The Surprising Power of Age-Dependent Taxes

Matthew Weinzierl*

April 28, 2008

Abstract

I study age-dependent labor income taxes as a partial reform in a Mirrleesian dynamic taxation environment. Despite its simplicity, age dependence generates a large welfare gain both in absolute size and relative to more complex, fully optimal policy. Using modern dynamic optimal tax methods, I provide a comprehensive theoretical and quantitative examination of age dependence and compare it to two alternative policies: an age-independent policy and a dynamic optimal policy.

First, I derive theoretical results for this partial reform that connect to and extend classic results and intuition from static and dynamic Mirrleesian tax analysis. I analytically characterize how age dependence improves on age-independent policy but falls short of the dynamic optimum on both the intratemporal consumption-leisure and intertemporal savings margins.

Second, using detailed individual wage data from the U.S. PSID, I calibrate and simulate the policy models, generating robust implications: age dependence (1) lowers marginal taxes on average and especially on high-income young workers, and (2) lowers average taxes on all young workers relative to older workers when private saving and borrowing are restricted. These results capture key features of the dynamic optimal policy.

Finally, I quantify the welfare gain from this partial reform. Age dependence yields a large welfare gain equal to between one and three percent of aggregate annual consumption, and it captures a substantial portion of the gain from reform to the dynamic optimal policy. A detailed decomposition reveals that improvements in efficiency and equity each account for approximately half of this gain. To further motivate its practicality, I add the constraint that age dependence must be Pareto improving. This more constrained reform generates nearly as large a welfare gain as does the Utilitarian-optimal policy.

Introduction

Individuals' wages change over their lifecycle. Starting with Golosov, Kocherlakota, and Tsyvinski (2003), this fact has inspired a recent surge of research on optimal taxation in dynamic economies that builds on the classic, static Mirrlees (1971) framework. Unifying the important insights of this research is the principle that optimal taxation in dynamic economies depends, other than in special cases,[1] on a taxpayer's history.

---

[1] Saez (2001) is the leading recent study of the classic Mirrlees framework. In the dynamic Mirrlees framework, Albanesi and Sleet (2006) shows that history dependence can be replaced with dependence on current wealth if shocks to skills are i.i.d.. Golosov and Tsyvinski (2006) shows that asset-testing can replace history dependence if shocks are permanent: i.e., an absorbing state. For general shock processes, Kocherlakota (2005) shows that history dependence is required.

While this principle points the way toward potentially powerful reforms, its near-term impact may be limited by the complexity of the reforms to existing policies that it implies.

This paper studies a simple reform to tax policy that responds to changes in wages over the lifecycle: age-dependent labor income taxes. Though not often recognized as such, age dependence is a component of Mirrleesian dynamic optimal taxation and thus can be analyzed as a *partial* reform on the path toward optimal policy. Taking that approach, this paper makes three contributions.

First, its status as a partial reform allows me to provide a comprehensive theoretical characterization of age dependence using the tools of modern dynamic Mirrleesian optimal tax research. I derive theoretical results for age-dependent policy that connect to and extend results and intuition from two alternative Mirrleesian policies, an age-independent optimal policy and a dynamic optimal policy. In a baseline model with deterministic wage paths and no private saving or borrowing, I show that age-dependent policy avoids discouraging the labor supply of the top earner at each age while age-independent policy cannot, extending a classic result from static Mirrlees research. Then, I analytically characterize the intertemporal consumption margin, where age-dependent policy satisfies a condition that improves on age-independent policy but falls short of the full optimum. I characterize how age dependence affects these margins in economic environments with stochastic wages and private saving and borrowing, as well.

Second, I provide a detailed quantitative study of age dependence using individual wage data from the U.S. Panel Study of Income Dynamics to calibrate and simulate policy. This is in contrast to most of the dynamic optimal taxation literature, in which illustrative numerical simulations are the rule, and it resembles the realistic calibrations by Saez (2001) of the static Mirrlees model and Golosov and Tsyvinski (2006) of optimal disability insurance. The numerical simulations generate specific implications for policy design. Two results largely robust across environments are that age dependence: (1) lowers marginal taxes on average and especially on high-income young workers, and (2) lowers average tax rates for young workers relative to older workers when private saving and borrowing are restricted. These results capture key features of the optimal dynamic policy.

Finally, the numerical simulations allow me to quantify and decompose the welfare gain from this partial reform. Age dependence yields a large welfare gain equivalent to between 1 and 3 percent of aggregate annual consumption, and this gain represents a substantial share of the potential gain from full reform to the dynamic optimal policy. A detailed decomposition reveals three main components of the gain. Efficiency and equity gains each account for a bit less than half of the total welfare gain, while intertemporal consumption smoothing accounts for approximately ten percent in the baseline model. Age dependence allows tax policy to be tailored to changes in the distribution of wages at each age and to transfer resources between age groups, avoiding inefficient distortions to labor effort and enabling more redistribution.[2]

Age dependence is a compelling example of the potential value of partial reforms. The notion of partial reform was formalized by Guesnerie (1977) in response to concerns raised by Feldstein (1976) about the practicality of static optimal tax design. Partial reforms are most valuable when they yield substantial welfare gains while raising fewer real-world concerns than fully optimal policy. A particularly strong case can be made for partial reforms that yield Pareto improvements, a characteristic stressed by Guesnerie. Thus, to further motivate the practicality of this reform, I consider a model in which I add the constraint that age dependence must be ex post Pareto improving. This constrained reform generates nearly as large

---

[2]A related policy is lifetime income taxation, where an individual pays taxes or receives refunds in each year based on their (currently) expected lifetime income. The age-dependent policy studied here subsumes that policy and adds the important ability to tailor marginal taxes by age. Moreover, lifetime income taxation also uses age as an argument in the tax function, so it provides little or no advantage in simplicity relative to fully age-dependent taxation.

a welfare gain as does the Utilitarian-optimal policy, underscoring the potential power of age dependence.

Though this paper is the first to give the partial reform of age dependence a comprehensive treatment in a modern dynamic Mirrleesian tax model,[3] the idea of having taxes depend on age has been around for some time and was even mentioned in passing by Mirrlees (1971). Despite this long history, age dependence was first rigorously analyzed in the important work of Kremer (2002), who demonstrated that marginal income taxes not conditioned on age are unlikely to be optimal and suggested that they should be lower for young workers, consistent with the results below. Kremer's analysis did not provide a full characterization or simulation of age dependent taxes and was limited to a static setting, limitations his paper acknowledges. Following on that work, Blomquist and Micheletto (2003) and Judd and Su (2006) provide results on age-dependent taxation in illustrative dynamic economies using theoretical and numerical approaches, respectively. The need for a more comprehensive analysis is underscored by the suggestion in the upcoming Mirrlees Review that age dependence is one of the most promising areas for the near-term reform of developed-country tax policy (see Banks, Diamond, and Mirrlees, 2007).[4]

The literature on age dependence may be relatively sparse because it has been assumed that age is merely another "tag", following Akerlof (1978), and that standard tagging analysis teaches us all we need to know about age dependence. That assumption is mistaken. A standard tag, such as gender, provides information on an individual's expected place in the distribution of lifetime income, but age reveals no information by itself. The difference is that a tag divides the population into mutually exclusive groups, while the entire population moves through all age groups. In other words, each age group has the same distribution of lifetime incomes. To provide useful information to the tax authority, age must be combined with data on an individual's current income and how incomes at each age relate to lifetime income. Therefore, age dependence cannot be fully understood with intuition from conventional tagging analysis.

Consistent with the small academic literature on age dependence, current policy in developed countries includes only uncoordinated and often unintentional age dependence.[5] In the United States, for example, the main dependencies on age work at cross purposes. Social Security and Medicare payroll taxes are levied on all individuals but, given the pay-as-you-go structure of these programs, these taxes place a larger effective burden on the young than on the old (see Feldstein and Samwick, 1992). On the other hand, the existing disability system effectively places a larger effective burden on the old than on the young, since the latter would receive a longer string of benefits if disabled. Some transfer programs, such as the child tax credit or Earned-Income Tax Credit, are more likely to benefit the young than the old, but they also mean higher marginal rates on the young who earn enough to be in the "phase-out" region of these benefits. Finally, the deductibility of mortgage interest and charitable donations are more likely to benefit the middle-aged, for whom renting is less common and incomes are higher, than the young or old. Together, these largely unintentional age dependencies are unlikely to mimic, and may often work in the oppositive direction of, the optimal age-dependent policy as studied in this paper.

Therefore, the analysis of this paper contributes to several literatures. Most directly, it provides a new, empirically-driven application of the tools of the dynamic Mirrleesian optimal tax literature surveyed in

---

[3] Age dependence has been studied in the Ramsey tax framework. For example, Erosa and Gervais (2002) study the effect of private asset accumulation and changes in the elasticity of labor supply with age on optimal linear taxes. As a representative agent framework, the conventional Ramsey approach neglects the redistributional role of taxation. I have simulated a Ramsey version of age-dependent taxation (i.e., age-specific linear taxes and lump-sum taxes or grants) in the heterogeneous-agent model economy discussed below, and the welfare gains from age-dependence are approximately one-tenth their size in the Mirrlees approach. The Ramsey approach cannot tailor marginal taxes to variation in the distribution of wages with age.

[4] The Mirrlees Review is the modern counterpart to the Meade Report of 1978, the influential review of taxation.

[5] One exception is Singapore's Provident Fund, a national retirement savings program that tailors personal contribution rates to age.

Golosov, Tsyvinski, and Werning (2006). It also extends the literature on partial reform, coinciding with recent work of Farhi and Werning (2007) on a different partial reform based on dynamic optimal policy. The paper's consideration of both deterministic and stochastic wages relates to the important ongoing debate over the determinants of lifecycle wage inequality and the appropriate policy responses to it: e.g., Storesletten, Telmer, and Yaron (2001) and Keane and Wolpin (1997). Finally, it adds to a recent literature that empirically estimates the value of conditioning taxes on personal characteristics, such as Alesina and Ichino (2007).

The paper proceeds as follows. Section 1 describes the social planner's problem in three policy scenarios for a baseline economy, i.e., with deterministic wage paths and no private saving or borrowing. For each policy, I analytically characterize the intratemporal and intertemporal distortions on private behavior and numerically simulate the structure of taxes. I also quantify the welfare implications of reform from the static optimal policy to age-dependent policy and the dynamic optimal policy. Section 1 serves as the reference point for the following three sections, in which I vary the assumptions about the economy to test the robustness of the baseline results. Section 2 allows individuals to save and borrow, Section 3 incorporates stochastic wage paths, and Section 4 combines these two variations. Throughout these extensions, the results from the baseline economy are largely robust. Section 5 discusses a range of specific topics that fall outside the main analysis of the paper, and Section 6 concludes.

1   Baseline economy

In this section I analyze age-dependent labor income taxes for a baseline economy characterized by two simplifying assumptions. First, individuals cannot transfer resources across periods: that is, they can neither save nor borrow. Second, each individual's lifetime wage path is deterministic, so that each individual knows in advance the exact path of wages it will have over its lifetime (i.e., there are no stochastic shocks to wages). These simplifying assumptions allow for cleaner analytical results, but I generalize the model to relax them in later sections of the paper. One assumption made throughout this paper, and throughout the dynamic optimal tax literature in general, is that wage paths are exogenous to individuals; I discuss the potential implications of this assumption in Section 5. I start by setting up the economy and then specifying the social planner's problem in three policy scenarios.

1.1   Setup

All individuals live and work for T periods,[6] indexed by $t = \{1, 2, ..., T\}$, and are members of the same generation.[7] Individuals are heterogeneous in their ability to earn income over their lifetimes. This ability comes in I types, indexed by $i = \{1, 2, ..., I\}$, with probabilities π^i so that $\sum_{i=1}^{I} \pi^i = 1$. At each age t, an individual of type i can earn a wage w_t^i for each unit of its labor effort, and each individual knows its full lifetime path of wages $\{w_t^i\}_{t=1}^{T}$ at time $t = 1$. Wages are not publicly observable. I refer to the present value of wages for type i as type i's *lifetime income-earning ability*, defined as $\sum_{t=1}^{T} \frac{w_t^i}{R^{t-1}}$ where R is the exogenous gross rate of return. Types are sorted so that $i = I$ refers to the type with the highest lifetime income-earning ability.

---

[6] I assume an exogenous date of entry into the labor market. This matches the quantitative analyses below, which consider age-dependent taxes designed to have minimal effects on incentives to obtain higher education or training. Age dependence could generate even larger welfare gains if it were designed to operate effectively on these incentives.

[7] In the Appendix, I discuss how the analysis generalizes largely unchanged to a setting with overlapping generations.

Production is structured as follows. Labor income y is the product of the wage and labor effort l, so $y = wl$. There is no capital in the economy.

All individuals have the same separable preferences over consumption c and labor effort l, where $l = \frac{y}{w}$. The utility U_t^i for individual i of age t is

$$U_t^i\left(c, y\right) = u\left(c\right) - v\left(\frac{y}{w_t^i}\right), \tag{1}$$

where I assume $u'\left(\cdot\right) > 0$, $u''\left(\cdot\right) < 0$, $v'\left(\cdot\right) > 0$, $v''\left(\cdot\right) > 0$. For an individual i, lifetime utility V^i is the discounted sum of its utility at each age:

$$V^i = \sum_{t=1}^{T} \beta^{t-1} U_t^i\left(c, y\right), \tag{2}$$

where individuals discount utility flows with the factor β. Social welfare W is a weighted sum of individual lifetime utilities:

$$W = \sum_{i=1}^{I} \pi^i \alpha^i V^i, \tag{3}$$

where α^i indicates a scalar Pareto weight on individual i. In general, the form of (3) allows us to consider any point along the Pareto frontier. I assume a form for Pareto weights (specified in the numerical simulations) in which an individual's weight is positive, bounded above by one, and not increasing in its lifetime income-earning ability.

1.2   Social planner's problem in three policy scenarios

Now I derive optimal taxes in three policy scenarios using the techniques of modern dynamic optimal tax analysis. In this approach, the tax problem is recast as a problem for a fictitious social planner that uses a direct mechanism to allocate resources (see Golosov, Tsyvinski, and Werning, 2006 for a review).[8]

The social planner maximizes social welfare (3) by offering a menu of income and consumption pairs to individuals. Individuals choose optimally from the menu, earn the assigned income, and receive the assigned consumption. Knowing this, the planner designs its menu of $\{c, y\}$ pairs intending each pair to be chosen by a specific individual. Because individuals differ in their lifetime income-earning ability and age, I write c_t^i and y_t^i for the pair intended for the individual of type i and age t.[9]

The planner maximizes social welfare subject to two types of constraints: a feasibility constraint and incentive constraints. The feasibility constraint is:

$$\sum_{i=1}^{I} \pi^i \sum_{t=1}^{T} R^{T-t}\left(y_t^i - c_t^i\right) = 0. \tag{4}$$

which says that the lifetime paths of income must fund the lifetime paths of consumption across all types.[10]

---

[8]I show the connection between these allocations and taxes in the Appendix: an appropriate nonlinear labor income tax system implements each policy scenario's allocations. In the Static Mirrlees policy, the labor income tax is a direct function of income only; in the Partial Reform it is a direct function of income and the taxpayer's current age; while in the Full Optimum it is a direct function of the lifetime path of incomes and the taxpayer's current age.

[9]In the language of the direct mechanism, each individual makes a report to the planner about its personal characteristics. The planner then assigns to each report a consumption and income pair, $\{c, y\}$.

[10]Note that taxation is purely redistributive. A positive net revenue requirement would imply larger tax distortions on average, likely increasing the welfare gain from age dependence calculated below.

The planner can transfer resources across time and earn or pay the gross rate R. I assume $\beta R = 1$ for simplicity.

Incentive constraints reflect that individuals choose from the planner's menu of $\{c, y\}$ pairs to maximize their utility. In this approach, these constraints take the form of inequalities ensuring that each individual chooses the allocation of c_t^i and y_t^i that the planner intended.[11]

Importantly, variations in the set of incentive constraints allow us to succinctly distinguish the planner's problem in three policy scenarios: Static Mirrlees, Partial Reform, and Full Optimum. I now state these planner's problems formally and discuss the differences between them.

The Static Mirrlees planner in the baseline model solves the following problem:

Problem 1 *(Static Mirrlees: Age-Independent)*

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i V^i,$$

subject to the feasibility constraint (4) and the incentive constraints

$$\beta^{t-1}\left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right)\right) \geq \beta^{t-1}\left(u\left(c_s^j\right) - v\left(\frac{y_s^j}{w_t^i}\right)\right) \tag{5}$$

for all $i, j \in \{1, 2, ..., I\}$ and $t, s \in \{1, 2, ..., T\}$.

These incentive constraints mean that the Static Mirrlees planner must guarantee that each individual of type i and age t chooses the allocation intended for it over that intended for any other individual of type j and any age s. To see this, note that each side of the inequality (5) equals period utility for an individual of type i and age t. The left-hand side is the utility this individual obtains by earning y_t^i and consuming c_t^i, while the right-hand side is the utility it obtains by earning y_s^j and consuming c_s^j. The inequality guarantees that this individual weakly prefers its $\{c, y\}$ allocation. I denote the multipliers on these constraints with $\left\{\mu_{s|t}^{j|i}\right\}_{s,t}^{i,j}$, where $\mu_{s|t}^{j|i}$ corresponds to the constraint preventing individual i of age t from preferring the allocation intended for individual j of age s.

The incentive constraints in (5) capture the restriction on the Static Mirrlees policy that each individual can choose among the same menu of income and consumption pairs, regardless of age. This is different from the requirement that two individuals with the same wage but different ages receive the same allocations of c and y. The latter is a stronger condition and restricts the Static Mirrlees planner more than is justified. Age may affect an individual's optimal choice of income and consumption, even if its wage is unchanged, and the Static Mirrlees planner can take advantage of this without making taxes age-dependent.[12]

The Partial Reform planner in the baseline model solves the following problem:

Problem 2 *(Partial Reform: Age-Dependent)*

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i V^i$$

---

[11] These incentive constraints reflect this approach's application of the Revelation Principle, by which we can restrict attention to incentive-compatible direct mechanisms, i.e., where individuals reveal their true types to the planner.

[12] In the language of the direct mechanism, individuals in the Static Mirrlees scenario report only their current wage, not their age. But, the planner can assign up to T different $\{c, y\}$ pairs to each reported wage, knowing that individuals of different ages may choose different allocations. The stronger alternative would further constrain the Static Mirrlees problem, magnifying the results of the paper.

subject to the feasibility constraint (4) and the incentive constraints

$$\beta^{t-1} \left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) \right) \geq \beta^{t-1} \left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^i}\right) \right) \tag{6}$$

for all $i, j \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$

Because of the Partial Reform planner's ability to restrict individuals to age-specific allocations, these incentive constraints say that the planner must guarantee only that each individual i of age t chooses the allocation intended for it over that intended for any other individual j of the *same* age t. To see this, notice that the right-hand side of (6) depends on c_t^j and y_t^j, so that both sides of the inequality are specific to age t (compare this to the Static Mirrlees planner, where the right-hand side depended on c_s^j and y_s^j). Formally, the set of constraints (6) is a subset of (5). This makes the set of incentive constraints in the Partial Reform planner's problem weakly easier to satisfy than the set in the Static Mirrlees planner's problem. I denote the multipliers on these constraints with $\left\{\mu_t^{j|i}\right\}_t^{i,j}$, where $\mu_t^{j|i}$ corresponds to the constraint preventing individual i of age t from preferring the allocation intended for individual j of age t.

In practical terms, age dependence means that taxes can be tailored to the wage distribution at each age and that transfers can be made between age groups. As we will see in the numerical results below, these turn out to be valuable tools because the distribution of wages varies with age.

