, and RwDW</td>
<td>whether the Respondent was able to report today’s date correctly, including the day of month, month, year, and day of week, respectively.</td>
<td>0-4</td>
</tr>
</tbody>
</table>
### B.4 Tables

Table 3: Elasticities of wages

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_w$</td>
<td>-0.13</td>
<td>-0.061</td>
<td>-0.066</td>
<td>-0.092</td>
<td>-0.16</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.099)</td>
<td>(0.099)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>39179</td>
<td>39179</td>
<td>39179</td>
<td>39179</td>
<td>39179</td>
<td>39179</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Elasticities $\epsilon_w$ are computed from linear regressions of changes in log hourly wages over four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. Column (4) includes log hourly wages. Column (5) includes a 5 piece cubic spline of log hourly wages. Column (6) includes a 10-piece linear spline of log hourly wages.
### Table 4: Elasticities of wages for job switchers

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon^w$</td>
<td>-0.31</td>
<td>-0.19</td>
<td>-0.19</td>
<td>-0.28</td>
<td>-0.32</td>
<td>-0.34</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>13958</td>
<td>13958</td>
<td>13958</td>
<td>13958</td>
<td>13958</td>
<td>13958</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Elasticities $\epsilon^w$ are computed from linear regressions of changes in log hourly wages over four years on changes in log marginal retention rates over two years (evaluated at the base year income data). The sample is restricted to those who switch jobs at least once between the baseline year and four years later. Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. Column (4) includes log hourly wages. Column (5) includes a 5 piece cubic spline of log hourly wages. Column (6) includes a 10-piece linear spline of log hourly wages.
### Table 5: Participation semi-elasticities

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>η^p</td>
<td>-0.039</td>
<td>0.063</td>
<td>0.044</td>
<td>0.037</td>
<td>0.013</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.058)</td>
<td>(0.056)</td>
<td>(0.062)</td>
<td>(0.062)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>72526</td>
<td>72526</td>
<td>72526</td>
<td>61526</td>
<td>61526</td>
<td>61526</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Semi-elasticities η^p are computed from linear regressions of changes of participation over four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. Column (4) includes log hourly wages. Column (5) includes a 5 piece cubic spline of log hourly wages. Column (6) includes a 10-piece linear spline of log hourly wages.
Table 6: "Rat race" externality estimates (4y/4y) − (1 − χ)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratio</td>
<td>0.034</td>
<td>-0.014</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(-0.0095, 2.42)</td>
<td>(-2.36, 0.027)</td>
<td>(-2.69, 0.024)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>observations</td>
<td>85706</td>
<td>85706</td>
<td>85706</td>
</tr>
</tbody>
</table>

Table 7: "Rat race" externality estimates (4y/4y) − (1 − χ) (cont.)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratio</td>
<td>-0.025</td>
<td>-0.12</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>(-47.4, 0.0071)</td>
<td>(-9.03, -0.044)</td>
<td>(-31.3, -0.076)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>85706</td>
<td>85706</td>
<td>85706</td>
</tr>
</tbody>
</table>

Notes. Bootstrapped bias-corrected confidence intervals in parentheses (with 2000 bootstrap replications). Estimates for externality (1 − χ, or one minus the labor wedge) are obtained from dividing the elasticity of wages by the participation semi-elasticity multiplied by one hundred. Elasticities of wages are computed from linear regressions of changes in log hourly wages over the next four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Participation semi elasticities are computed from regressing changes in participation over the next four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. In the bottom table, column (1) includes log hourly wages. Column (2) includes a 5 piece cubic spline of log hourly wages. Column (3) includes a 10-piece linear spline of log hourly wages.
Table 8: 'Rat race' externality estimates (4y/2y) \(-(1 - \chi)\)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratio</td>
<td>-0.017</td>
<td>-0.0045</td>
<td>-0.0049</td>
</tr>
<tr>
<td></td>
<td>(-0.22, 0.019)</td>
<td>(-0.029, 0.013)</td>
<td>(-0.030, 0.013)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>observations</td>
<td>85706</td>
<td>85706</td>
<td>85706</td>
</tr>
</tbody>
</table>