Finally, the Full Optimum planner in the baseline model solves the following problem:

Problem 3 *(Full Optimum: Age-Dependent and History-Dependent)*

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i V(i)$$

subject to the feasibility constraint (4) and the incentive constraints

$$\sum_{t=1}^{T} \beta^{t-1} \left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) \right) \geq \sum_{t=1}^{T} \beta^{t-1} \left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^i}\right) \right). \tag{7}$$

for all $i, j \in \{1, 2, ..., I\}$.

These incentive constraints reflect the Full Optimum planner's ability to make, and commit to, history-dependent allocations.[13] History dependence allows the planner to hold an individual to the lifetime path of allocations intended for a single type at all ages. Thus, the Full Optimum planner must guarantee only that each individual i chooses the *lifetime path* of allocations intended for it over that intended for any other type j. This is apparent from (7) in that each side of the inequality is a discounted sum of period utilities over individual i's lifetime. The left-hand side is i's lifetime utility if it chooses its intended allocations $\left(c_t^i, y_t^i\right)$ at each age t, while the right-hand side is i's lifetime utility from claiming the allocations intended for type j at each age. I denote the multipliers on these constraints with $\left\{\mu^{j|i}\right\}^{i,j}$, where $\mu^{j|i}$ corresponds to the constraint preventing individual i from preferring the lifetime allocation intended for individual j.

Using history dependence to satisfy incentives on a lifetime basis can be a powerful tool for the planner. For example, suppose the planner wants to give individual i a generous allocation later in life in exchange

---

[13] The assumption that the planner can commit to a path of allocations is standard in the dynamic optimal tax literature. Bisin and Rampini (2005) study the impacts of relaxing that assumption.

for a "bad" allocation early in life. In the Full Optimum, the planner can offer that path of allocations to the individual because it can make later allocations dependent on earlier ones. In the Static Mirrlees or Partial Reform scenarios, such a path is not sustainable. In those scenarios, individuals know that the planner cannot reward early sacrifice because it cannot use history dependence, so they will not accept the bad allocation early in life.

These three sets of incentive constraints, and their corresponding policy scenarios, lie along a continuum of sophistication,

Least sophisticated   **Most sophisticated**

$\longleftrightarrow$

Static Mirrlees   Partial Reform   Full Optimum

where the precise meaning of "sophistication" depends on the characteristics of the economy being considered. In this baseline model, where there is no private saving or borrowing, a policy's sophistication depends on two features: age dependence and history dependence.[14] The Static Mirrlees policy is neither age-dependent nor history-dependent. At the other extreme, in the tradition of recent work on optimal dynamic taxation, the Full Optimum policy is both age-dependent and history-dependent.[15] Partial Reform policy strikes a middle ground between the Static Mirrlees and Full Optimum, being age-dependent but not history-dependent.

We can see this relationship between the three policies formally. The transition between the Static Mirrlees incentive constraints (5) and the Full Optimum incentive constraints (7) can be decomposed into two steps. First, the planner replaces all the s subscripts in (5) with t subscripts, limiting individuals to claiming only those allocations intended for individuals of the same age. This adds age dependence. Second, the planner adds the discounting and summations over t that appear in (7), using its ability to commit to lifetime allocations. This adds history dependence. The Partial Reform policy takes only the first of these two steps.

Before analyzing these models in detail, I note that the Static Mirrlees policy is not designed to match the detailed structure of existing tax policy on labor income. Rather, it is the optimal policy constrained by two characteristics of existing tax policy: age independence and history independence. Our comparison of the Partial Reform to this Static Mirrlees policy rather than to existing tax policies allows us to isolate the potential for age dependence to generate welfare gains by itself, relative to an age-neutral benchmark.[16]

1.3 Analytical results

Now, I compare the allocations chosen by the social planner in the Static Mirrlees, Partial Reform, and Full Optimum policies along two margins: the intratemporal margin between consumption and leisure and the intertemporal margin between consumption in one period and the next. In particular, I derive theoretical results on these margins that connect to and extend classic results and intuition from static and dynamic Mirrleesian tax analysis. The importance assigned to these margins is due to their relation, conceptually

---

[14] This informal discussion of sophistication does not apply to the models of the economy (discussed later) that include private saving and borrowing. Rather than specifying a variant on sophistication for each model economy, we rely on the formal incentive constraints to clarify the distinctions.

[15] History dependence cannot be avoided in this baseline economy by using wealth to encode past wages, such as in Albanesi and Sleet (2006), because individuals cannot accumulate wealth.

[16] In addition to the simulations of optimal policy presented below, I have simulated a policy approximating the current U.S. tax system. Using the data and parameterization described in the paper, I apply the 1999 U.S. income tax schedule to labor earnings and simulate individuals' behavior. The welfare gain from that policy to the Static Mirrlees policy is equivalent to a 6.5 percent increase in aggregate annual consumption.

and formally, to marginal taxes on labor income and taxes on capital income. I evaluate average taxes, a key dimension along which the three policy scenarios differ, in the section using numerical simulations. The Mirrlees model, and this extension of it, are not well-suited to discussing average taxes analytically.

1.3.1 Intratemporal distortions

First, I compare the distortions to individuals' choices of how much income to earn: i.e., distortions on their marginal choices between consumption and leisure. I begin with a definition:

Definition 1 (Intratemporal Distortion) *The intratemporal distortion for an individual of type i and age t is denoted $\tau(i,t)$ and equals*

$$\tau(i,t) = 1 - \frac{v'\left(\frac{y_t^i}{w_t^i}\right)}{w_t^i u'\left(c_t^i\right)}. \tag{8}$$

Denote $\tau^{SM}(i,t)$, $\tau^{PR}(i,t)$, and $\tau^{FO}(i,t)$ as the intratemporal distortions for an individual of type i and age t in the solutions to the Static Mirrlees (SM), Partial Reform (PR), and Full Optimum (FO) planner's problems.

In expression (8), positive $\tau(i,t)$ distorts the individual's choice away from work (and consumption) and toward leisure. If $\tau(i,t) = 0$, the individual sets the marginal utility from an extra unit of consumption equal to the marginal disutility of earning it, so there is no distortion on this margin.

With these preliminaries, we can now turn to the results. The following proposition serves as a benchmark:

Proposition 1 (Intratemporal Benchmark) *Let the utility function for all $i \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$ be defined by (1), and let $\beta R = 1$. If $w_s^i = w_t^i$ for all $i \in \{1, 2, ..., I\}$ and $s, t \in \{1, 2, ..., T\}$, then*

1. *$\tau^{SM}(i,s) = \tau^{SM}(i,t)$, $\tau^{PR}(i,s) = \tau^{PR}(i,t)$, and $\tau^{FO}(i,s) = \tau^{FO}(i,t)$ for all $i \in \{1, 2, ..., I\}$ and $s, t \in \{1, 2, ..., T\}$.*

2. *$\tau^{SM}(i,t) = \tau^{PR}(i,t) = \tau^{FO}(i,t)$ for each $i \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$.*

Proof. In Appendix. ∎

This proposition is about an economy in which each individual has a constant wage throughout its lifetime: $w_s^i = w_t^i$ for all $i \in \{1, 2, ..., I\}$ and $s, t \in \{1, 2, ..., T\}$. This means that the distribution of wages in the population is the same at every age. In such an economy, each individual faces the same distortion at all ages in each policy.[17] For example, in the Static Mirrlees policy, $\tau^{SM}(i,s) = \tau^{SM}(i,t)$ for all $i \in \{1, 2, ..., I\}$ and $s, t \in \{1, 2, ..., T\}$. In addition, each individual faces the same distortion at a given age in all three policy scenarios, so $\tau^{SM}(i,t) = \tau^{PR}(i,t) = \tau^{FO}(i,t)$ for each $i \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$.

Intuitively, an unchanging wage distribution means that each age is a replica of the others, so that the planner's best solution for one age will be the best solution for all ages. This means constant intratemporal distortions for each individual over time. It also means that the three planners are solving equivalent problems, so they choose the same pattern of intratemporal distortions.

As this proposition implies, any differences between the three policy scenarios' intratemporal distortions rely on individuals' lifecycle wage paths not being constant. Before characterizing these distortions in general, I focus on a specific distortion: that at the top of the income distribution.

---

[17]Werning (2007a) proves a similar result for the optimal dynamic policy.

The top marginal distortion In this subsection, I focus on a classic result from static optimal tax analysis: the top earner in the economy should face *no* intratemporal distortion.[18] The intuition for the classic result is that an intratemporal distortion has both a cost and a benefit. The cost is that it causes individuals to work and consume differently than they would without taxes, leading to either lower utility or less efficiently-provided utility for these individuals. The benefit is that it enables the planner to collect more tax revenue from higher earners, increasing the extent of redistribution.[19] At the top of income distribution, this benefit is zero (there are no higher earners from whom to collect more tax revenue). Thus, a distortion on the top earner solely discourages effort, and it is avoided in the optimal policy.

   The following proposition describes how this classic result applies to a dynamic economy under this paper's three policy scenarios:

Proposition 2 (Top Marginal Distortion) *In the baseline economy, for*

1. *if $w_t^i \geq w_t^j$ for all $j \in \{1, 2, ..., I\}$, then $\tau^{PR}(i, t) \leq 0$ and $\tau^{FO}(i, t) \leq 0$,*

2. *if $w_t^i \geq w_t^j$ for all $j \in \{1, 2, ..., I\}$ and for all $t \in \{1, 2, ..., T\}$, and if $a^i \leq a^j$ for all $j \in \{1, 2, ..., I\}$, then $\tau^{PR}(i, t) = 0$ and $\tau^{FO}(i, t) = 0$ for all $t \in \{1, 2, ..., T\}$, and $\tau^{SM}(i, t) = 0$ for t such that $w_t^i \geq w_s^i$ for all $s \in \{1, 2, ..., T\}$.*

Proof. In Appendix. ■

   The first part of this proposition states that the highest wage earner at each age faces a nonpositive intratemporal distortion whenever the social planner can condition taxes on age: that is, in the Partial Reform and Full Optimum policy scenarios. In particular, the distortion on the top earner at each age could be negative in these scenarios. The second part of this proposition states that an individual who is the highest wage earner at all ages faces no intratemporal distortion at any age in the Partial Reform and Full Optimum scenarios, but only at its peak-earnings age in the Static Mirrlees scenario.

   The first part of this proposition is a surprising theoretical deviation from the classic static optimal tax result because it allows for the possibility that, with age dependence, the top earner at a given age may face a *negative* distortion.[20] Intuitively, suppose an individual has the highest wage within its current age but also has a low lifetime income-earning potential. An example might be an entertainer or athlete whose earnings power is temporarily high at a young age. The planner wants to assign this individual both high income and high consumption: the former because it is a productive worker, and the latter because its welfare weight is relatively large. A negative marginal distortion makes this possible. With it, the planner can levy high average taxes on individuals who earn less at the current age but more over their lifetimes, for example consultants or lawyers, while using the negative distortion to reduce the tax burden on the top current earner. The lower earners will not be willing to earn enough income to qualify for the negative

---

[18] Diamond (1998) and Saez (2002) show that this result depends on the shape of the wage distribution. With a bounded wage distribution such as that used throughout the paper, the zero top rate result always holds. In the Appendix, I show that the results presented here have analogues in a model with a wage distribution more similar to that which they use.

[19] It does this by taking advantage of differences in the marginal disutility of income across individuals with different wages. A distortion on a given individual i harms i less than it harms a higher-skilled individual j claiming i's allocations. While i suffers lower consumption due to the distortion, it also enjoys working less. For j, the loss in consumption exacerbates the main disadvantage to claiming i's allocation, while the decrease in required income is of less value because j is already working much less than it would to earn its own allocation. Thus, the planner can assign a larger lumpsum transfer to i while discouraging j from claiming i's allocation.

[20] The optimality of a negative top distortion has also been suggested by Judd and Su (2006), but for a different reason. In their model with multiple dimensions of heterogeneity, the interaction of wage and labor supply elasticity differences can justify negative top distortions. Judd and Su also perform an illustrative simulation of age-dependent taxation.

distortion, so the planner is able to target its resources.[21]

The second part of the proposition highlights a key difference in intratemporal distortions between the Static Mirrlees and the two more sophisticated scenarios. To see why it holds, consider an individual i^* with the highest wage at all ages. As mentioned above, the potential benefit of a positive marginal distortion on any individual is that it makes possible the collection of more tax revenue from higher earners. If there are no higher earners than i^* at each age, positive distortions on i^* have costs but no benefits, so they are avoided by the planners with access to age-dependent taxes. The Static Mirrlees planner, on the other hand, faces a more difficult problem. While no other individuals of the same age earn more than i^*, some individual of a different age (perhaps i^* itself) earns more than i^* for each age except the age at which i^*'s earnings peak. Thus, in order to collect revenue from the highest earners across all age groups, the planner will use distortions on i^* at all but its peak-earnings age.[22] A simple numerical illustration of this is provided in the Appendix for the interested reader.

Though the second part of the proposition is about a special topic, the top marginal distortion, it highlights the fundamental limitation of the Static Mirrlees policy relative to the Partial Reform and Full Optimum: the former's inability to hold individuals to age-specific tax schedules. The effects of this limitation ripple throughout the wage distribution, allowing the more sophisticated planners to distort individuals' labor efforts less than the Static Mirrlees planner. Later, we will quantify these effects and their impact on social welfare in the numerical simulations below. Next, however, we turn to characterizing intratemporal distortions other than at the top of the income distribution.

General characterization   In this subsection, I begin by giving formal expressions for the intratemporal distortions in each scenario. As in classical Mirrleesian analysis, characterizing intertemporal distortions in general is difficult due to their dependence on the details of the wage distribution. Thus, the formal expressions below are not in closed form. Nevertheless, we can use them to determine the key forces driving distortions and to address some natural questions about the pattern of distortions in each policy.[23] To simplify the results, I assume disutility takes the isoelastic form

$$v\left(\frac{y_t^i}{w_t^i}\right) = \frac{1}{\sigma}\left(\frac{y_t^i}{w_t^i}\right)^{\sigma}. \tag{9}$$

where the parameter $\frac{1}{\sigma-1}$ gives the constant-consumption elasticity of labor supply.

---

[21] In the language of the mechanism, a worker with a lower current wage but a higher lifetime income-earning potential may be tempted to claim the allocation assigned to this top current earner. To prevent this, the planner makes such a claim more costly for the lower-skilled worker by negatively distorting the top earner, increasing its pre-tax income. Following a similar logic, a negative distortion is, in principle, possible in the Static Mirrlees for the top earner in the entire economy across ages if that top earner is, at other ages, a low earner.

[22] In the language of the direct mechanism, no other individual of the same age wants to mimic i^*. The only reason to distort an individual is to discourage mimicking by others, so the age-dependent policies leave i^* undistorted. In the Static Mirrlees, some individual of a different age (perhaps i^* itself) earns more than i^* for each age except the one at which i^*'s earnings peak. Thus, i^* will be a tempting target for mimicking, and the Static Mirrlees planner must distort it at all but its peak-earnings age.

[23] Similar conclusions to those below, as well as closed form results for intratemporal distortions, are obtainable for a continuous wage distribution from an analysis using the Hamiltonian methods familiar from static Mirrlees analysis such as Saez (2002) or Salanie (2003). Such an analysis for the baseline model can be found in the Appendix. The method does not extend to the dynamic settings with private saving considered later in this paper (Case 2 and Case 4), so I use the alternative method, conventional in the dynamic optimal tax literature, throughout the paper.

In the Static Mirrlees scenario, the intratemporal distortion on worker of type i and age t is:

$$\tau^{SM}(i,t) = \frac{\sum_{s=1}^{T} \sum_{j=1}^{I} \left(1 - \left(\frac{w_t^i}{w_s^j}\right)^{\sigma}\right) \beta^{s-t} \mu_{t|s}^{i|j}}{\alpha^i \pi^i + \sum_{s=1}^{T} \sum_{j=1}^{I} \mu_{s|t}^{j|i} - \sum_{s=1}^{T} \sum_{j=1}^{I} \left(\frac{w_t^i}{w_s^j}\right)^{\sigma} \beta^{s-t} \mu_{t|s}^{i|j}}. \tag{10}$$

where, as stated after (5), $\mu_{t|s}^{i|j}$ is the multiplier on the incentive constraint preventing individual i of age t from claiming the allocation of any other individual j of age s.

In the Partial Reform scenario, it is:

$$\tau^{PR}(i,t) = \frac{\sum_{j=1}^{I} \left(1 - \left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\right) \mu_t^{i|j}}{\alpha^i \pi^i + \sum_{j=1}^{I} \mu_t^{j|i} - \sum_{j=1}^{I} \left(\frac{w_t^i}{w_t^j}\right)^{\sigma} \mu_t^{i|j}}. \tag{11}$$

where $\mu_t^{i|j}$ is the multiplier on the incentive constraint preventing individual i of age t from claiming the allocation of any other individual j of the same age t.

In the Full Optimum scenario, it is:

$$\tau^{FO}(i,t) = \frac{\sum_{j=1}^{I} \left(1 - \left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\right) \mu^{i|j}}{\alpha^i \pi^i + \sum_{j=1}^{I} \mu^{j|i} - \sum_{j=1}^{I} \left(\frac{w_t^i}{w_t^j}\right)^{\sigma} \mu^{i|j}}.$$

where $\mu^{i|j}$ is the multiplier on the incentive constraint preventing individual i from claiming the lifetime allocation of any other individual j.

One lesson we learn from these expressions is that the tradeoff familiar from static Mirrlees analysis remains relevant in this baseline dynamic model. As discussed after Proposition 2, increasing the distortion on individual i of age t has a cost and a benefit to the planner. Consider the expression for $\tau^{PR}(i,t)$, the intratemporal distortion on individual i of age t in the Partial Reform policy (similar analysis holds for the Static Mirrlees and Full Optimum). This distortion is increasing in $\mu_t^{i|j}$ when $\mu_t^{i|j}$ is positive for j with a higher wage than i at age t. Why? Recall that $\mu_t^{i|j}$ is the cost in terms of social welfare to the Partial Reform planner of ensuring that j prefers its allocation to i's when of age t. The planner has two tools to use in satisfying j's incentives: it can allocate resources to j that it would prefer to allocate to those lower in the earnings distribution, or it can distort the allocations to lower types in order to make them less tempting to j. Because the planner has redistributive tastes (or, because utility is concave) and the latter tool allows for a more egalitarian distribution of resources, a distortion on i can be beneficial despite its cost in reducing i's labor effort. The more costly it is to satisfy j's incentives (i.e., the larger is $\mu_t^{i|j}$) the more the planner is willing to distort i rather than transfer resources to j.

We can also use these expressions to consider more specific topics. Next, I address three natural questions about the structure of intratemporal distortions in an age-dependent tax policy.

First, does an individual face rising or falling intratemporal distortions over its lifecycle? In the Static Mirrlees scenario, the tax schedule is constant over the lifecycle, so the answer to the question depends only on two factors: (1) the shape of that tax schedule, and (2) whether the individual moves up or down in the income distribution over its lifetime. Of course, these factors depend on the individual's wage path, the distribution of all wages in the economy, and the planner's preferences, so we can say very little about the lifecycle path of intratemporal distortions in general. The answer is yet more complicated in the Partial

Reform scenario, because that planner sets policy based on an individual's wage relative to its own age's wage distribution. Thus, the path of distortions for an individual is sensitive not only to the individual's wage path but also to the wage paths of all individuals of the same age, further complicating the general characterization of the lifecycle path of distortions. I provide a numerical illustration of this sensitivity in the Appendix.

Second, how would variation in the elasticity of labor supply with age affect the intratemporal distortions an individual faces? If an isolated individual is more elastic at one age than at all other ages, it is more responsive to distortions at that age. This affects the tradeoff facing planners that can use age dependence, who respond by lowering the intratemporal distortion on the individual at that age. In contrast, the Static Mirrlees planner is unable to respond because it cannot hold individuals to age-specific tax schedules. General results are difficult for broader differences in elasticity, such as when all individuals are more elastic at some ages than at others. Intuitively, if all individuals are more elastic, a larger distortion on some may be optimal to enable redistribution of income without discouraging effort among the high-skilled.[24] Nevertheless, in Section 5 I find that the intuition from the individual case carries through for a calibrated numerical simulation with elasticity differences by age, so that intratemporal distortions are lower at ages with more elastic labor supply in the Partial Reform and Full Optimum scenarios.

Third, if two individuals earn the same current income at different ages, who should face the higher distortion on that income? Intuitively, suppose a manual laborer earns the same income when middle-aged as a professional earns when young. In the Static Mirrlees, these workers are treated identically (i.e., they face the same distortion) because they have the same position in the overall earnings distribution. In contrast, the Partial Reform and Full Optimum planners are able to distinguish between these workers and consider each worker's place in its own age-specific distribution, so they need not face the same distortion. It is ambiguous who will face the larger distortion, because the relative benefit from distorting the two workers is ambiguous. Distorting the young professional raises more tax revenue from a population of higher earners that is both smaller (implying a smaller distortion) and richer in lifetime income (implying a larger distortion) than the population of higher earners above the manual laborer. A numerical illustration of this is provided in the Appendix.

As is standard in the Mirrleesian optimal tax literature, in Section 1.4 I turn to calibrated simulations to further explore these questions and others. Before doing so, however, I discuss a second set of analytical results, these on the intertemporal consumption margin.

1.3.2 Intertemporal Distortions

In this section, I analytically characterize the intertemporal consumption margin in an age-dependent policy and compare it to well-known results from the dynamic optimal tax literature. I show that age-dependent policy satisfies a condition that improves on age-independent policy but falls short of the full optimum. The recent development of a dynamic Mirrlees literature has highlighted this margin because optimal distortions to it, most prominently characterized by Golosov, Kocherlakota, and Tsyvinski (2003), have renewed interest in the taxation of capital after a long period during which the Chamley-Judd result of zero optimal capital taxation held sway.

Consider an individual's problem of maximizing lifetime utility (2) given a wage path $\left\{w_t^i\right\}_{t=1}^T$ and the lifetime budget constraint $\sum_{t=1}^T R^{T-t}\left(y_t^i - c_t^i\right) = 0$. Continue to assume $\beta R = 1$. This individual's optimal

---

[24] This ambiguity is in part due to the distinction between intratemporal distortions and marginal taxes (which include changes in lump-sum grants and taxes) in a model with a discrete wage distribution. As shown in the Appendix, age-dependent marginal rates are always inversely correlated with labor supply elasticities for a continuous wage distribution.

choice of consumption satisfies, for each $(t, t+1)$ pair:

$$u'\left(c_t^i\right) = u'\left(c_{t+1}^i\right).$$ (12)

This is the familiar intertemporal Euler condition that sets the marginal utility from consumption equal across periods. Expression (12) represents an undistorted intertemporal margin.

I analyze the extent to which the planners' chosen allocations distort this intertemporal margin. Recall that individuals cannot save or borrow in the baseline economy. The following Proposition serves as a benchmark:

Proposition 3 (Intertemporal Benchmark) *Let the lifetime utility function for all* $i \in \{1, 2, ..., I\}$ *be defined by (2), and let* $\beta R = 1$. *If* $w_s^i = w_t^i$ *for all* $i \in \{1, 2, ..., I\}$ *and* $s, t \in \{1, 2, ..., T\}$, *then* $u'\left(c_t^i\right) = u'\left(c_{t+1}^i\right)$ *for all* $i \in \{1, 2, ..., I\}$ *and* $t, t+1 \in \{1, 2, ..., T\}$ *in the Static Mirrlees, Partial Reform, and Full Optimum planner's problems.*

Proof. In Appendix. ∎

This proposition, a parallel to Proposition 1 above, states than no individual faces an intertemporal distortion in any of the three policy scenarios if each individual has a constant wage over its lifetime.[25] Intuitively, if each age is a replica of the next, the allocations to each individual will be the same at each age.

While this proposition provides a useful benchmark, we are interested in more realistic settings, namely with changing wage distributions. Solving the planner's problems as stated at the start of this section, we can obtain the following results.

The Full Optimum planner's allocations satisfy, for all $i \in \{1, 2, ..., I\}$ and $t, t+1 \in \{1, 2, ..., T\}$:

$$u'\left(c_t^i\right) = u'\left(c_{t+1}^i\right),$$ (13)

a classic Atkinson and Stiglitz (1976) result: i.e., the Full Optimum policy does not distort intertemporal allocations.[26] This result depends on the Full Optimum planner's ability to use history-dependent allocations, as is made clear by the contrast between expression (13) and the result for the Partial Reform planner to which I now turn.

The Partial Reform planner's allocations satisfy, for individual i and ages $t, t+1$:

$$u'\left(c_t^i\right) = \left(\frac{\pi^i \alpha^i + \sum_{j=1}^{I} \mu_{t+1}^{j|i} - \sum_{j=1}^{I} \mu_{t+1}^{i|j}}{\pi^i \alpha^i + \sum_{j=1}^{I} \mu_t^{j|i} - \sum_{j=1}^{I} \mu_t^{i|j}}\right) u'\left(c_{t+1}^i\right).$$ (14)

The ratio in parentheses in (14) is generally different from one, implying that the Partial Reform planner imposes a distortion on the intertemporal margin. To see this, recall that $\mu_t^{j|i}$ and $\mu_{t+1}^{j|i}$ are the multipliers on the incentive constraints preventing type i from claiming type j's allocation at ages t and $t+1$, respectively. Unless $\mu_t^{j|i} = \mu_{t+1}^{j|i}$ for all i, j, and t, the ratio in parentheses in (14) is not equal to one. Intuitively, whenever the incentive problems facing the planner differ across ages, the Partial Reform planner generally fails to satisfy the intertemporal Euler equation. The Partial Reform planner must satisfy incentives at

---

[25] Werning (2007a) proves a similar result for the optimal dynamic policy.

[26] Note that wages are deterministic, so there is no reason for an intertemporal distortion along the lines of Rogerson (1985) or Golosov, Kocherlakota, and Tsyvinski (2003). In Sections 3 and 4, I analyze generalizations of the baseline model where wages are stochastic.

each age, so changing wage distributions make the allocation for an individual differ across ages, violating the intertemporal Euler equation.

The ability to satisfy the standard Euler equation and provide smooth consumption is a substantial advantage for the Full Optimum planner relative to the Partial Reform (or Static Mirrlees[27]) planner. Consider two workers who have similar lifetime incomes in present value: say, an engineer and a lawyer. The engineer has a relatively flat earnings profile over its lifetime, while the lawyer has low earnings when young but a steep earnings profile that eventually raises its earnings well above the engineer's. Because these workers have similar lifetime incomes, the planner weighs them equally in the social welfare function. It would like to give them both smooth paths of consumption that have similar present values while concentrating their labor supply on the ages at which they are most productive. The Full Optimum planner is able to achieve these goals, even though this includes giving the engineer the same consumption as the lawyer when young but having the engineer earn much more income. The Full Optimum planner achieves this by promising the engineer that the situation will be reversed when they are older: the engineer will have the same consumption as the lawyer but will have to earn much less.

The Partial Reform and Static Mirrlees planners cannot achieve the smoothed consumption paths obtained by the Full Optimum planner because they cannot make history-dependent allocations. In other words, they cannot promise to reward sacrifice at a later date. Suppose the Partial Reform planner tried to match the Full Optimum's allocation just described for the engineer and lawyer. Knowing that the planner would not be able to reward her for earning high income and accepting low consumption, the young engineer will be tempted to work less and claim the lower average tax rates intended for the lawyer. Similarly, the lawyer will know that the planner cannot prevent him from working less and claiming the lower average tax rates intended for the engineer when old. Thus, the planner cannot tailor labor effort while smoothing consumption. Instead, it more closely aligns the lawyer's consumption to its earnings path and skews the engineer's consumption in the other direction. This encourages the lawyer to work more when old and the engineer to work more when young, solving the incentives problem. But, it requires intertemporal distortions. A simple numerical example of this two-type economy is provided in the Appendix.

While at a disadvantage to the Full Optimum, the Partial Reform allocations do satisfy what I will call the "Symmetric Inverse Euler Equation." The following proposition states the result.

Proposition 4 (Symmetric Inverse Euler) *Let the lifetime utility function for all $i \in \{1, 2, ..., I\}$ be defined by (2). Then, the solution to the Partial Reform planner's problem satisfies:*

$$\sum_{i=1}^{I} \frac{\pi^i}{u'(c(i,t))} = \sum_{i=1}^{I} \frac{\pi^i}{u'(c(i,t+1))}, \qquad (15)$$

for any $t, t+1 \in \{1, 2, ..., T\}$.

Proof. In Appendix. ∎

This "Symmetric Inverse Euler Equation" guarantees that resources are being allocated efficiently between age groups, as it equalizes across ages the cost (in consumption) of increasing welfare. Though not as powerful a restriction as the intertemporal Euler equation, the Symmetric Inverse Euler Equation is nevertheless an achievement of Partial Reform that the Static Mirrlees planner cannot replicate. Because it cannot restrict individuals to age-specific tax schedules, the Static Mirrlees planner cannot make these efficient transfers across ages.

---

[27] I omit the Static Mirrlees planner's intertemporal result for brevity, but it is an intuitive modification of (14).

1.3.3 Summary

The analytical results above make clear that an age-dependent policy differs substantially from an age-independent one and resembles in many important ways a fully optimal policy. They show how theoretical results for age-dependent policy in a dynamic economy with lifecycle wage paths connect to and extend prominent results from the static and dynamic optimal tax literature. Moreover, they identify the factors at play in determining the pattern of both intratemporal and intertemporal distortions under age dependence.

As with most Mirrleesian analyses, however, these results can pin down only select characteristics of tax policy, and on many questions they provide only ambiguous theoretical guidance. To provide a more general characterization of policy and to answer many of these questions, I turn to numerical simulations.

1.4 Numerical results

The conventional, static Mirrlees optimal tax literature was linked to calibrated simulations of optimal policy from its beginnings. In contrast, the dynamic Mirrleesian literature has thus far used mostly abstract, illustrative simulations to reveal the effects of its recommendations.[28] In this section, I calibrate the dynamic optimal tax problems specified above to detailed individual wage data from the U.S. PSID, simulate and characterize policy, and quantify the welfare impacts of reform. I begin by describing the data and parameter specification.

1.4.1 Data

The structure of the baseline model matches a well-known existing empirical literature on lifecycle income distributions and wage paths, beginning with Fullerton and Rogers (1993) and including recent work such as Altig, Auerbach, Kotlikoff, Smetters, and Walliser (2001) and Diamond and Tung (2006)[29]. My approach follows that literature.

The goal is to construct representative lifetime wage paths for I groups of individuals classified according to lifetime income-earning potential. The construction of the required data can be divided into four steps. First, I limit the data to household heads from the U.S. PSID core sample for the years 1968-2001 and collect data on their income, hours worked, age, race, gender, and education for each year they are a head of household. Second, I calculate reported real wages for each observation by dividing reported labor income by reported hours (a potentially noisy measure of the wage but the best one available) and inflating or deflating the data with the CPI to put all wages in 1999 dollars.[30] Third, I remove potentially problematic observations by eliminating all those for which reported annual hours were less than 500 or greater than 5,824, for which reported labor income was zero but hours were positive, or for which the nominal wage implied by earnings and hours was less than half the applicable minimum wage in that calendar year. After these adjustments, the dataset contains approximately 155,000 observations on just over 19,000 individuals with an average of 8.1 years observed per person. Fourth, I use these data to estimate a weighted (by the PSID sample weights) individual fixed-effects regression of the log wage on a quartic in age and interaction terms that multiply education, gender, and race with both age and age-squared. With the results of this

---

[28] An exception is Golosov and Tsyvinski (2006), who calibrate their disability model.

[29] Thanks to John Diamond and Joyce Tung for providing helpful advice on the construction of the dataset.

[30] When using an empirical income distribution to infer the distribution of skills and simulate optimal taxes, it is important to back out the effects of the current tax system on income as in Saez (2001). With data on income and hours, it is possible to calculate wages directly, instead.

estimation, I predict a wage path for each individual. The resulting dataset contains the predicted wage paths over a full lifecycle as well as the observed wage paths for the sample of household heads.[31]

In the baseline economy, individuals are classified into types $i = \{1, 2, ..., I\}$, where i indexes lifetime income-earning potential. The natural empirical counterpart to this is the amount of income an individual can earn over its lifetime, given its lifecycle wage path. As Fullerton and Rogers (1993) express it,

$$\text{Potential Lifetime Income} = \sum_{t=1}^{T} \frac{(\hat{w}_t \times 4,000)}{(1+r)^{t-1}}$$

where T is set to 55 to include predicted income over the age range 21-75, $\hat{w}_t$ is the predicted wage rate at age t, the number 4,000 represents maximum feasible hours of work in a year (e.g., 80 hours per week for 50 weeks), and I set the gross rate of return $1 + r = 1.05$.[32] After calculating potential lifetime income for each individual, I divide the sample into I quantiles and assign an index $i = \{1, ..., I\}$ to each individual. For the baseline model, I use $I = 10$.

The final step is to calculate a representative wage path (using observed, not predicted, wages) for each lifetime income group. I retain only those individuals who report a wage in at least four years to limit the influence of atypical observations. This eliminates approximately 13,000 observations from the data corresponding to approximately 7,500 infrequently observed individuals. Due to computational speed considerations, I calculate representative wage paths for each income group by grouping data into the three main decades of work life.[33] Within each type i, I calculate the average real wage over the ranges 30-39, 40-49, and 50-59. These averages form the series of wages $\{w_t^i\}_{t=1}^{T}$ for $i = \{1, 2, ..., I\}$ and $t = \{1, 2, 3\}$. The results are shown in Table 1 along with demographic data describing the income groups.

Table 1: Data for simulation of baseline model											
						Lifetime income decile					
		Bottom	2	3	4	5	6	7	8	9	Top
Wage paths	Age range										
Average wage by age range ($1999)	30-39	6.76	8.83	10.74	12.55	14.46	16.52	18.78	21.38	25.08	36.14
	40-49	7.02	9.81	11.73	13.94	15.93	18.23	20.48	23.83	28.79	46.52
	50-59	6.76	9.92	12.09	13.82	15.79	18.36	20.33	24.06	27.92	48.12
Descriptive data											
Proportion race=white		0.35	0.46	0.49	0.62	0.65	0.67	0.72	0.79	0.80	0.91
Proportion with college degree		0.03	0.05	0.07	0.09	0.14	0.17	0.20	0.28	0.37	0.60
Proportion gender=male		0.50	0.64	0.72	0.79	0.84	0.86	0.90	0.90	0.93	0.96
Pareto weight (calculated)		1.00	0.99	0.98	0.97	0.96	0.94	0.93	0.92	0.90	0.83

Source: PSID core sample household heads 1968-2001; author's calculations as described in paper.

Figure 1 plots these wage paths.

---

[31] Controlling for time and cohort effects, as studied in Heathcote, Storesletten, and Violante (2005), does not materially affect the results. For time effects, I control for year dummies in the estimation of individual fixed effects and in the calculation of representative wage paths for each type. For cohort effects, I perform the same calculations as for the entire sample on 5-year cohort subsets. In both cases, the resulting wage paths closely match those shown in the paper.

[32] The specific choices of $T = 55$, a maximum of 4,000 hours, and $1 + r = 1.05$ are all of minor importance, in that very few individuals would be classified into different income deciles if they were modified.

[33] Using decade-average income smooths out fluctuations in annual incomes that could make age dependence less powerful. This possibility is best examined in the stochastic wage paths extension below (Case 3), where wages are more free to fluctuate across ages. Simulations in that model suggest that this paper's results are robust to this concern.

Figure 1: Wage Paths by Lifetime-Income Decile
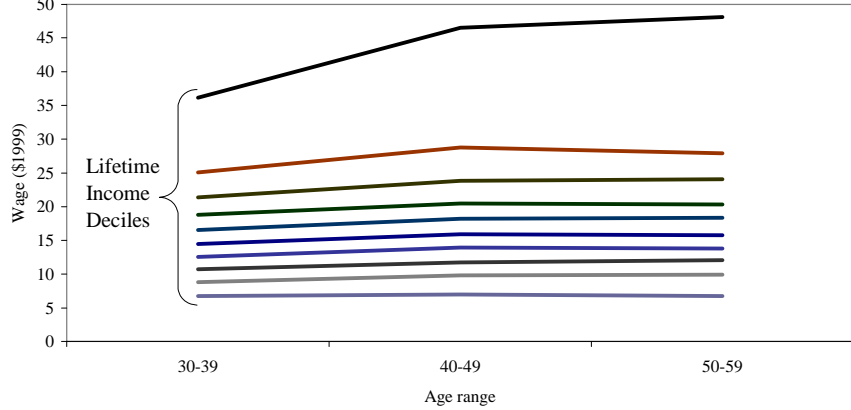
Source: U.S. PSID core sample household heads 1968-2001; author's calculations as described in paper.

The wage paths shown in Figure 1 are similar to those implied by the results of Fullerton and Rogers (1993), shown in their Table 4.11. Now I discuss the specification of the models' parameter values.

1.4.2 Parameter specification

I assume the period utility function

$$U\left(c,l\right) = \ln c - \frac{1}{\sigma}l^{\sigma},$$

and set $\sigma = 3$, which implies a constant-consumption elasticity of labor supply of 0.5. The results described below are robust to alternative parameterizations, with the welfare gains from age dependence increasing when utility from consumption is more concave and the elasticity of labor supply is greater. I use an annual gross rate of return of five percent, which implies that $R = (1.05)^{10}$ because I am using decade-long age ranges.

For the Pareto weights, I assume a form that is not increasing in an individual's lifetime income-earning ability. Formally, the Pareto weight on individual i is:

$$\alpha^i = \exp\left(-\rho \frac{\sum_{t=1}^T \frac{w_t^i}{R^{t-1}} - \sum_{t=1}^T \frac{w_t^1}{R^{t-1}}}{\sum_{i=1}^I \sum_{t=1}^T \pi^i \frac{w_t^i}{R^{t-1}}}\right). \tag{16}$$

where $\rho \geq 0$. Note that $\alpha^i = 1$ for type $i = 1$, who has the lowest lifetime income-earning ability (defined above as the present value of wages).

The parameter ρ allows us to vary the redistributive tastes of the planner: i.e., the extent to which Pareto weights decline with i. If ρ equals zero, the planner is Utilitarian, and all Pareto weights equal one. For larger ρ, Pareto weights decline with i. My benchmark will be $\rho = 0.1$, implying moderate redistributive tastes for the planner. The weights corresponding to the data described above are shown in Table 1 above. They decline from 1.00 for the bottom income decile to 0.83 for the top decile.

With these data and parameters, I simulate the policy models. Now, I turn to the results of these simulations.

1.4.3 Simulation results

I focus on four outputs from the numerical simulations: intratemporal distortions, average tax rates, intertemporal distortions, and welfare. For each, I compare the results under the Static Mirrlees, Partial Reform, and Full Optimum policy scenarios. Together, these four outputs allow me to provide a detailed characterization of optimal policy under each scenario and to quantify the welfare gains of reform from the Static Mirrlees to the more sophisticated policies.