Table 9: "Rat race" externality estimates (4y/2y) \(-(1 - \chi)\) (cont.)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratio</td>
<td>-0.0072</td>
<td>-0.015</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(-0.043, 0.0086)</td>
<td>(-0.12, 0.0045)</td>
<td>(-0.13, 0.0053)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>85706</td>
<td>85706</td>
<td>85706</td>
</tr>
</tbody>
</table>

Notes. Bootstrapped bias-corrected confidence intervals in parentheses (with 2000 bootstrap replications). Estimates for externality \((1 - \chi\), or one minus the labor wedge) are obtained from dividing the elasticity of wages by the participation semi-elasticity multiplied by one hundred. Elasticities of wages are computed from linear regressions of changes in log hourly wages over the next four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Participation semi elasticities are computed from regressing changes in participation over the next two years on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. In the bottom table, column (1) includes log hourly wages. Column (2) includes a 5 piece cubic spline of log hourly wages. Column (3) includes a 10-piece linear spline of log hourly wages.
Table 10: Mental status scores

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta^m$</td>
<td>0.14</td>
<td>-0.82</td>
<td>-1.08</td>
<td>-1.20</td>
<td>-1.30</td>
<td>-1.31</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.51)</td>
<td>(0.51)</td>
<td>(0.54)</td>
<td>(0.54)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>year f.e.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>marital status</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>hourly wages</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>linear</td>
<td>c. spline</td>
<td>l. spline</td>
</tr>
<tr>
<td>observations</td>
<td>16027</td>
<td>13187</td>
<td>13187</td>
<td>11030</td>
<td>11030</td>
<td>11030</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. Semi-elasticities $\eta^m$ are computed from linear regressions of mental status scores two years in the past on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample for those who are working in the baseline year and four years ahead. Each column includes different sets of controls: year fixed effects, marital status dummies, and hourly wages. Column (4) includes log hourly wages. Column (5) includes a 5 piece cubic spline of log hourly wages. Column (6) includes a 10-piece linear spline of log hourly wages.
Table 11: Elasticities of wages

<table>
<thead>
<tr>
<th></th>
<th>-8y</th>
<th>-6y</th>
<th>-4y</th>
<th>-2y</th>
<th>+2y</th>
<th>+4y</th>
<th>+6y</th>
<th>+8y</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_w$</td>
<td>0.12</td>
<td>-0.0089</td>
<td>-0.014</td>
<td>0.055</td>
<td>0.036</td>
<td>-0.16</td>
<td>-0.27</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.075)</td>
<td>(0.062)</td>
<td>(0.047)</td>
<td>(0.076)</td>
<td>(0.099)</td>
<td>(0.12)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>N</td>
<td>17898</td>
<td>25425</td>
<td>34685</td>
<td>45922</td>
<td>53931</td>
<td>39179</td>
<td>29253</td>
<td>21119</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. Elasticities $\epsilon_w$ are computed from linear regressions of changes in log hourly wages over k-years on changes in log marginal retention rates over two years (evaluated at the base year income data). All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Table 12: Elasticities of wages for job switchers

<table>
<thead>
<tr>
<th>y</th>
<th>( \epsilon^w )</th>
<th>(SE)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>-8y</td>
<td>0.36</td>
<td>(0.16)</td>
<td>10040</td>
</tr>
<tr>
<td>-6y</td>
<td>0.19</td>
<td>(0.13)</td>
<td>11950</td>
</tr>
<tr>
<td>-4y</td>
<td>0.023</td>
<td>(0.12)</td>
<td>12375</td>
</tr>
<tr>
<td>-2y</td>
<td>0.30</td>
<td>(0.12)</td>
<td>9430</td>
</tr>
<tr>
<td>+2y</td>
<td>-0.34</td>
<td>(0.19)</td>
<td>10996</td>
</tr>
<tr>
<td>+4y</td>
<td>-0.43</td>
<td>(0.18)</td>
<td>13958</td>
</tr>
<tr>
<td>+6y</td>
<td>-0.43</td>
<td>(0.18)</td>
<td>13703</td>
</tr>
<tr>
<td>+8y</td>
<td>-0.43</td>
<td>(0.21)</td>
<td>11737</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. Elasticities \( \epsilon^w \) are computed from linear regressions of changes of log hourly wages over k-years on changes in log marginal retention rates over two years (evaluated at the base year income data). For each column, the sample is restricted to those who switch jobs at least once between the baseline year and k years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
<table>
<thead>
<tr>
<th></th>
<th>-8y</th>
<th>-6y</th>
<th>-4y</th>
<th>-2y</th>
<th>+2y</th>
<th>+4y</th>
<th>+6y</th>
<th>+8y</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta^p$</td>
<td>-0.039</td>
<td>-0.021</td>
<td>-0.048</td>
<td>-0.039</td>
<td>0.10</td>
<td>0.010</td>
<td>0.032</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.047)</td>
<td>(0.062)</td>
<td>(0.068)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>N</td>
<td>20522</td>
<td>29053</td>
<td>39482</td>
<td>51895</td>
<td>72301</td>
<td>61526</td>
<td>53964</td>
<td>46435</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Semi-elasticities $\eta^p$ are computed from linear regressions of changes of hours wages over k-years years on changes in log marginal retention rates over two years (evaluated at the base year income data). All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Table 14: Mental status scores