Intratemporal distortions First, consider intratemporal distortions as defined in (8). Table 2 lists these distortions by age and lifetime income decile, while Figures 2a, 2b, and 2c plots them against annual income.[34] The lines in the figures connect discrete points corresponding to the I types at each age; the somewhat jagged pattern of distortions is a consequence of this discreteness.

Table 2: Intratemporal distortions in baseline model simulation

Intratemporal distortion						Lifetime income decile						
	Age range	Bottom	2	3	4	5	6	7	8	9	Top	Average
Static Mirrlees	30-39	0.26	0.34	0.39	0.45	0.41	0.43	0.34	0.40	0.34	0.31	
	40-49	0.34	0.35	0.33	0.34	0.36	0.28	0.32	0.23	0.50	0.04	**0.324**
	50-59	0.26	0.37	0.39	0.32	0.34	0.29	0.31	0.25	0.45	0.00	
Partial Reform	30-39	0.25	0.32	0.33	0.34	0.34	0.32	0.30	0.29	0.33	0.00	
	40-49	0.32	0.33	0.39	0.36	0.37	0.33	0.36	0.36	0.41	0.00	**0.309**
	50-59	0.34	0.35	0.33	0.36	0.40	0.30	0.39	0.31	0.43	0.00	
Full Optimum	30-39	0.27	0.33	0.35	0.36	0.35	0.34	0.32	0.31	0.36	0.00	
	40-49	0.30	0.32	0.37	0.35	0.36	0.32	0.34	0.33	0.39	0.00	**0.303**
	50-59	0.32	0.34	0.32	0.35	0.38	0.29	0.36	0.29	0.40	0.00	

---

[34]Income from the simulation results is converted to annual U.S. dollars as follows. The median annual hours worked in the data is 2,070 per year, while the Partial Reform planner has the corresponding worker exert 0.84 units of labor effort. This implies that a worker exerting one unit of labor effort per period in the model would work approximately 2,477 hours per year. I use this number as the benchmark for normal hours per year, and multiply the simulation results for income by it to obtain annual income as shown.

Figure 2a: Intratemporal Distortions in Static Mirrlees, Baseline Model
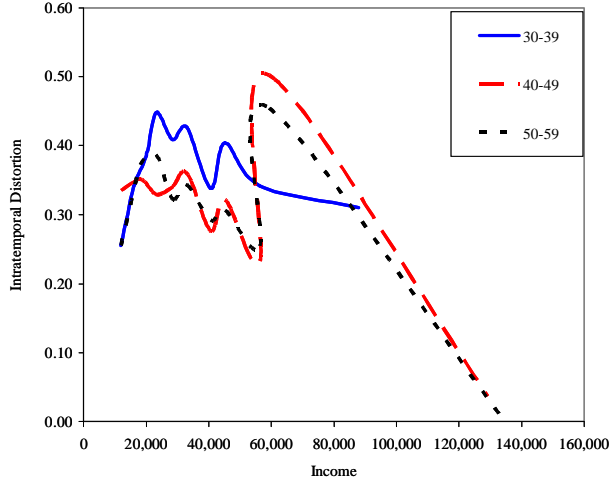
Figure 2b: Intratemporal Distortions in Partial Reform, Baseline Model
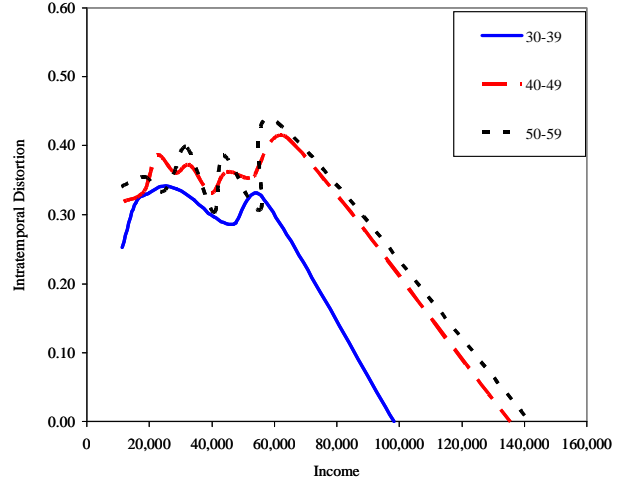
Figure 2c: Intratemporal Distortions in Full Optimum, Baseline Model
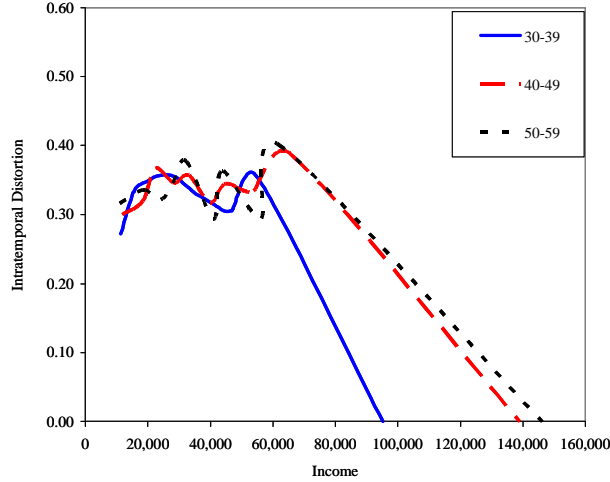
The most striking difference between the scenarios is the treatment of the highest-income young (individuals in the 30-39 age range). The Static Mirrlees policy substantially distorts their intratemporal choice, while the Partial Reform and Full Optimum policies do not.[35] This is the numerical counterpart to Proposition 2, in which we saw that the classic result of no marginal distortion at the top of the income distribution fails to extend to the Static Mirrlees policy within age groups. To repeat the intuition, a distortion allows the planner to collect more tax revenue from higher earners. Distortions are therefore valuable on all except the top earner across ages in the Static Mirrlees policy. But, they bring no benefits for age-dependent policies when levied on the top earners in each age group, because individuals are restricted to age-specific tax schedules.

More generally, the use of intratemporal distortions decreases as the sophistication of policy increases. This is most apparent for policy toward individuals when they are young, where Table 2 shows that distortions

---

[35] The Static Mirrlees distortions differ across age groups even though its taxes are age-independent because the intratemporal distortion $\tau(i,t)$ depends on an individual's wage. Two individuals of different ages with different wages who choose the same income and consumption allocation will have different implied distortions.

are everywhere lower under the Partial Reform policy than under the Static Mirrlees policy. The intuition for this is that a distortion at a given income when individuals are young raises less revenue from higher earners than does the same distortion when they are older, since wage disparities rise with age.[36] Outside this example, the differences between the policies' distortions are less systematic, in that distortions are lower for some individuals at some ages and higher for others as the sophistication of the policy increases. Nevertheless, a telling summary statistic is provided in the final column of Table 2: the unweighted average marginal distortion across all types and ages is 0.324 in the Static Mirrlees policy, 0.309 in the Partial Reform policy, and 0.303 in the Full Optimum policy. If we weight distortions by income, the income-weighted average marginal distortion falls from 0.278 in the Static Mirrlees policy to 0.247 in the Partial Reform policy and 0.241 in the Full Optimum policy. Moreover, these differences do not reflect the improved pattern of distortions in the more sophisticated policies. The combination of lower average distortions and better-designed distortions encourages labor effort under the more sophisticated policies, raising total output and the efficiency of the economy. We will see the welfare impact of these efficiency gains below.

Average taxes Average tax rates are substantially affected by age dependence, as well.[37] Table 3 lists the average tax rates for each scenario.

Table 3: Average tax rates in baseline model simulation

Average tax rate (in percent)						*Lifetime income decile*					
	Age range	Bottom	2	3	4	5	6	7	8	9	Top
	30-39	-110.4	-68.3	-46.1	-33.7	-19.0	-10.1	0.0	4.1	11.4	27.3
Static Mirrlees	40-49	-110.4	-54.9	-34.0	-19.0	-10.1	0.0	4.1	11.4	13.3	31.7
	50-59	-110.4	-54.9	-34.0	-19.0	-10.1	0.0	4.1	11.4	13.3	30.7
	30-39	-141.4	-89.1	-59.0	-40.5	-26.4	-15.1	-6.3	0.6	7.3	24.5
Partial Reform	40-49	-113.3	-53.1	-34.9	-17.0	-7.7	1.5	6.3	13.2	20.0	38.3
	50-59	-126.1	-51.8	-28.1	-17.9	-9.2	2.7	5.5	14.7	17.9	40.7
	30-39	-130.9	-85.3	-56.4	-39.4	-25.1	-16.0	-5.2	-0.4	6.1	16.8
Full Optimum	40-49	-122.6	-56.5	-39.0	-18.1	-8.4	1.3	5.7	12.9	21.7	43.0
	50-59	-138.3	-55.8	-28.3	-19.6	-11.6	4.0	3.3	16.6	17.2	45.8

Figures 3a, 3b, and 3c plot average tax rates against annual income for each policy.

---

[36]In the language of the direct mechanism, when the highest earner among the young faces no marginal distortion, smaller distortions are needed to prevent the highest earner from mimicking lower earners. This chain reaction lowers distortions in general on the young.

[37] An individual's average tax rate is defined as the ratio $\frac{y-c}{c}$.

Figure 3a: Average Tax Rates in Static Mirrlees, Baseline Model
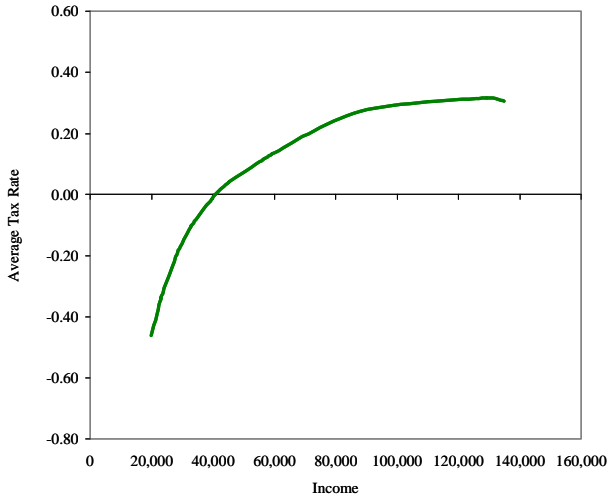
Figure 3b: Average Tax Rates in Partial Reform, Baseline Model
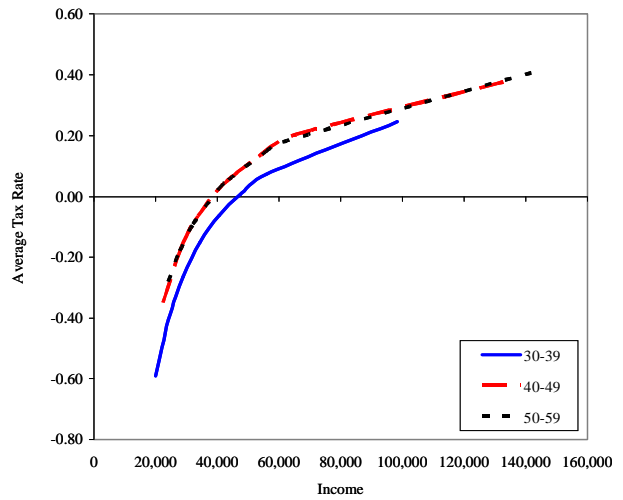
Figure 3c: Average Tax Rates in Full Optimum, Baseline Model
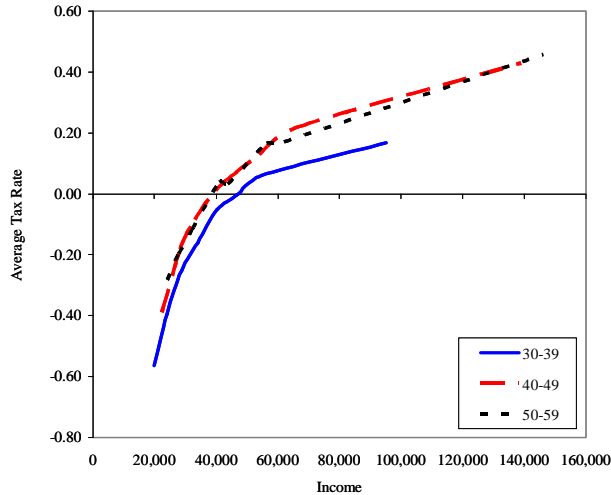
In the Static Mirrlees, the average tax schedule is the same for all age groups. In the Partial Reform and Full Optimum, separate average tax schedules face workers in their thirties, forties, and fifties.

Workers face lower average taxes in their thirties and (to a lesser extent) fifties than in their forties under the more sophisticated policies. The magnitude of the difference across ages can be substantial. For example, in the Partial Reform policy, a middle-income worker earning a constant income over his lifetime faces an average tax rate in his thirties that is more than 7 percentage points lower than in his forties.

Why do the more sophisticated planners use lower average taxes on individuals when they are young? The data show wages rising from the thirties to the forties in all income groups. Individuals want to borrow against future wages to raise consumption when young, but in this baseline economy they cannot transfer resources across periods. Age-dependent tax policy can substitute for private borrowing by lowering average taxes when wages are low: i.e., in workers' thirties.[38] The Static Mirrlees planner cannot do so, because it

---

[38] The simulation results show that age dependence makes consumption *less* smooth for low earners. The reason for this was discussed in the section on intertemporal distortions: when smoothing high-earners' consumption, the planner skews low-earners' consumption in the opposite direction to most efficiently satisfy incentives.

cannot target lower average taxes at an age group.

Intertemporal distortions Next, Table 4 shows the intertemporal distortions under each policy. Here, the main advantage of history dependence is made clear, as only the Full Optimum policy avoids these distortions entirely, providing fully smoothed consumption to all workers.

Table 4: Intertemporal distortions in baseline model simulation											
Data show the ratio: u'(c(t)) / u'(c(t+1))						*Lifetime income decile*					
	Age range	Bottom	2	3	4	5	6	7	8	9	Top
	30-39	1.00	1.05	1.06	1.09	1.07	1.12	1.07	1.15	1.03	1.38
Static Mirrlees	40-49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06
	50-59	…	...	...	...	...	...	...	...	...	...
	30-39	0.93	0.96	0.95	0.97	0.97	0.99	0.97	0.98	1.00	1.12
Partial Reform	40-49	0.99	1.00	1.02	1.00	0.99	1.01	0.99	1.02	0.97	1.01
	50-59	...	...	...	...	...	...	...	...	...	...
	30-39	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Full Optimum	40-49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50-59	...	...	...	...	...	...	...	...	...	...

While not smoothing for each worker, the Partial Reform policy smooths in aggregate across ages, as shown formally in the Symmetric Inverse Euler Equation above, expression (15). This improves on the Static Mirrlees policy, which we can see in Table 4 by noting that the distortions to the intertemporal margin are smaller, on average, in the Partial Reform than in the Static Mirrlees policy.

Welfare gain and decomposition Finally, I quantify and identify the sources of the welfare gain from reform. For each policy, Table 5 lists overall social welfare and lifetime utility by income decile, and Figure 4a plots social welfare.[39]

Table 5: Welfare in baseline model simulation											
	Social Welfare (consumption equivalents*)	Lifetime Utility (consumption equivalents**)									
						Lifetime income decile					
		Bottom	2	3	4	5	6	7	8	9	Top
Static Mirrlees	2.585	2.29	2.36	2.42	2.48	2.53	2.60	2.66	2.75	2.87	3.13
Partial Reform	2.605	2.34	2.40	2.45	2.50	2.56	2.62	2.68	2.76	2.86	3.06
Full Optimum	2.606	2.34	2.40	2.45	2.50	2.56	2.62	2.68	2.76	2.86	3.06

* The value for consumption that, if provided freely to all workers at each age, would generate the same social welfare as the actual allocation.
** The value for consumption that, if provided freely to the worker at each age, would generate the same lifetime utility as the worker obtains with the actual allocation.

Figure 4a: Social Welfare Comparisons, Baseline Model

---
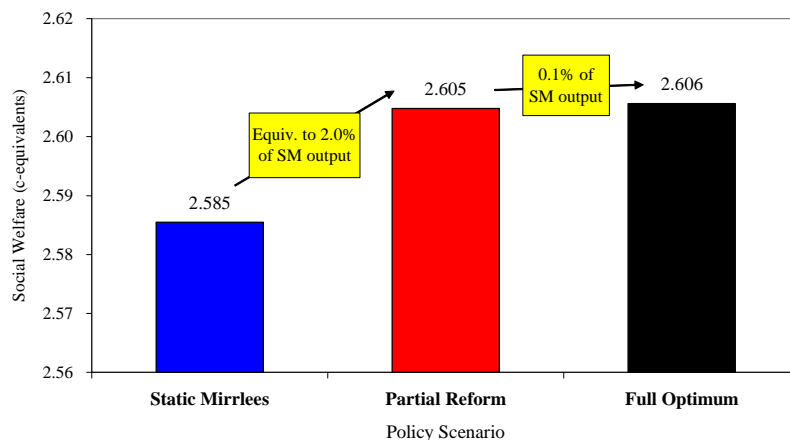
[39] I report social welfare and utility levels in consumption equivalent units. See the notes to Table 5 for specifics.

The main findings are that age dependence generates a large welfare gain in absolute size and that it captures nearly all of the gain from full reform to a more complex, optimal dynamic policy. Moreover, age dependence yields a more equal distribution of utility than is possible under Static Mirrlees policy.

First, the increase in welfare due to age dependence alone is large, equivalent to a 2.0 percent increase in aggregate consumption or roughly $200 billion in current U.S. dollars, annually. Specifically, if the Static Mirrlees planner received a windfall enabling it to increase each individual's consumption by 2.0 percent while holding labor effort fixed, welfare in the Static Mirrlees policy would equal that in the Partial Reform policy. Below, I provide a detailed decomposition of this large welfare gain.

Second, the gain from this Partial Reform captures 96 percent of the gain from reform to the Full Optimum. Specifically, the additional welfare gain from history dependence is small, at 0.1 percent of aggregate consumption, so that the total gain due to reform from the Static Mirrlees to the Full Optimum is 2.1 percent of aggregate consumption. As discussed more below, the advantage of history dependence is minimal in this baseline model because many of the empirical wage paths shown in Figure 1 have similar shapes.

Finally, this welfare gain is particularly pronounced among the low-skilled, so that the Partial Reform and Full Optimum policies achieve more egalitarian (and nearly identical) distributions of lifetime utility across income groups. Table 5 and Figure 4b show the lifetime utility levels of each income group.

Figure 4b: Utility Comparisons, Baseline Model
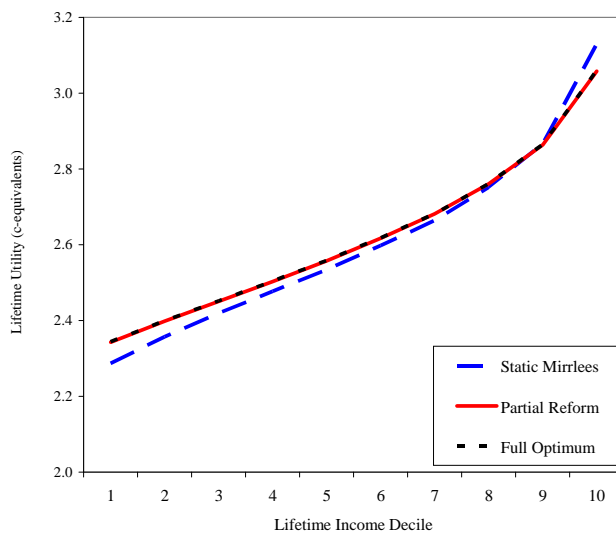
The Static Mirrlees policy provides higher utility to only the top income group, while the more sophisticated policies substantially increase utility among individuals in the bottom half of the income distribution.

What drives this large and equitably-distributed welfare gain from age dependence? Figure 4c shows the results of a welfare gain decomposition that attributes the gain from Partial Reform to improvements in efficiency, equity (due to concave utility of consumption, convex disutility of labor, and the influence of Pareto weights), and consumption-smoothing. I now discuss each of the components in turn.[40]

---

[40] All told, my estimated gains from these components sum to slightly more than the total welfare gain, so that I explain approximately 105 percent of the welfare gain. This overestimate is attributable to gaps between the estimates yielded by the experiments described below, which are necessarily imperfect, and the components' true effects.

Figure 4c: Decomposition of Welfare Gains from Partial Reform
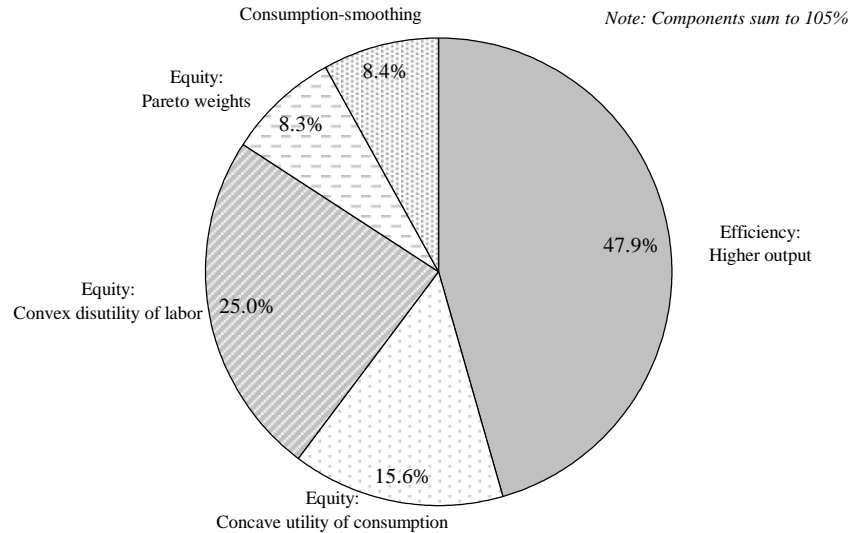
Nearly half of the welfare gain from age dependence is because the economy is more efficient under age-dependent taxes. In Table 2, we saw that the average marginal distortion to labor effort is lowered by adding age dependence to a static Mirrleesian policy. This encourages more effort, so that total output is 2.5 percent higher under the Partial Reform policy than under the Static Mirrlees policy. This output must be produced as well as consumed, however, so its welfare impact is not equal to that of a resource windfall equalling 2.5 percent of income. Instead, I calculate the net benefit of this higher output using a simple thought experiment. Take the Static Mirrlees allocation of income and consumption and suppose that each individual were required to earn and allowed to consume 2.5 percent more at each age. The welfare gain of reform from this modified version of the Static Mirrlees policy to the Partial Reform policy is slightly more than 1.0 percent of aggregate consumption. Thus, the increased output due to efficiency gains account for approximately 48 percent of the total welfare gain from age dependence.

Most of the remaining welfare gain from age dependence is due to a more equitable distribution of resources than under an age-independent policy. In particular, the Partial Reform planner allocates consumption to individuals with higher marginal utilities of consumption and requires production from individuals with lower marginal disutilities of income than does the Static Mirrlees planner. I separately estimate the welfare impacts of each of these two factors.

To estimate the effect of the distribution of consumption, consider an experiment in which each individual's consumption path under the Static Mirrlees policy is scaled to provide the same share of total consumption (in present value) as under the actual Partial Reform policy. This hypothetical allocation replicates the Partial Reform's allocation of consumption across individuals while holding fixed the Static Mirrlees level of total consumption.[41] Specifically, it raises the present value of consumption for the lowest income group by almost 3 percent relative to the actual Static Mirrlees policy, an increase offset by lower present values of consumption for higher income groups. Because utility from consumption is concave, this hypothetical Static Mirrlees policy yields higher welfare than does the actual Static Mirrlees policy, and the consumption-equivalent welfare gain from this hypothetical Static Mirrlees to the Partial Reform is only 1.7

---

[41]Note that it also violates the incentive constraints on the Static Mirrlees problem, which is why the Static Mirrlees planner could not achieve this hypothetical allocation even though it satisfies the feasibility constraint.

percent of Static Mirrlees output. Thus, the distribution of consumption in accordance with the Partial Reform policy accounts for approximately 15 percent of the welfare gain from age dependence (0.3 of the 2.0 percent of total consumption-equivalent gain).

To estimate the impact of allocating required income to those with lower marginal disutilities of labor effort, I consider an analogous experiment to that for consumption. I scale the income required from each individual in the Static Mirrlees to equal (in present value) the same share of total income as in the Partial Reform. Similar calculations to those for consumption imply that the distribution of required income in accordance with the Partial Reform policy accounts for approximately 25 percent of the welfare gain from age dependence.

The welfare impact of these more redistributive allocations of consumption and income is magnified by the assumption that Pareto weights decline as we move up the income distribution. To gauge the importance of this assumption, I calculate social welfare with uniform weights. The consumption-equivalent difference between the Static Mirrlees and Partial Reform scenarios implies that declining Pareto weights account for approximately 8 percent of the gain from age dependence. A nearly identical estimate is obtained when I solve the planner's problems in each policy scenario assuming uniform Pareto weights, i.e., a pure Utilitarian planner.[42]

Finally, age dependence allows for more efficient intertemporal allocations, i.e., more consumption-smoothing, than in the Static Mirrlees. Consider an experiment in which each individual's present value of consumption under the Static Mirrlees policy is allocated across ages as it is in the actual Partial Reform policy. This hypothetical Static Mirrlees policy achieves higher social welfare than the true Static Mirrlees policy, implying that the Partial Reform's increased consumption-smoothing accounts for about 8 percent of the gain from age dependence.

Why do these gains capture nearly all of the gain from reform to the Full Optimum? The Full Optimum's advantage over Partial Reform is history dependence. History dependence is most valuable when wage paths cross or have substantially different slopes, as history-independent policies then have to address incentive problems that vary substantially by age. In contrast, the Full Optimum planner's ability to track individuals allows it to target redistribution and smooth consumption despite differently-sloped wage paths. In the data used for the baseline simulation, as shown in Figure 1, most of the wage paths have similar shapes, thereby reducing the benefit from history dependence. In Section 3, I consider an extension of the baseline model that incorporates crossing wage paths, as guided by the data, by allowing wage paths to be stochastic rather than deterministic. As discussed there, the Partial Reform policy continues to capture nearly all of the gain from the Full Optimum in that extension.

The quantitative simulations have therefore revealed several key lessons about how age-dependent taxes differ from age-independent taxes that, as we will see, are largely robust to the extensions considered in the rest of the paper. First, age dependence lowers the marginal intratemporal distortion on the highest-earning young workers. This is a specific example of the benefits from being able to tailor marginal distortions to the wage distribution at each age and avoid cross-age incentive problems. Second, average taxes are lower for younger workers than older workers, by about seven percentage points across the income distribution in this simulation. In this baseline model, lower average taxes act as a substitute for private consumption-smoothing, a motivation that is absent when individuals can save and borrow privately. In fact, the optimal path of average taxes over the lifecycle will be indeterminate with private saving and borrowing, but that

---

[42] Note that the welfare gains from age dependence do not depend on funding gains for the poor with losses by the rich. In Section 5, I show that age dependence generates only slightly smaller welfare gains when constrained to be Pareto-improving.

indeterminacy allows for a similar pattern of average taxes to be optimal. Third, the welfare gains from age dependence are large, here equal to about two percent of aggregate consumption. These gains are evenly divided between improvements in efficiency and equity. Finally, welfare gains from Partial Reform capture a substantial share (here, nearly all) of the potential gain from a reform to the Full Optimum, and they do so while providing a more egalitarian outcome.

1.5 Summary of baseline model

In this section, I theoretically and numerically characterized age-dependent taxation in a baseline model. I found that age dependence substantially improves on age-independent policy along the two margins that dominate theoretical Mirrleesian tax analysis: the intratemporal consumption-leisure margin and the intertemporal savings margin. Then, I used detailed individual wage data to show the effects of age dependence quantitatively. Age dependence affects the use of marginal distortions as well as the pattern of average taxes, deviating dramatically from the optimal age-independent policy and mimicking key features of the dynamic optimal policy. These improvements generate a large welfare gain, estimated at 2 percent of aggregate annual consumption, and capture nearly all of the gains from reform to the dynamic optimum. In the next several sections, I show that these results are largely robust to extensions of this baseline model to more complicated economic environments.

2 Case 2: Model with private saving and borrowing

In this section, I examine how the results from the baseline model are affected by allowing individuals to save and borrow across ages. Private saving and borrowing generate an important new set of incentive problems for policy, in that individuals may now subsidize consumption with after-tax income earned at a different age.[43] This affects the marginal tradeoffs facing individuals at each age and, therefore, the optimal policy toward them.

Before showing how the three policy scenarios respond to private transfers of resources across periods, it is important to clarify my assumptions on how the three policy scenarios *can* respond. In particular, I need to specify whether capital taxation is available to each policy and what forms it can take. While it is natural to assume that there are no restrictions on capital taxation for the Full Optimum policy, it is less clear what the appropriate assumption is for the Static Mirrlees and Partial Reform policies.

I assume that the Static Mirrlees and Partial Reform planners can neither tax nor subsidize private saving or borrowing in any way. This is a conservative assumption when gauging the power of age dependence, in that it maximizes both the potential for private saving and borrowing to undermine the baseline results and the relative power of the Full Optimum, which has unlimited flexibility in taxing and subsidizing intertemporal transfers. For example, if I allow the Partial Reform and Static Mirrlees policies to include a 15 percent tax rate on capital income (resembling the current U.S. system for capital gains and dividends), the absolute and relative sizes of the welfare gains from Partial Reform increase relative to the results below.

Thus, Partial Reform is defined consistently throughout the paper: it always means only that labor income taxes can depend on age. One interesting extension to this paper's analysis would be to consider age-dependent linear capital taxation, which would increase the potential power of age dependence.[44]

---

[43] A technical note: I assume that savings and debt are observable to the planner. The term "private" indicates private sector, not "hidden," which has a specific meaning in the optimal tax literature.

[44] However, age-dependent linear capital taxes would sacrifice some of the practical advantage of simple age-dependence, and

Now, I follow the structure of Section 1 and specify the social planner's problem in each policy scenario.

2.1 Social planner's problem in three policy scenarios

As in the baseline model, the social planner specifies a menu of bundles to maximize social welfare subject to feasibility and incentive constraints. With private saving and borrowing, however, these bundles are of pre-tax income and after-tax income, not consumption, because after-tax income in a given year may be used by individuals for consumption at any age. Thus, I distinguish between after-tax income, denoted x_t^i for individual i of age t, and consumption, denoted c_t^i as before.

The planner assigns after-tax income to each pre-tax income to maximize social welfare, which is the same as in expression (3) from the baseline model. The feasibility constraint is similar to the baseline model's, though I replace consumption with after-tax income:

$$\sum_{i=1}^{I} \pi^i \sum_{t=1}^{T} R^{T-t} \left(y_t^i - x_t^i \right) = 0 \tag{17}$$

For the incentive constraints, we need notation that reflects the individual's ability to claim a wage different from its true wage at any age, be assigned another individual's allocations, and transfer resources across ages.[45] Let $W_T^{j(s_t)} = \left\{ w_{s_1}^{j_{s_1}}, w_{s_2}^{j_{s_2}}, ..., w_{s_T}^{j_{s_T}} \right\}$ denote a T-period path of wages corresponding to individuals of type j_{s_t} and age s_t, where s_t can vary across t. Thus, $W_T^i = \left\{ w_1^i, w_2^i, ..., w_T^i \right\}$ denotes the true path of wages for individual i. Then, $\left\{ y \left(w_s^{j_{s_t}} \right) \right\}_{t=1}^{T}$ and $\left\{ x \left(w_s^{j_{s_t}} \right) \right\}_{t=1}^{T}$ are the sequence of pre-tax income and after-tax income allocations assigned to an individual who claims the wage sequence $W_T^{j(s_t)}$.

Using this notation, the Static Mirrlees planner in Case 2 solves the following problem:

Problem 4 *(Case 2 Static Mirrlees: Age-Independent)*

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i V^i,$$

subject to the feasibility constraint (17) and the incentive constraints

$$\sum_{t=1}^{T} \beta^{t-1} \left(u \left(c_t \left(W_T^i \right) \right) - v \left(\frac{y \left(w_t^i \right)}{w_t^i} \right) \right) \geq \sum_{t=1}^{T} \beta^{t-1} \left(u \left(c_t \left(W_T^{j(s_t)} \right) \right) - v \left(\frac{y \left(w_{s_t}^{j_{s_t}} \right)}{w_t^i} \right) \right)$$

for all $i, j_{s_t} \in \{1, 2, ..., I\}$ and all $W_T^{j(s_t)} = \left\{ w_{s_1}^{j_{s_1}}, w_{s_2}^{j_{s_2}}, ..., w_{s_3}^{j_{s_3}}, ... w_{s_T}^{j_{s_T}} \right\}$, where

$$\left\{ c_t \left(W_T^{j(s_t)} \right) \right\}_{t=1}^{T} = \arg\max_{\{c_t\}} \left\{ \begin{array}{l} \sum_{t=1}^{T} \beta^{t-1} \left(u \left(c_t \right) - v \left(\frac{y \left(w_{s_t}^{j_{s_t}} \right)}{w_t^i} \right) \right) \\ s.t. \sum_{t=1}^{T} R^{T-t} \left(x \left(w_{s_t}^{j_{s_t}} \right) - c_t \right) = 0 \end{array} \right\}$$

is the consumption path individual i chooses when it claims the sequence of wage levels $W_T^{j(s_t)}$.

---

their effectiveness would likely be seriously undermined by avoidance behavior.

[45] Recall that these planner's problems are structured as direct mechanisms, in which an individual claims (or reports) a wage level and receives an allocations based on that claim.

Though more complicated, these incentive constraints are closely related to those from the baseline model. They reflect that individuals are free to choose any path of after-tax incomes, including those that are intended for individuals of different types and ages, and transfer them across periods using saving and borrowing in order to maximize their lifetime utility. The constraints ensure that each individual prefers its own path of wages W_T^i to any other path $W_T^{j(s_t)}$.[46]

As in the baseline model, the Partial Reform planner has the advantage of conditioning taxes on age. To express its problem, let $W_T^{j(t)} = \left\{ w_1^{j_1}, w_2^{j_2}, ..., w_T^{j_T} \right\}$ denote a path of wages corresponding to individual j_t at each age t. Note the notation j_t rather than j_{s_t} as in the Static Mirrlees policy. This indicates that the Partial Reform planner can restrict an individual to claiming only wages of others of the same age, not all ages. Then, $\left\{ y \left(w_t^{j_t} \right) \right\}_{t=1}^T$ and $\left\{ x \left(w_t^{j_t} \right) \right\}_{t=1}^T$ are the sequence of pre-tax income and after-tax incomes assigned to an individual who claims the wage sequence $W_T^{j(t)}$.

Using this notation, the *Partial Reform* planner in Case 2 solves the following problem:

Problem 5 *(Case 2 Partial Reform: Age-dependent)*

$$\max_{\{x,y\}} \sum_{i=1}^I \pi^i \alpha^i V^i,$$

subject to the feasibility constraint: (17) and the incentive constraints

$$\sum_t \beta^{t-1} \left(u \left(c_t \left(W_T^i \right) \right) - v \left(\frac{y \left(w_t^i \right)}{w_t^i} \right) \right) \geq \sum_t \beta^{t-1} \left(u \left(c_t \left(W_T^{j(t)} \right) \right) - v \left(\frac{y \left(w_t^{j_t} \right)}{w_t^i} \right) \right)$$

for all $i, j_t \in \{1, 2, ..., I\}$ and all $W_T^{j(t)} = \left\{ w_1^{j_1}, w_2^{j_2}, ..., w_T^{j_T} \right\}$, where

$$\left\{ c_t \left(W_T^{j(t)} \right) \right\}_{t=1}^T = \arg\max_{\{c_t\}} \left\{ \begin{array}{c} \sum_t \beta^{t-1} \left(u \left(c_t \right) - v \left(\frac{y \left(w_t^{j_t} \right)}{w_t^i} \right) \right) \\ s.t. \sum_t R^{T-t} \left(x \left(w_t^{j_t} \right) - c_t \right) = 0 \end{array} \right\}$$

is the consumption path individual i chooses when it claims (i.e., reports) the sequence of wage levels $W_T^{j(t)}$.

As in the baseline model, the Partial Reform planner's incentive constraints are easier to satisfy than the Static Mirrlees planner's because each individual must prefer its path of wages W_T^i to only the set of wage paths composed of wages corresponding to individuals of the same age at each age t. As in the baseline, this reflects the Partial Reform planner's ability to restrict individuals to age-specific tax schedules.

Finally, the planner's problem for the Full Optimum scenario is unchanged from the baseline model. Because it can link allocations across an individual's lifetime, the Full Optimum planner spreads the after-tax income received by an individual over its lifetime optimally, leaving the individual's optimal choice undistorted. Recall that this was shown formally in result (13) from the baseline model. Thus, consumption equals after-tax income at each age for each individual in the Full Optimum, so $c_t^i = x_t^i$ for all $i \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$. The individuals' intertemporal optimization problem is irrelevant, and the planner's problem remains one of specifying optimal consumption and income bundles.

---

[46] The summation over t in these incentive constraints does not imply that the Static Mirrlees planner is allowed to make history-dependent allocations. As in the baseline model, an individual's choice of an allocation at age t will have no effect on the planner's allocations to it at age $t + 1$. The summation is needed because the individuals can independently link periods.

Therefore, the *Full Optimum* planner in the Case 2 model solves the following problem:

Problem 6 *(Case 2 Full Optimum: Age-Dependent and History-Dependent)*

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i V(i)$$

subject to the feasibility constraint (4) and the incentive constraints (7).

As in the baseline model, the Full Optimum planner has the benefit of making allocations on a lifetime basis, so it can make promises or threats to individuals that encourage behavior at one age with consequences at another. The other two policy scenarios cannot make these promises or threats because they lack history dependence: that is, they cannot keep track of individuals as they age.

The presence of private saving and borrowing complicates the incentives problems facing the Static Mirrlees and Partial Reform planners. I now turn to showing the effects of these more complicated incentive problems on the characteristics of taxes.

2.2 Analytical results

In the baseline model, I derived theoretical results connected to classic results on the intratemporal and intertemporal margins from the static and dynamic optimal tax literatures. In this model with private saving and borrowing, the second of these margins goes undistorted in all three policy scenarios.[47] Therefore, I focus my analysis in this section on the intratemporal margin.

The impact of private saving and borrowing on intratemporal distortions can best be seen in a simple example. Consider an economy with only two worker types, $i = \{L, H\}$ for low and high skilled, and two ages $t = \{1, 2\}$. Suppose that the high-skilled type H is always higher-skilled than the low-skilled type L, so that $w_1^H > w_1^L$ and $w_2^H > w_2^L$. Finally, assume that the utility function takes the following simple form:

$$U(c,y) = \ln c - \frac{1}{\sigma} \left(\frac{y}{w}\right)^\sigma.$$

The Full Optimum policy's treatment of the high-skilled worker is unchanged by private saving and borrowing. We know from Proposition 2 *(Top Marginal Distortion)* that, in the baseline model's Full Optimum policy, the high-skilled worker in this example would face a zero intratemporal distortion at both ages. The Full Optimum planner's problems are equivalent in Case 2 and the baseline model, so the high-skilled worker also faces zero distortions at both ages in the Full Optimum policy in Case 2.