<table>
<thead>
<tr>
<th></th>
<th>-8y</th>
<th>-6y</th>
<th>-4y</th>
<th>-2y</th>
<th>+2y</th>
<th>+4y</th>
<th>+6y</th>
<th>+8y</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_m$</td>
<td>0.32</td>
<td>-0.034</td>
<td>-0.56</td>
<td>0.65</td>
<td>-0.77</td>
<td>-1.31</td>
<td>-1.19</td>
<td>-0.61</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td>(0.61)</td>
<td>(0.45)</td>
<td>(0.38)</td>
<td>(0.45)</td>
<td>(0.54)</td>
<td>(0.57)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>N</td>
<td>4362</td>
<td>5784</td>
<td>9566</td>
<td>13055</td>
<td>14705</td>
<td>11030</td>
<td>8264</td>
<td>5005</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. For columns (5) to (8), semi-elasticities $\eta^p$ are computed from linear regressions of memory scores two years in the past over years on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample for those who are working in the baseline year and k-years ahead. For columns (1) to (4), semi-elasticities $\eta^p$ are computed from linear regressions of memory scores ($2+|k|$) years in the past over years on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample for those who are working in the baseline year and $|k|$-years before. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
### Table 15: Income percentiles

<table>
<thead>
<tr>
<th></th>
<th>wg, 4y</th>
<th>wg (j.s.) 4y</th>
<th>partic. 2y</th>
<th>mental st. 2y</th>
</tr>
</thead>
<tbody>
<tr>
<td>bottom third</td>
<td>-0.14</td>
<td>-0.40</td>
<td>0.087</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.25)</td>
<td>(0.082)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>middle third</td>
<td>-0.064</td>
<td>-0.37</td>
<td>0.19</td>
<td>-1.09</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.26)</td>
<td>(0.078)</td>
<td>(0.82)</td>
</tr>
<tr>
<td>upper third</td>
<td>-0.29</td>
<td>-0.22</td>
<td>0.035</td>
<td>-1.18</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.44)</td>
<td>(0.073)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Observations</td>
<td>39179</td>
<td>13958</td>
<td>72301</td>
<td>14705</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Elasticities and semi-elasticities are computed from linear regressions of different dependent variables on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column corresponds to a different dependent variable: the first column reports results where the dependent variable are changes in log hourly wages over the next four years; the second column additionally restricts the sample to respondents who switched jobs over the same period. The third column reports results for changes in labor force participation over two years. The fourth column has a dependent variable mental status scores two years in the past, and restricts the sample to those who are working in the baseline year and two years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Table 16: Occupations