In the Partial Reform planner's solution to the this two-type example, the expression for the intratemporal distortion on the high-skilled worker when it is young $(t = 1)$ is:

$$\tau^{PR}\left(y_1^H\right) = \frac{\mu^{HL|HH}}{\pi^H \alpha^H + \mu^{LH|HH} + \mu^{LL|HH}} \left(\frac{c_1^{HH}}{c_1^{HL}} - 1\right) \tag{18}$$

where I use $\mu^{ij|kk}$ to denote the Lagrange multiplier on the incentive constraint preventing type k from claiming the series of wages $W_T^{jt} = \left\{w_1^i, w_2^j\right\}$, and all other notation is as in the baseline model.

Intuitively, the distortion in (18) is positive if the young, high-skilled worker is tempted to save some of its after-tax income from the first period and use that savings to raise its consumption while working less

---

and claiming the tax treatment of the low-skilled worker later in life. The Partial Reform planner uses this distortion to raise the marginal utility of consumption for the high-skilled worker when young, discouraging it from the strategy of oversaving and working less later in life. Formally, $\mu^{HL|HH} > 0$ because the incentive constraint preventing the high-skilled worker from claiming the $\{w_1^H, w_2^L\}$ wage path binds, and $\frac{c_1^{HH}}{c_1^{HL}} > 1$ because a high-skilled worker claiming the $\{w_1^H, w_2^L\}$ wage path can subsidize its consumption in the second period with savings from the first. These in turn imply, through (18), a positive intratemporal distortion on the young high-skilled worker. Analogous conditions for the high-skilled old worker and the low-skilled worker, as well as an illustrative numerical example, are provided in the Appendix.

There is an enlightening relationship between this distortion and the well-known Inverse Euler Equation in stochastic dynamic optimal tax models, first noted in Rogerson (1985) and prominently explored by Golosov, Kocherlakota, and Tsyvinski (2003). The Inverse Euler Equation suggests that policy ought to distort the after-tax return to saving (ex post, as shown in Kocherlakota, 2005) in order to counteract the temptation people face in the presence of skill shocks to oversave and falsely claim a low skill level, and thus lower taxes, in the future. The intratemporal distortion above works toward the same goal, even though wages are fully deterministic in this economy. The reason is that the Partial Reform planner, lacking history dependence, must act as if wages are stochastic even when they are not. Though, by assumption, it has no way to tax savings in the Partial Reform, it uses intratemporal distortions to try to achieve the same results.

Unlike the Full Optimum planner in a stochastic economy, the Partial Reform planner in this economy must worry about the possibility of overborrowing as well as oversaving. Because it cannot make future allocations contingent on past behavior, the Partial Reform planner cannot directly discourage individuals from working less when young and subsidizing consumption with borrowing. In fact, given the rising wages paths apparent in the data, overborrowing is a more prominent concern for the planner than is oversaving.

As illustrated in result (18), the use of intratemporal distortions to substitute for intertemporal distortions puts into jeopardy, in principle, the result from the baseline model that age dependence lowers marginal disortions on the highest-earning young workers. To check whether that result holds in Case 2, and to test the robustness of the other main lessons from the baseline model, I turn to numerical simulations.

2.3 Numerical results

For the numerical simulation of this model, I use the same data and approach as in the baseline economy simulation. The only difference in procedure is that computational considerations cause me to simplify the setting by reducing the number of types of individuals to $I = 5$. Therefore, I classify individuals into lifetime income quintiles rather than deciles. Table 6 shows the wage paths for the five income groups, descriptive demographic data, and the calculated Pareto weights that range from 1.00 to 0.87.

Table 6: Data for simulation of Case 2: private saving and borrowing

		Lifetime income quintile				
		Bottom	*2*	*3*	*4*	*Top*
Wage paths	*Age range*					
Average wage by age range ($1999)	30-39	7.92	11.65	15.49	20.10	30.60
	40-49	8.38	12.84	17.11	22.18	37.91
	50-59	8.07	12.96	17.03	22.22	38.59
Descriptive data						
Proportion race=white		0.41	0.55	0.66	0.76	0.85
Proportion with college degree		0.04	0.08	0.15	0.24	0.49
Proportion gender=male		0.57	0.75	0.85	0.90	0.94
Pareto weight (calculated)		1.00	0.98	0.96	0.93	0.87

Source: PSID core sample household heads 1968-2001; author's calculations as described in paper.

To preview the results presented below, marginal distortions on the labor supply of high-income young workers are lowered by age dependence, and Partial Reform yields a large welfare gain that captures a substantial (though somewhat smaller than in the baseline model) share of the potential gain from a fully optimal reform. While lower average taxes on the young remain optimal, average tax rates are no longer uniquely determined, so alternative patterns of average rates are also optimal. Now, I will describe each of these results in more detail.

First, the intratemporal distortions are shown in Table 7 and plotted in Figures 5a, 5b, and 5c.

Table 7: Intratemporal distortions in Case 2 model simulation

Intratemporal distortion		*Lifetime income quintile*					
	Age range	*Bottom*	*2*	*3*	*4*	*Top*	*Simple average*
Static Mirrlees	30-39	0.23	0.27	0.31	0.29	0.30	
	40-49	0.35	0.41	0.40	0.41	0.11	**0.316**
	50-59	0.27	0.43	0.39	0.42	0.16	
Partial Reform	30-39	0.26	0.32	0.31	0.31	0.11	
	40-49	0.27	0.32	0.31	0.35	0.11	**0.272**
	50-59	0.29	0.31	0.32	0.35	0.14	
Full Optimum	30-39	0.27	0.31	0.30	0.31	0.00	
	40-49	0.28	0.31	0.30	0.34	0.00	**0.242**
	50-59	0.29	0.30	0.30	0.34	0.00	

Figure 5a: Intratemporal Distortions in Static Mirrlees, Case 2
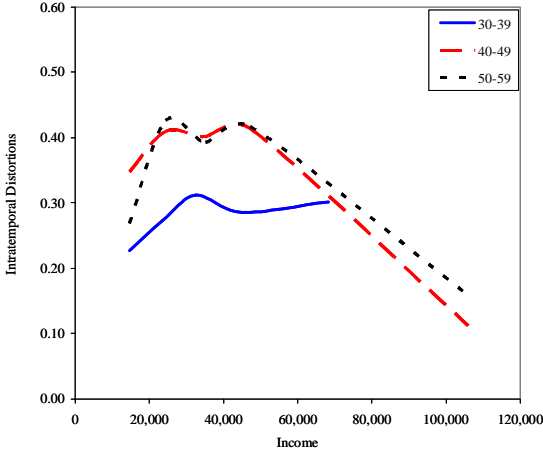
Figure 5b: Intratemporal Distortions in Partial Reform, Case 2
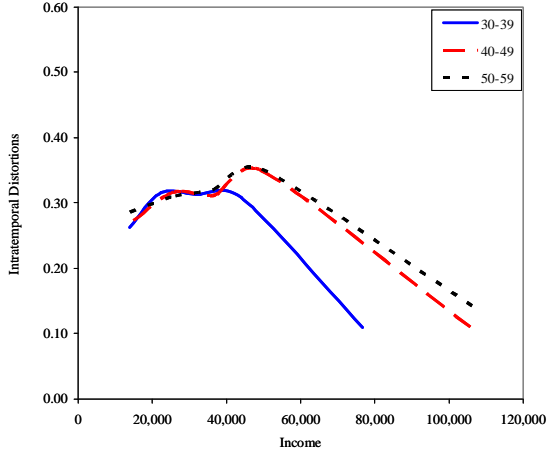
Figure 5a: Intratemporal Distortions in Full Optimum, Case 2
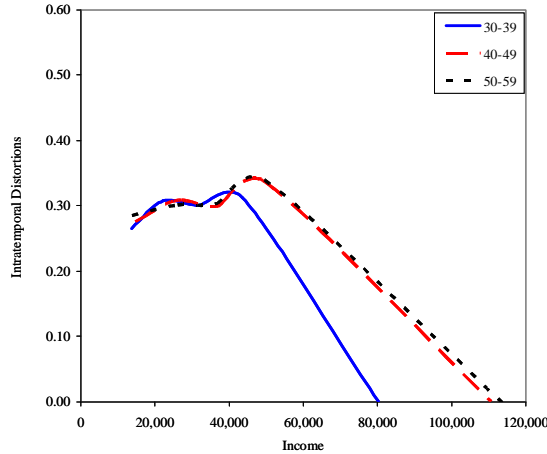
As in the baseline model, the high-income young continue to face larger distortions to labor supply in the age-independent Static Mirrlees than in the more sophisticated policies. This is in spite of result (18), which showed that the Partial Reform policy uses intratemporal distortions to discourage deviation strategies in which individuals save or borrow to supplement consumption while working less to claim a more generous tax treatment at another age. Thus, the Partial Reform policy's ability to identify the position of top earners within their age group, which we learned in the baseline model reduces the optimal distortion on them, overwhelms the incentive problems raised by private saving and borrowing. Meanwhile, the Full Optimum planner can avoid such incentive problems directly by taxing intertemporal transfers, so it has no need to distort the consumption-leisure choice of the top earner in each age group, just as in the baseline economy.

Importantly, the use of intratemporal distortions in general decreases as the sophistication of policy increases, just as in the baseline model. The final column of Table 2 shows that the unweighted average marginal distortion is 0.316 in the Static Mirrlees policy, 0.272 in the Partial Reform policy, and 0.242 in the Full Optimum policy. The difference is greater when we weight distortions by income: the income-weighted average marginal distortion falls from 0.279 in the Static Mirrlees policy to 0.232 in the Partial Reform policy and 0.169 in the Full Optimum policy. As in the baseline, the combination of lower average distortions and a better-designed pattern of distortions encourages labor effort under the more sophisticated policies, raising

total output and the efficiency of the economy. We will see the welfare impacts of these efficiency gains below.

Optimal average taxes in this model are, in contrast to the baseline model, indeterminate in the Partial Reform policy. As individuals can freely transfer resources across ages, any pattern of average taxes under this policies can be replaced with another that transfers resources lump-sum from one age to another without affecting any individual's choices, therefore not affecting aggregate welfare, the allocation's feasibility, or incentive constraints.

This implies that average taxes may, without loss of generality, take a form that resembles that from the baseline model. For instance, Table 8 shows one set of optimal average tax rates in the Partial Reform model alongside the optimal rates for the Static Mirrlees and Full Optimum policies. The average tax rate for individual i of age t is $\frac{y_t^i - x_t^i}{x_t^i}$.

Table 8: Average tax rates in Case 2 model simulation						
Average tax rate (in percent)			*Lifetime income quintile*			
	Age range	*Bottom*	*2*	*3*	*4*	*Top*
Static Mirrlees	30-39	-87.3	-34.4	-11.7	3.9	19.9
	40-49	-87.3	-31.3	-7.4	5.9	27.9
	50-59	-87.3	-31.3	-7.4	5.9	27.9
Partial Reform	30-39	-147.4	-71.0	-34.1	-11.1	13.9
	40-49	-54.0	-7.6	10.3	20.4	37.9
	50-59	-70.8	-10.6	6.5	18.2	37.3
Full Optimum	30-39	-117.8	-52.9	-21.8	-2.5	15.6
	40-49	-101.5	-32.3	-4.9	10.0	38.8
	50-59	-114.6	-29.7	-5.8	10.0	40.3

These example average tax rates are plotted against lifetime income in Figures 6a, 6b, and 6c.

Figure 6a: Average Tax Rates in Static Mirrlees, Case 2
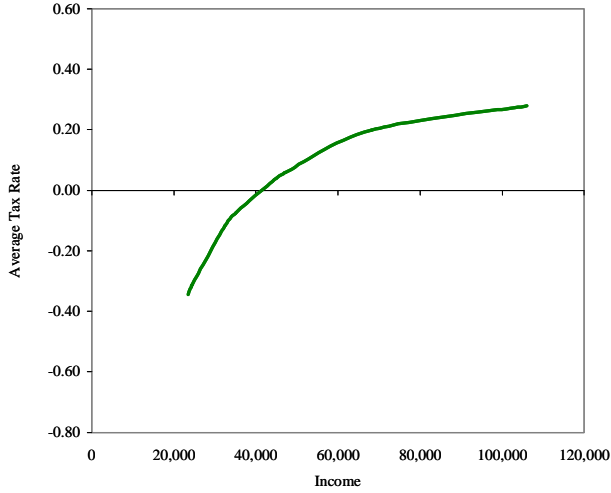
Figure 6b: Average Tax Rates in Partial Reform, Case 2
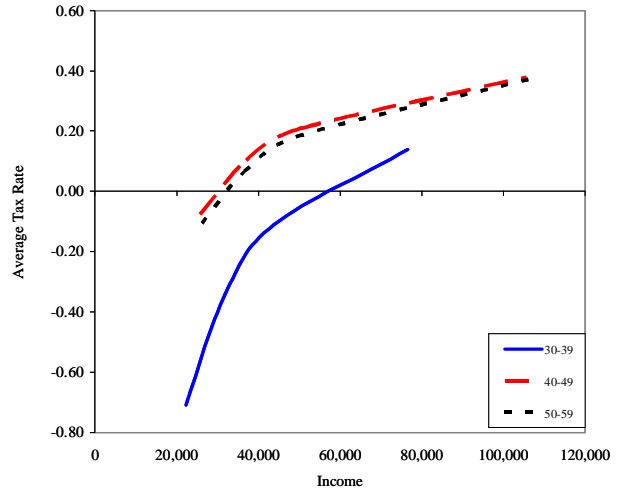
Figure 6c: Average Tax Rates in Full Optimum, Case 2
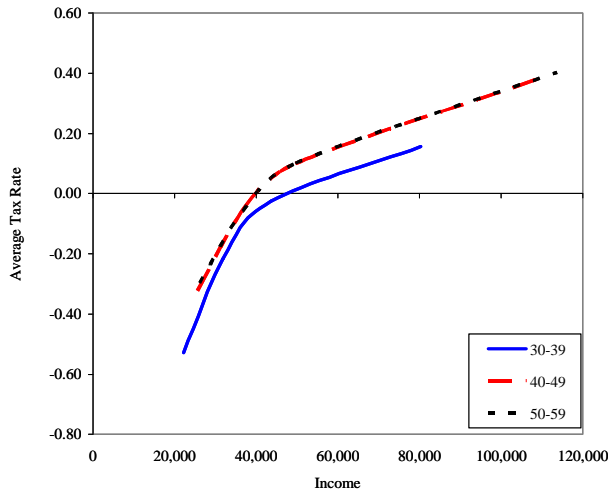
It is important to emphasize, however, that the Partial Reform schedules in Figure 6b are not the unique optimal average tax rate schedules. To see why, consider transferring one unit of after-tax income from each type of worker in its thirties to each type of worker in its forties under the Partial Reform policy. As individuals can transfer resources across time using the same technology as the planner, the paths of consumption corresponding to each path of labor effort are unchanged, so each worker's choices are unaffected by this transfer. Thus, a very different pattern of optimal taxes, such as with higher average taxes on the young than on the old, would also be optimal in this setting.[48]

Next, the Partial Reform continues to capture a large absolute welfare gain and a substantial share of the potential gains from more comprehensive reform. Table 9 shows social welfare and each income quintile's lifetime utility under the three policies.

---

[48] I am grateful to Ivan Werning for pointing out this result.

Table 9: Welfare in Case 2 model simulation

	Social Welfare (consumption equivalents*)	Lifetime Utility (consumption equivalents**)				
		Lifetime income quintile				
		Bottom	2	3	4	5
Static Mirrlees	2.60	2.35	2.46	2.57	2.71	3.00
Partial Reform	2.62	2.41	2.50	2.59	2.71	2.96
Full Optimum	2.63	2.44	2.52	2.61	2.72	2.91

* The value for consumption that, if provided freely to all workers at each age, would generate the same social welfare as the actual allocation.

** The value for consumption that, if provided freely to the worker at each age, would generate the same lifetime utility as the worker obtains with the actual allocation.

Figure 7a plots social welfare under the three policies.

Figure 7a: Social Welfare Comparisons, Case 2
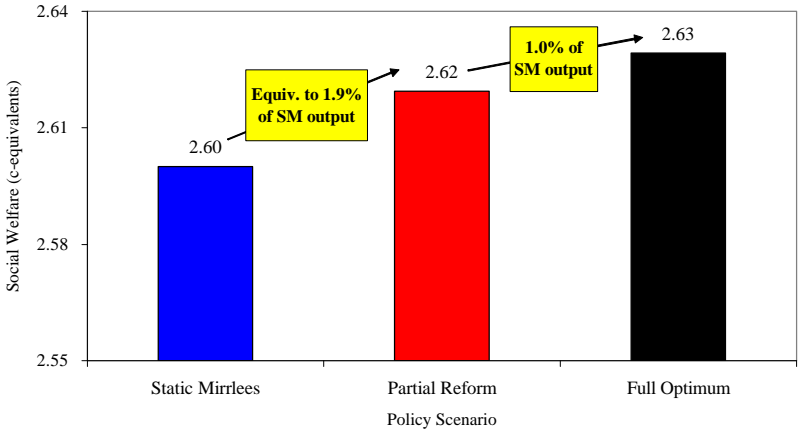
Reform from the Static Mirrlees policy to the Partial Reform policy yields a gain of 1.9 percent of aggregate consumption, nearly the same gain as in the baseline model. Because the Full Optimum planner is better able to respond to the new incentive problems introduced by private saving and borrowing, this gain makes up somewhat less of the gain from full reform. Nevertheless, it captures a substantial share, 67 percent, of that potential gain. More sophisticated capital taxation would magnify the power of age dependence. For instance, in an extension with a uniform 15 percent tax on capital income (not shown), the welfare gain from Partial Reform rises to 2.2 percent of aggregate consumption, comprising 72 percent of the potential gain from the Full Optimum.

Finally, as in the baseline model, the Partial Reform's higher overall social welfare is also shared more equally among the individuals in the population. Table 9 and Figure 7b show the lifetime utility for each income group when individuals can save and borrow.

Figure 7b: Utility Comparisons, Case 2
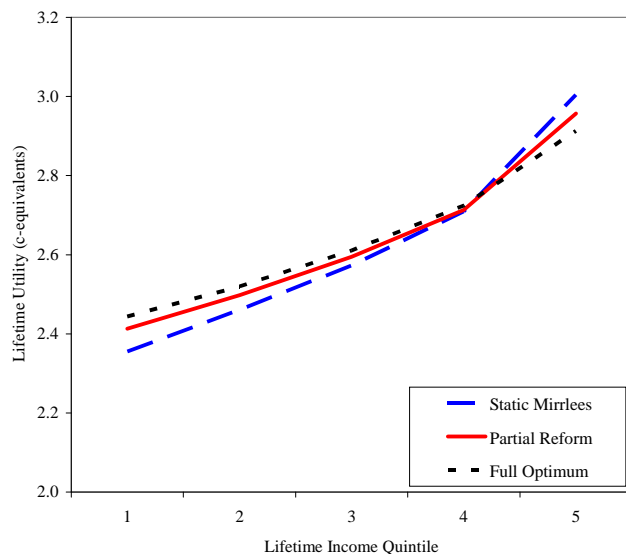
The Partial Reform policy produces a more egalitarian distribution of utility than does the Static Mirrlees, though not as egalitarian as the Full Optimum.

Thus, the results from the baseline model are largely robust to the addition of private saving and borrowing. Even when individuals are free to transfer after-tax income across periods, age dependence is a powerful reform. In the next section, I test the robustness of this result to relaxing a second assumption in the baseline model.

3 Case 3: Model with stochastic wage paths

In this section, I return to a setting in which individuals cannot transfer resources across time, but I explore a new variation on the baseline model by modeling wage paths as stochastic rather than deterministic. Stochastic wage paths generalize the baseline model in two important ways. First, they mean that individuals and the planner are uncertain about their future wages. As shown in Rogerson (1985) and Golosov, Kocherlakota, and Tsyvinski (2003), this uncertainty affects individuals' labor supply and saving and borrowing behavior, with important implications for dynamic optimal policy. Second, they allow for substantially more heterogeneity in wage paths. In the baseline model, individuals were assigned to types by lifetime income-earning potential, and each type was assumed to have a single representative wage path. In this section and the next, such types play no role, and two individuals with similar lifetime incomes may have very different wage paths.

The extent to which individuals' wage paths are determined over time due to stochastic shocks, rather than at the start of their working lives, is the subject of substantial recent research. Keane and Wolpin (1997) and Storesletten, Telmer, and Yaron (2001), provide evidence that stochastic shocks account for as little as 10 percent and as much as 40 percent of total variation in wage paths, respectively. Guvenen (2007) finds evidence that individuals undergo substantial learning over time about the shape of their wage path. The larger the role of uncertainty in wage paths, the more important it is to understand how stochasticity affects the power of age dependence.

I model stochastic wages as a simple Markov process. At each age of working life, individuals are distributed among an age-specific set of discrete wage levels. A separate transition matrix links each age's wage distribution to the next, so that a transition matrix between ages t and $t+1$ determines the distribution of all individuals with a given wage at age t among the set of wage levels at age $t+1$. This simple Markov approach yields a transparent and computationally tractable representation of the dynamic uncertainty and heterogeneity in wage paths from the data.

As with the baseline model, I begin with a theoretical analysis of the policy scenarios as social planners' problems. The key result is Proposition 5 (Baseline and Case 3 Equivalence), which shows that the Static Mirrlees and Partial Reform planners' problems in this Case 3 model are identical to their problems in an appropriately-specified baseline model from Section 1. Then, I use numerical simulations to show that the quantitative results from the baseline model carry through to this model with stochastic wage paths.

3.1 Social planner's problem in three policy scenarios

As in the baseline model, I work with social planners' problems in three policy scenarios: the Static Mirrlees, Partial Reform, and Full Optimum.

I begin by defining some notation. Denote an individual's true path of wages as $W_T^{i(t)} = \left\{ w_1^{i_1}, w_2^{i_2}, ... w_t^{i_t}, ... w_T^{i_T} \right\}$ and let $\pi^{i(t)}$ denote the population proportion represented by this individual. Using these probabilities, let $\pi_t^j = \sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)}$ denote the probability of wage level w_t^j at age t. It is the sum of the population proportions of the individuals whose wage paths equal w_t^j at age t. Note that $\sum_{j=1}^{I} \pi_t^j = 1$ for all t. Denote the transition matrix between ages t and $t+1$ as $P_{t,t+1}$, whose element (m,n) is:

$$P_{t,t+1}(m,n) = \Pr(w_{t+1}^n | w_t^m).$$

In words, $P_{t,t+1}(m,n)$ is the probability that an individual with wage w_t^m will have wage w_{t+1}^n. Thus, the population proportion of an individual with wage path $W_T^{i(t)}$ can also be written $\pi^{i(t)} = \prod_{t=1}^{T-1} \pi_1^i P_{t,t+1}(i_t, i_{t+1})$.

The structure of each planner's problem is the same as in the previous sections. To maximize social welfare, each planner offers a menu of income and consumption pairs to individuals. In the Static Mirrlees and Partial Reform policies, the planner offers a pair $\left\{ c_t^j, y_t^j \right\}$ as the consumption and income intended to be chosen by an individual with wage w_t^j at age t. In the Full Optimum policy, the allocations can be history-dependent.[49]

Social welfare depends on the Pareto weights assigned to individuals with different wage paths. To determine these welfare weights, the planner uses its (complete) knowledge of the stochastic structure of wages. It calculates all possible lifetime income-earning potentials as determined by the truthful wage paths $W_T^{i(t)}$ and assign Pareto weights $\alpha\left(W_T^{i(t)}\right)$ to each of them, just as in the baseline model. Thus, $\alpha\left(W_T^{i(t)}\right)$ indicates a scalar Pareto weight on the individual with the wage path $W_T^{i(t)}$. Using these Pareto weights, define the term

$$\alpha_t^j = \frac{\sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)} \alpha\left(W_T^{i(t)}\right)}{\sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)}}, \tag{19}$$

---

[49] The Static Mirrlees and Partial Reform allocations may, in principle, be different for two individuals with different histories but the same current wage, in that these individuals could choose different (c, y) pairs. In this Case 3 model, however, individual decisions depend only on their current wage, so this possibility is irrelevant.

as the expected Pareto weight on an individual of age t with wage j. Expression (19) is the probability-weighted average of the Pareto weights on individuals with wage paths that include w_t^j when they are of age t. For instance, individuals with the first-period wage w_1^j will go on to have a variety of wage paths. The weight α_1^j captures the probability-weighted average of their eventual Pareto weights.

Using this notation, I now state the planner's problems in the Static Mirrlees and Partial Reform policy scenarios. After discussing an important result on these policies, I then state the Full Optimum planner's problem.

The Static Mirrlees planner in the Case 3 model solves the following problem:

Problem 7 *(Case 3 Static Mirrlees: Age-Independent)*

$$\max_{\{c,y\}} \sum_{j=1}^{I} \sum_{t=1}^{T} \beta^{t-1} \pi_t^j \alpha_t^j \left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^j}\right) \right) \tag{20}$$

subject to feasibility

$$\sum_{j=1}^{I} \sum_{t=1}^{T} R^{T-t} \pi_t^j \left(y_t^j - c_t^j\right) = 0. \tag{21}$$

and incentive constraints:

$$\beta^{t-1} \left(u\left(c_t^{i_t}\right) - v\left(\frac{y_t^{i_t}}{w_t^{i_t}}\right) \right) \geq \beta^{t-1} \left(u\left(c_s^k\right) - v\left(\frac{y_s^k}{w_t^{i_t}}\right) \right). \tag{22}$$

for all $i_t, k \in \{1, 2, ..., I\}$ and $t, s \in \{1, 2, ..., T\}$.

As in the baseline model, the Static Mirrlees incentive constraints reflect that each individual must prefer the allocation intended for its wage level and age to that intended for any other wage level of any age. Formally, this is captured by the use of $\left(c_t^{i_t}, y_t^{i_t}\right)$ on the left-hand side of (22) and $\left(c_s^k, y_s^k\right)$ on the right-hand side.

Next, the Partial Reform planner in the Case 3 model solves the following problem:

Problem 8 *(Case 3 Partial Reform: Age-Dependent)*

$$\max_{\{c,y\}} \sum_{j=1}^{I} \sum_{t=1}^{T} \beta^{t-1} \pi_t^j \alpha_t^j \left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^j}\right) \right) \tag{23}$$

subject to feasibility (21) and incentive constraints:

$$\beta^{t-1} \left(u\left(c_t^{i_t}\right) - v\left(\frac{y_t^{i_t}}{w_t^{i_t}}\right) \right) \geq \beta^{t-1} \left(u\left(c_t^k\right) - v\left(\frac{y_t^k}{w_t^{i_t}}\right) \right). \tag{24}$$

for all $i_t, k \in \{1, 2, ..., I\}$ and $t \in \{1, 2, ..., T\}$.

As in the baseline model, the Partial Reform planner's incentive constraints are a subset of the Static Mirrlees planner's and are, therefore, easier to satisfy. The Partial Reform planner need only satisfy incentives within age groups. Formally, the right-hand side of (24) depends on $\left(c_t^k, y_t^k\right)$ rather than $\left(c_s^k, y_s^k\right)$ as in the Static Mirrlees problem.

With these statements of the Static Mirrlees and Partial Reform planner's problems in Case 3, I now establish an equivalence between these planners' problems and those in an appropriately-chosen baseline economy with deterministic wage paths. The following proposition gives the result:

Proposition 5 (Baseline and Case 3 Equivalence) *Consider a stochastic wage path $W_T^{i(t)}$ and a deterministic wage path W_T^j. If, for each $i(t)$, there exists a j such that $\pi^{i(t)} = \pi^j$ and $w_t^{i_t} = w_t^j$ for all t, where $w_t^{i_t} \in W_T^{i(t)}$ and $w_t^j \in W_T^j$, then:*

- *the solution to the Static Mirrlees planner's problem in the baseline model is also the solution to the Case 3 Static Mirrlees planner's problem,*

- *the solution to the Partial Reform planner's problem in the baseline model is also the solution to the Case 3 Partial Reform planner's problem.*

Proof. In Appendix. ■

This proposition considers a deterministic economy that has the same set of individual wage paths as the stochastic economy will have, ex post. Its implications hold because, in each of these policy scenarios, the objective function, feasibility constraint, and incentive constraints are the same with stochastic or deterministic wage paths. Thus, the same distributions of c and y solve the planner's problems in each model.

The key to this result is that adding stochasticity in wages fails to change the problem for either the planner or the individuals in the Static Mirrlees and Partial Reform scenarios. For these planners, being restricted to history-independence plays the same role as wage stochasticity. When a planner cannot track individuals across ages, it must satisfy incentives age-by-age, rather than over a lifetime, so it is already setting policy as if wages were stochastic. For individuals, the addition of stochasticity has no effect on their incentives relative to the baseline model because, in the Case 3 model, individuals cannot transfer resources between periods and utility is separable across periods. Each age involves an isolated optimization for the individual, so the stochasticity of wages is irrelevant.

In contrast, the Full Optimum planner's problem is substantially affected by stochasticity. Stochastic wages multiply the number of possible wage paths, each of which is assigned a history-dependent allocation at each age. In particular, let $\left\{ c_t^i\left(W_{t-1}^{j(t)}\right), y_t^i\left(W_{t-1}^{j(t)}\right)\right\}$ be the allocation of consumption and pre-tax income intended for an individual of age t who has reported the (possibly false) wage path $W_{t-1}^{j(t)} = \left\{w_1^{j_1}, w_2^{j_2}, \ldots w_{t-1}^{j_{t-1}}\right\}$ and who reports the current wage w_t^i. The Full Optimum planner's incentive constraints must guarantee that individuals would rather reveal their true wage path $W_T^{i(t)}$, *age by age*, rather than any other path, taking into account that individuals know the true transition matrices.

The Full Optimum planner in the Case 3 model solves the following problem:

Problem 9 *(Case 3 Full Optimum: Age-Dependent and History-Dependent)*

$$\max_{\{c,y\}} \left\{ \sum_{i(t)} \pi^{i(t)} \alpha\left(W_T^{i(t)}\right) \sum_t \beta^{t-1} \left(u\left(c_t^{i_t}\left(W_{t-1}^{i(t)}\right)\right) - v\left(\frac{y_t^{i_t}\left(W_{t-1}^{i(t)}\right)}{w_t^{i_t}}\right)\right)\right\},$$

subject to feasibility

$$\sum_{i(t)} \pi^{i(t)} \sum_{t=1}^T R^{T-t}\left(y_t^{i_t}\left(W_{t-1}^{i(t)}\right) - c_t^{i_t}\left(W_{t-1}^{i(t)}\right)\right) = 0.$$

and incentive constraints, which are defined recursively. First, for the last working age, T, and for all i, j :

$$u\left(c_T^{i_T}\left(W_{T-1}^{j(t)}\right)\right) - v\left(\frac{y_T^{i_T}\left(W_{T-1}^{j(t)}\right)}{w_T^{i_T}}\right) \geq u\left(c_T^j\left(W_{T-1}^{j(t)}\right)\right) - v\left(\frac{y_T^j\left(W_{T-1}^{j(t)}\right)}{w_T^{i_T}}\right)$$

Next, for all i, j and $t < T$:

$$\left[U\left(W_{t-1}^{j(t)}, w_t^{i_t}\right) + \beta \sum_{i_{t+1}} P_{t,t+1}\left(i, i_{t+1}\right) U\left(W_{t-1}^{j(t)}, w_t^{i_t}, w_{t+1}^{i_{t+1}}\right)\right.$$

$$\left. \geq U\left(W_{t-1}^{j(t)}, w_t^j\right) + \beta \sum_{i_{t+1}} P_{t,t+1}\left(i, i_{t+1}\right) U\left(W_{t-1}^{j(t)}, w_t^j, w_{t+1}^{i_{t+1}}\right)\right]$$

where $U\left(W_{t-1}^{j(t)}, w_t^{i_t}\right)$ is the period utility at age t of an individual reporting the sequence of wages defined by $\left(W_{t-1}^{j(t)}, w_t^{i_t}\right)$, so

$$U\left(W_{t-1}^{j(t)}, w_t^{i_t}\right) = u\left(c_t^{i_t}\left(W_{t-1}^{j(t)}\right)\right) - v\left(\frac{y_t^{i_t}\left(W_{t-1}^{j(t)}\right)}{w_t^{i_t}}\right)$$

$$U\left(W_{t-1}^{j(t)}, w_t^{i_t}, w_{t+1}^{i_{t+1}}\right) = u\left(c_{t+1}^{i_{t+1}}\left(W_{t-1}^{j(t)}, w_t^{i_t}\right)\right) - v\left(\frac{y_{t+1}^{i_{t+1}}\left(W_{t-1}^{j(t)}, w_t^{i_t}\right)}{w_{t+1}^{i_{t+1}}}\right)$$

This planner's problem is the stochastic analogue of the Full Optimum problems in the baseline and Case 2 models. Intuitively, the incentive constraints have two components. First, they ensure that, no matter the previous path of wage claims, individuals want to reveal their true wage in the last period of working life, age T. Second, they ensure that truth-telling is optimal at each age t prior to the final working period, no matter the previous path of claims, given that truth-telling is optimal at the next age $t+1$. These two steps guarantee that truthtelling is optimal at all ages for all individuals.

Now, I work with the planner's problems specified above to characterize policy analytically and numerically.

3.2 Analytical results

Proposition 5 (Baseline and Case 3 Equivalence) showed that the Static Mirrlees and Partial Reform planners' problems in a stochastic-wage economy with no private saving or borrowing are equivalent to their problems in an appropriately-chosen deterministic economy. As such, the analysis of their allocations is a straightforward extension of the analysis of the baseline model. The Full Optimum planner's problem does not yield simple conditions on optimal intratemporal distortions, but it does yield an important result on the intertemporal margin that I compare to the less sophisticated models.

3.2.1 Intratemporal distortions

The Partial Reform and Static Mirrlees intratemporal distortions are directly related to those in the baseline model. Using the notation defined above and, to make the results cleaner, the assumption that the disutility of labor takes the isoelastic form of expression (9), the Static Mirrlees policy's intratemporal distortion on

an individual with wage $w_t^{i_t}$ is:

$$\tau^{SM}(i_t, t) = \frac{\sum_{s=1}^{T} \sum_{j=1}^{I} \left(1 - \left(\frac{w_t^{i_t}}{w_s^j}\right)^{\sigma}\right) \beta^{s-t} \mu_{t|s}^{i_t|j}}{\sum_{j(t):w_t^{j_t}=w_t^{i_t}} \pi^{j(t)} \alpha\left(W_T^{j(t)}\right) + \sum_{s=1}^{T} \sum_{j=1}^{I} \mu_{s|t}^{j|i_t} - \sum_{s=1}^{T} \sum_{j=1}^{I} \left(\frac{w_t^{i_t}}{w_s^j}\right)^{\sigma} \beta^{s-t} \mu_{t|s}^{i_t|j}}.$$

where $\mu_{t|s}^{i|j}$ is the multiplier on the incentive constraint preventing an individual of age t with wage $w_t^{i_t}$ from claiming the wage w_s^j of an individual of age s, and all other notation is as specified above. The Partial Reform policy's intratemporal distortion on an individual with wage $w_t^{i_t}$ is:

$$\tau^{PR}(i_t, t) = \frac{\sum_{j=1}^{I} \left(1 - \left(\frac{w_t^{i_t}}{w_t^{j_t}}\right)^{\sigma}\right) \mu_t^{i_t|j}}{\sum_{j(t):w_t^{j_t}=w_t^{i_t}} \pi^{j(t)} \alpha\left(W_T^{j(t)}\right) + \sum_{j=1}^{I} \mu_t^{j|i_t} - \sum_{j=1}^{I} \left(\frac{w_t^{i_t}}{w_t^{j_t}}\right)^{\sigma} \mu_t^{i_t|j}}.$$

where $\mu_t^{i|j}$ is the multiplier on the incentive constraint preventing an individual of age t with wage $w_t^{i_t}$ from claiming the wage w_t^j of an individual of the same age, and all other notation is as specified above.

These expressions are identical[50] to the baseline model results (10) and (11), so wage stochasticity has little effect on the theoretical characterization of intratemporal distortions in these policies. The Partial Reform planner retains its advantage over the Static Mirrlees planner in limiting individuals to age-specific tax schedules.

Next, I characterize distortions to the intertemporal consumption margin.

3.2.2 Intertemporal distortions

Unlike in the baseline economy, the Full Optimum policy distorts the intertemporal margin in this model. Echoing the well-known result shown by Rogerson (1985) and Golosov, Kocherlakota, and Tsyvinski (2003), the Full Optimum allocation is described by an Inverse Euler Equation. For individual i of age t, this expression is:

$$\frac{1}{u'\left(c_t^i\left(W_{t-1}^i\right)\right)} = \sum_{j=1}^{I} \frac{P_{t,t+1}(i,j)}{u'\left(c_{t+1}^j\left(W_t^i\right)\right)}, \tag{25}$$

The Inverse Euler Equation sets the consumption cost to the planner of providing marginal utility for an individual at age t equal to the expected consumption cost of providing marginal utility for the same individual at age $t+1$. If wage paths were deterministic, the expectation would be degenerate and the allocations would satisfy the standard Euler equation. With stochastic wage paths, the most efficient way for the planner to satisfy incentives is to allocate consumption according to (25).

The Partial Reform policy fails to satisfy the Inverse Euler Equation because it lacks history dependence. Rather, its satisfies what I called in Section 1 the Symmetric Inverse Euler Equation:

$$\sum_{i=1}^{I} \frac{\pi_t^i}{u'\left(c_t^i\right)} = \sum_{i=1}^{I} \frac{\pi_{t+1}^i}{u'\left(c_{t+1}^i\right)}, \tag{26}$$

---

[50] The Pareto weight terms seem to differ, but they are the expected weights given the currently-observed wage. A deterministic model with the same set of ex post wage paths as the stochastic model would have the same expected Pareto weights given a currently-observed wage.

for any $t, t+1 \in \{1, 2, ..., T\}$. Recall that this condition means the planner has equalized across ages the consumption cost of increasing welfare. As with the intratemporal distortion, stochasticity does not affect the intertemporal distortions in the Partial Reform (or Static Mirrlees) policies because, for these planners, being restricted to history-independence plays the same role as wage stochasticity.

As in the baseline model, the efficiency of intertemporal allocations in Case 3 increases with the policy's sophistication. The Full Optimum planner's ability to satisfy the Inverse Euler Equation (25) is an advantage over the Partial Reform planner, while the latter's ability to satisfy the Symmetric Inverse Euler Equation improves on the Static Mirrlees policy. The Static Mirrlees planner cannot target resources to specific ages, so it cannot equalize the cost of raising social welfare across ages.

Now, I turn to numerical simulations to test the robustness of the baseline model's quantitative results.