<table>
<thead>
<tr>
<th></th>
<th>wg.</th>
<th>wg (j.s.)</th>
<th>partic.</th>
<th>mental st.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managerial, sales, clerical</td>
<td>0.14</td>
<td>-0.10</td>
<td>0.10</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.29)</td>
<td>(0.087)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>Professional</td>
<td>-0.43</td>
<td>-0.45</td>
<td>0.11</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.54)</td>
<td>(0.12)</td>
<td>(0.83)</td>
</tr>
<tr>
<td>Other services</td>
<td>0.21</td>
<td>0.17</td>
<td>-0.18</td>
<td>-4.69</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.42)</td>
<td>(0.14)</td>
<td>(1.95)</td>
</tr>
<tr>
<td>Farming, forestry, mechanics,</td>
<td>-0.038</td>
<td>-1.22</td>
<td>0.050</td>
<td>-2.60</td>
</tr>
<tr>
<td>construction</td>
<td>(0.43)</td>
<td>(0.79)</td>
<td>(0.16)</td>
<td>(1.70)</td>
</tr>
<tr>
<td>Operators</td>
<td>-0.49</td>
<td>-0.83</td>
<td>0.090</td>
<td>-2.59</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.38)</td>
<td>(0.13)</td>
<td>(1.38)</td>
</tr>
<tr>
<td>Observations</td>
<td>39179</td>
<td>13958</td>
<td>72301</td>
<td>14705</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. Elasticities and semi-elasticities are computed from linear regressions of different dependent variables on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column corresponds to a different dependent variable: the first column reports results where the dependent variable are changes in log hourly wages over the next four years; the second column additionally restricts the sample to respondents who switched jobs over the same period. The third column reports results for changes in labor force participation over two years. The fourth column has a dependent variable mental status scores two years in the past, and restricts the sample to those who are working in the baseline year and two years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Table 17: Education levels

<table>
<thead>
<tr>
<th></th>
<th>wg.</th>
<th>wg (j.s.)</th>
<th>partic.</th>
<th>mental st.</th>
</tr>
</thead>
<tbody>
<tr>
<td>high school or less</td>
<td>-0.27</td>
<td>-0.28</td>
<td>0.018</td>
<td>-2.38</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.23)</td>
<td>(0.067)</td>
<td>(0.76)</td>
</tr>
<tr>
<td>more than high school</td>
<td>-0.054</td>
<td>-0.39</td>
<td>0.18</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.26)</td>
<td>(0.062)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>Observations</td>
<td>39179</td>
<td>13958</td>
<td>72301</td>
<td>14705</td>
</tr>
</tbody>
</table>

Notes. Robust standard errors in parentheses. Elasticities and semi-elasticities are computed from linear regressions of different dependent variables on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column corresponds to a different dependent variable: the first column reports results where the dependent variable are changes in log hourly wages over the next four years; the second column additionally restricts the sample to respondents who switched jobs over the same period. The third column reports results for changes in labor force participation over two years. The fourth column has a dependent variable mental status scores two years in the past, and restricts the sample to those who are working in the baseline year and two years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
### Table 18: Income percentiles and education

<table>
<thead>
<tr>
<th></th>
<th>wg.</th>
<th>wg (j.s.)</th>
<th>partic.</th>
<th>mental st.</th>
</tr>
</thead>
<tbody>
<tr>
<td>bottom third, highschool or less</td>
<td>-0.28</td>
<td>-0.24</td>
<td>-0.020</td>
<td>-1.78</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.27)</td>
<td>(0.099)</td>
<td>(1.14)</td>
</tr>
<tr>
<td>middle third, highschool or less</td>
<td>-0.21</td>
<td>-0.33</td>
<td>0.10</td>
<td>-3.21</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.38)</td>
<td>(0.10)</td>
<td>(1.22)</td>
</tr>
<tr>
<td>upper third, highschool or less</td>
<td>-0.37</td>
<td>-0.30</td>
<td>-0.042</td>
<td>-2.50</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.88)</td>
<td>(0.16)</td>
<td>(1.63)</td>
</tr>
<tr>
<td>bottom third, more than highschool</td>
<td>0.15</td>
<td>-0.74</td>
<td>0.31</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.50)</td>
<td>(0.14)</td>
<td>(1.13)</td>
</tr>
<tr>
<td>middle third, more than highschool</td>
<td>0.093</td>
<td>-0.40</td>
<td>0.30</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.33)</td>
<td>(0.11)</td>
<td>(1.07)</td>
</tr>
<tr>
<td>upper third, more than highschool</td>
<td>-0.26</td>
<td>-0.20</td>
<td>0.061</td>
<td>-0.77</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.50)</td>
<td>(0.081)</td>
<td>(0.64)</td>
</tr>
<tr>
<td>Observations</td>
<td>39179</td>
<td>13958</td>
<td>72301</td>
<td>14705</td>
</tr>
</tbody>
</table>

**Notes.** Robust standard errors in parentheses. Elasticities and semi-elasticities are computed from linear regressions of different dependent variables on changes in log marginal retention rates over two years (evaluated at the base year income data). Each column corresponds to a different dependent variable: the first column reports results where the dependent variable are changes in log hourly wages over the next four years; the second column additionally restricts the sample to respondents who switched jobs over the same period. The third column reports results for changes in labor force participation over two years. The fourth column has a dependent variable mental status scores two years in the past, and restricts the sample to those who are working in the baseline year and two years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
B.5 Figures

Figure 2: State variation in real marginal tax rates

Approximate % of Joint Tax Filings Experiencing an Absolute Change in State Marginal Rate >= 2%

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of jointly-filing taxpayers experiencing an absolute marginal tax change larger than 2%.
Figure 3: State variation in real marginal tax rates

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of jointly-filing taxpayers experiencing a marginal tax change larger than 2%.
Figure 4: State variation in real marginal tax rates

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of taxpayers experiencing a marginal tax decrease larger than 2%.
Figure 5: State variation in real marginal tax rates

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of single taxpayers experiencing an absolute marginal tax change larger than 2%.
Figure 6: State variation in real marginal tax rates

Approximate % of Single Tax Filings Experiencing a Change in State Marginal Rate ≥ 2%

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of single taxpayers experiencing a simulated marginal tax increase larger than 2%.
Figure 7: State variation in real marginal tax rates

Notes. Percentages were calculated using the NBER tax simulator and the publicly available HRS RAND longitudinal file, imputing each state distribution of income from the national distribution of income. Shades are proportional to the percentage of single taxpayers experiencing a simulated marginal tax decrease larger than 2%.
Figure 8: Elasticity of wages over different horizons

Notes. 95% confidence intervals. Elasticities are computed from linear regressions of changes in log hourly wages over k-years on changes in log marginal retention rates over two years (evaluated at the base year income data). All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Figure 9: Elasticity of wages for job switchers over different horizons

Notes. 95% confidence intervals. Elasticities are computed from linear regressions of changes of log hourly wages over k-years on changes in log marginal retention rates over two years (evaluated at the base year income data). For each column, the sample is restricted to those who switch jobs at least once between the baseline year and k years ahead. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Figure 10: Semi-elasticities of participation over different horizons

Notes. 95% confidence intervals. Semi-elasticities are computed from linear regressions of changes of hours wages over k-years years on changes in log marginal retention rates over two years (evaluated at the base year income data). All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Notes. 95% confidence intervals. For the periods (k) between 2 and 8, semi-elasticities are computed from linear regressions of memory scores two years in the past over years on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample for those who are working in the baseline year and k-years ahead. Similarly, for the periods (k) between -6 to -2, semi-elasticities are computed from linear regressions of memory scores (2+|k|) years in the past over years on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample for those who are working in the baseline year and |k|-years before. All specifications include year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Figure 12: Elasticity of wages over different hourly wages percentiles

Notes. 95% bootstrap confidence intervals. Elasticities are computed from local linear regressions of changes in log hourly wages over four years on changes in log marginal retention rates over two years (evaluated at the base year income data), including year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Figure 13: Semi-elasticities of participation over different hourly wages percentiles

Notes. 95% bootstrap confidence intervals. Semi-elasticities are computed from linear regressions of changes of hours wages over two years on changes in log marginal retention rates over two years (evaluated at the base year income data), including year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Notes. 95% bootstrap confidence intervals. Estimates are computed from the ratio of elasticities of wages and semi-elasticities of participation. Elasticities of participation are computed from local linear regressions of changes in log hourly wages over four years on changes in log marginal retention rates over two years (evaluated at the base year income data). Semi-elasticities are computed from linear regressions of changes of hours wages over two years on changes in log marginal retention rates over two years (also evaluated at the base year income data). Each regression includes year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.
Figure 15: Semi-elasticities of mental status scores over different hourly wages percentiles

Notes. 95% bootstrap confidence intervals. Semi-elasticities are computed from linear regressions of memory scores 2 years in the past over years on changes in log marginal retention rates over two years (evaluated at the base year income data), restricting the sample to those who are working in the baseline year and 2 years ahead. Each regression includes year fixed effects, marital status dummies, and a 10-piece linear spline of log hourly wages.