3.3 Numerical results

In this section, I simulate the Case 3 planners' problems. First, I discuss the construction of the required data and the parameter specification. Then, I describe the results of the simulations.

3.3.1 Data and Parameters

As stated earlier, I model stochastic wage paths as a simple Markov process in which individuals are distributed among age-specific sets of discrete wage levels and move between these wage levels over time according to transition matrices linking each age to the next. As with the Case 2 model, computational considerations cause me to limit the size of the simulation. I set the number of wage levels at each age to $I = 4$ and continue to use $T = 3$ to represent the three decades of working life.

The simulations require a wage distribution for each age and transition matrices between ages. For the wage distributions, I use the data on household heads from the PSID core sample as described in Section 1, though I restrict the sample to individuals who were observed at least twice in each age range (i.e., 30-39, 40-49, and 50-59 years of age). This eliminates approximately 23,000 observations representing 2,000 sparsely-observed individuals, leaving us with 120,000 observations representing over 9,500 individuals with an average of over twelve observations each. With this sample I calculate the 5th, 35th, 65th, and 95th percentile wages within each age range and use these as the four wage levels among which individuals stochastically move. These wage levels are shown in Table 10.

Transition matrices are calculated from the data as follows. I assign percentile rankings to individual wages at each age, and for each age range (i.e., 30-39, 40-49, 50-59), I average each individual's percentile rank for the observed years. Within each age range, I then sort individuals according to this average rank and group them into wage quartiles. This assigns each individual to a wage quartile in each age range, allowing me to calculate empirical transition probabilities that populate the transition matrices in Table 10.

Table 10: Data for simulation of Case 3 (stochastic wage paths)

Wage levels		*Percentile of wage distribution*			
	Age range	*5th*	*35th*	*65th*	*95th*
Wage for specified	30-39	5.46	12.33	19.20	35.95
percentile and age	40-49	5.60	13.18	21.44	43.41
range ($1999)	50-59	5.02	12.13	20.45	43.88
	Initial pbb	0.25	0.30	0.27	0.18

Transition matrices		*Wage quartile in 40-49 age range*			
		Bottom	*2*	*3*	*Top*
	Bottom	0.71	0.26	0.02	0.00
Wage quartile in 30-	*2*	0.18	0.56	0.23	0.02
39 age range	*3*	0.03	0.22	0.56	0.19
	Top	0.01	0.03	0.19	0.77

		Wage quartile in 50-59 age range			
		Bottom	*2*	*3*	*Top*
	Bottom	0.66	0.31	0.03	0.00
Wage quartile in 40-	*2*	0.13	0.61	0.25	0.01
49 age range	*3*	0.02	0.14	0.60	0.24
	Top	0.00	0.02	0.15	0.83

Source: PSID core sample household heads 1968-2001; author's calculations.

While the most common movement is no movement across quartiles, about 40 percent of moderate wage earners and between 20 and 30 percent of low and high wage earners switch quartiles in each transition.

The simulations also require parameterization. Other than for the Pareto weights, I use the parameters specified in Section 1. Pareto weights are assigned to each (ex post) lifetime path of wages using the expression (16) from the baseline model simulation, yielding weights similar to the baseline case. The maximum weight is 1.00 and applies to the individual receiving the lowest wage level at each age. The minimum weight is 0.83 and applies to the individual receiving the highest wage level at each age.

3.3.2   Simulation Results

The simulation results for the Case 3 planner's problems reinforce the lessons of the baseline and Case 2 simulations. As in the baseline model, I examine intratemporal distortions, average tax rates, intertemporal distortions, and social welfare.

First, consider intratemporal distortions. The Static Mirrlees and Partial Reform distortions are shown in Table 11.

Table 11: Intratemporal distortions in Case 3 model simulation

Intratemporal distortion		*Wage quartile in each age range*				
	Age range	*Bottom*	*2*	*3*	*Top*	*Simple average*
	30-39	0.35	0.30	0.27	0.26	
Status Quo	40-49	0.40	0.43	0.47	0.01	0.276
	50-59	0.16	0.27	0.39	0.00	
	30-39	0.29	0.28	0.28	0.00	
Partial Reform	40-49	0.34	0.34	0.35	0.00	0.248
	50-59	0.35	0.35	0.38	0.00	

Figures 8a and 8b plot these distortions against lifetime income.[51]

---

[51] To avoid confusion, the Full Optimum distortions are not shown in the table because they are not readily comparable to the other two scenarios, as they are not functions of current income only.

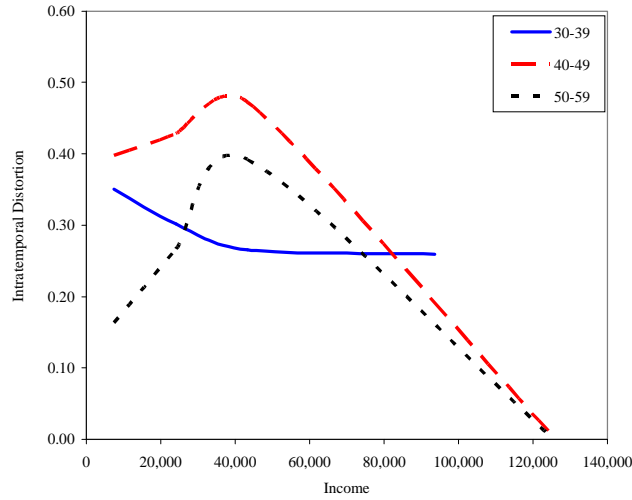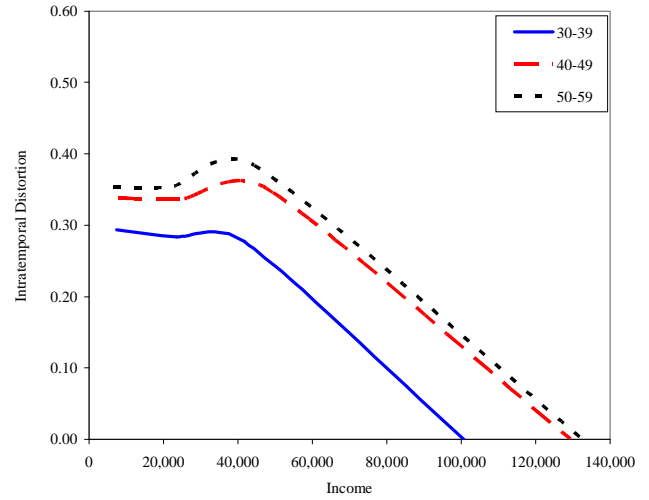**Figure 8a: Intratemporal Distortions in Static Mirrlees, Case 3**   **Figure 8b: Intratemporal Distortions in Partial Reform, Case 3**

The striking disparity in the treatment of the high-income young that we saw in previous models (Figures 2a and 2b or Figures 5a and 5b) is apparent here as well, so that high-skilled young workers are inefficiently discouraged from working by an age-independent tax system. This is consistent with Proposition 5 (Baseline and Case 3 Equivalence), which implied that the characteristics of the Partial Reform and Static Mirrlees policies in the baseline model were likely to carry over to the Case 3 model.

Moreover, the Partial Reform policy again uses marginal distortions less overall than does the Static Mirrlees policy. The unweighted average distortion is 0.276 in the Static Mirrlees, compared to 0.248 in the Partial Reform policy. The difference in the average distortion increases when weighted by income, from 0.186 in the Static Mirrlees to 0.128 under Partial Reform.

The results on average tax rates also resemble those from the baseline model. The average tax rates for the Static Mirrlees and Partial Reform results are given in Table 12.

Table 12: Average Tax Rates in Case 3 model simulation					
Average tax		*Wage quartile in each age range*			
	Age range	*Bottom*	*2*	*3*	*Top*
	30-39	-293.7	-42.9	1.5	34.8
Status Quo	40-49	-293.7	-42.9	1.5	35.7
	50-59	-293.7	-42.9	1.5	35.3
	30-39	-346.3	-63.5	-10.9	28.3
Partial Reform	40-49	-303.0	-41.9	6.2	40.5
	50-59	-363.3	-53.3	4.1	43.0

These average tax rates are plotted against lifetime income in Figures 9a and 9b.
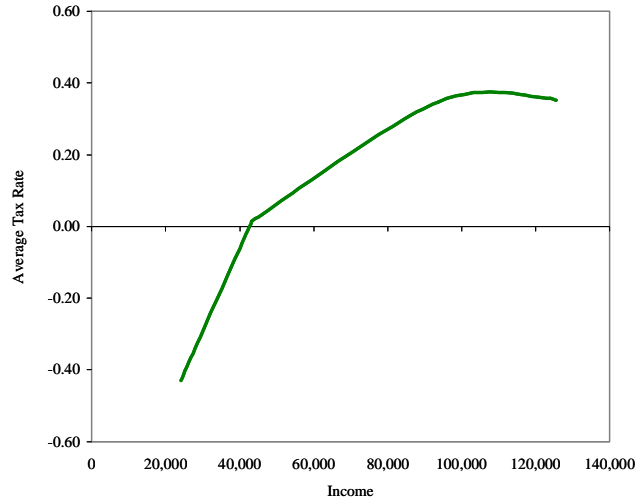
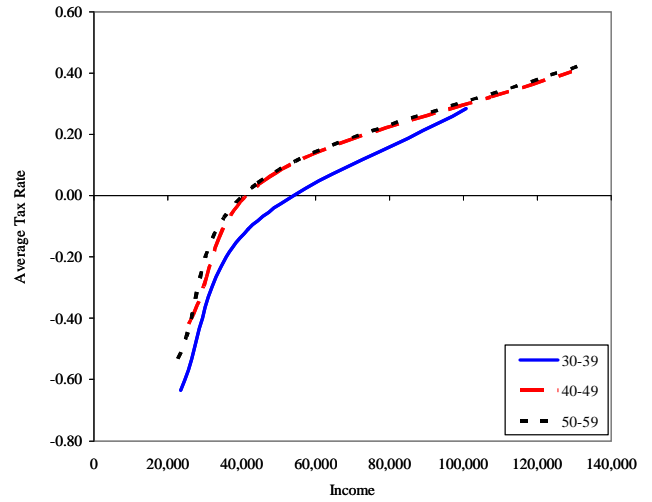**Figure 9b: Average Tax Rates in Static Mirrlees, Case 3**   **Figure 9b: Average Tax Rates in Partial Reform, Case 3**

As in the baseline model, the Partial Reform policy lowers average tax rates on workers in their thirties. The size of the gap in rates in the middle of the income distribution between the young and peak earners resembles that in the baseline case, and the intuition is the same. In Case 3, individuals cannot borrow against their higher expected future wages, so tax policy can substitute for private borrowing.

I also compare the intertemporal distortions under each scenario. Table 13 shows the ratio

$$\frac{c_t^i}{\sum_{j=1}^{I} P_{t,t+1}(i,j) c_{t+1}^j (W_t^i)},$$

which rearranges the Inverse Euler Equation in expression (25) when utility of consumption is logarithmic, for the Partial Reform and Static Mirrlees policies.

Table 13: Intertemporal distortions in Case 3 model simulation

		Wage quartile in each age range			
	Age range	Bottom	2	3	Top
	30-39	0.94	0.95	0.89	0.84
Status Quo	40-49	0.93	0.95	0.84	1.07
	50-59	…	…	…	…
	30-39	1.01	1.01	0.95	1.04
Partial Reform	40-49	0.97	0.99	0.91	1.11
	50-59	…	…	…	…

Though not shown in the table, this ratio is equal to one in the Full Optimum for each individual (aside from some numerical noise in the extremely unlikely paths). Apparent from the table is that the deviations of this ratio from one are larger for the Static Mirrlees planner than for the Partial Reform planner. Note, in particular, the substantial distortions on the high-skilled young by the Static Mirrlees planner.

Finally, Partial Reform continues to capture a large absolute and relative welfare gain. Table 14 shows overall social welfare and lifetime utility for individuals with four representative wage paths under the three policies.

Table 14: Welfare in Case 3 model simulation

	Social Welfare (consumption equivalents*)	Lifetime Utility (consumption equivalents**) Wage paths			
		Always lowest wage	*2,2,2*	*3,3,3*	*Always highest wage*
Status Quo	2.466	2.47	2.54	2.68	2.96
Partial Reform	2.488	2.53	2.58	2.70	2.91
Full Optimum	2.489	2.53	2.58	2.70	2.90

* The value for consumption that, if provided freely to all workers at each age, would generate the same social welfare as the actual allocation.

** The value for consumption that, if provided freely to the worker at each age, would generate the same lifetime utility as the worker obtains with the actual allocation.

Figure 10a plots social welfare under the three policies.

Figure 10a: Social Welfare Comparisons, Case 3
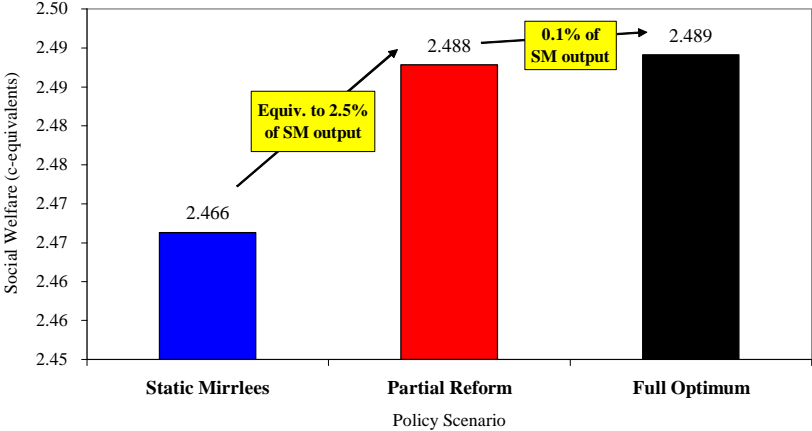
Age dependence yields a welfare gain equivalent to a 2.5 percent increase in aggregate consumption relative to the Static Mirrlees policy in Case 3. Moreover, this gain captures 95 percent of the welfare gain from reform to the Full Optimum policy. Both of these results are similar in magnitude to the results from the baseline and Case 2 models.

Again, the utility gains are especially substantial for the lower-income workers in this model. In Table 14 and Figure 10b, I show the lifetime utilities of four individuals with wage paths that stay in the same quartile for all three periods.

Figure 10b: Utility Comparisons, Case 3
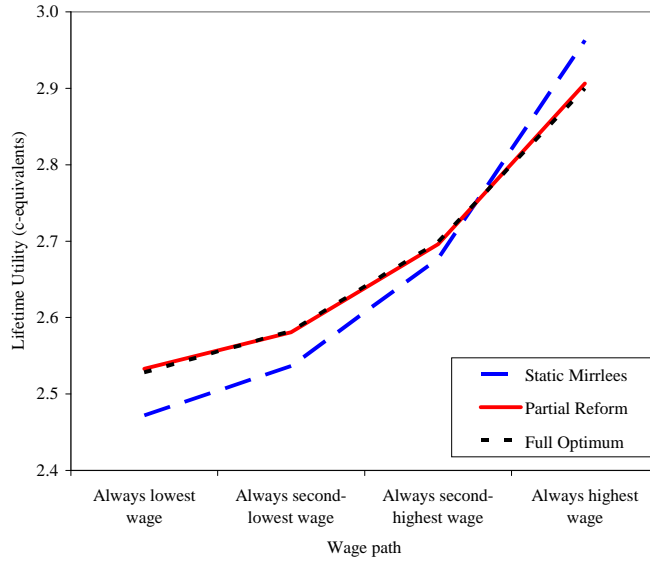
As in previous models, age dependence makes the distribution of lifetime utility more equal while raising overall welfare.

As mentioned at the start of Section 3, an important reason to allow for stochastic wage paths is that they capture heterogeneity in wage paths conditional on lifetime income. The same topic was raised in Section 1 when discussing why the Partial Reform captures such a large share of the gain from the Full Optimum policy. Why don't stochastic wage paths undermine the relative case for Partial Reform? The result relies on two factors. First, the stochastic nature of these heterogeneous wage paths weakens the Full Optimum policy, because it now has to satisfy incentives repeatedly rather than only once. This narrows the gap between the Full Optimum and the history-independent policies that were already forced to satisfy incentives at each age. Second, the empirical magnitude of wage path heterogeneity (i.e., wage path crossing) is not large enough to change the main results. At each age, current income is a powerful enough predictor of lifetime income-earning ability that the Partial Reform policy can still redistribute on a lifetime basis by redistributing within ages.[52]

Thus, the results from the baseline model are robust to the inclusion of wage stochasticity, at least to the extent implied by the data used in these analyses and for a parsimonious specification of the stochastic process. In the next section, I again relax the assumption that individuals cannot save or borrow, but now in the context of stochastic wages.

---

[52]This result does not rely on the use of incomes averaged over decade-long age ranges. Though these averages smooth incomes, the correlations between income at each age and lifetime income-earning ability in these data are nearly as large as those between the decade-long average incomes and lifetime income-earning ability. Moreover, I have simulated a version of the Case 3 model in which I use five-year age ranges. The welfare gain from Partial Reform in fact increases relative to the main results, as the wider differences between age groups raise the value of age-dependent marginal distortions and transfers across age groups.

4 Case 4: Model with stochastic wage paths and private saving and borrowing

In this section, I combine the variations on the baseline model examined separately in the previous two sections and consider a model with both stochastic wages and private saving and borrowing. I make the same modifications to the model in Case 3 as I did to the baseline model when specifying the model in Case 2. In particular, I retain the assumption from Case 2 that the Static Mirrlees and Partial Reform planners cannot tax intertemporal transfers by individuals.

As in the previous models, I consider a social planning problem for each policy, where a planner maximizes social welfare subject to feasibility and incentive compatibility constraints. When individuals can transfer resources across periods, the planners in the Static Mirrlees and Partial Reform scenarios do not control consumption directly. Thus, these planners specify pre-tax income and after-tax income bundles in Case 4, just as in Case 2. The objective function for these two policies is

$$\max_{\{x,y\}} \left\{ \sum_{i(t)} \pi^{i(t)} \alpha \left(W_T^{i(t)} \right) \sum_{t=1}^{T} \beta^{t-1} \left(u \left(c_t^{i_t} \right) - v \left(\frac{y_t^{i_t}}{w_t^{i_t}} \right) \right) \right\}, \tag{27}$$

and the feasibility constraint is:

$$\sum_{i(t)} \pi^{i(t)} \sum_{t=1}^{T} R^{T-t} \left(y_t^{i_t} - x_t^{i_t} \right) = 0. \tag{28}$$

Both of these expressions include, as in Case 2, after-tax income x, and all other notation is as before.[53]

As usual, variations in the incentive constraints allow us to distinguish between policy scenarios. The combination of private access to capital markets and stochasticity makes these incentive constraints quite complicated, however. Therefore, I relegate them to the Appendix.

In words, the incentive constraints for the Static Mirrlees and Partial Reform scenarios reflect that an individual can choose a separate deviation strategy, including saving and borrowing, for each possible true path of wages. So, the Static Mirrlees incentive constraints must ensure that each individual prefers its allocation to any of the other allocation streams it might claim. If there are T periods and I wage levels, the number of these other streams is $\left[(IT)^{1+I(T-1)} - 1 \right]$. That is, in the first period, an individual can claim any of IT wages, including his own. When planning for the second period, for each of the I possible second period wages, he can claim any of the IT wages again. The Partial Reform incentive constraints are, as usual, a subset of the Static Mirrlees scenario's because the planner can make age-dependent allocations. Each individual in the Partial Reform must be prevented from claiming allocation streams other than her own that number only $\left[I^{1+I(T-1)} - 1 \right]$, as she can claim I, not IT, wages for each wage level she receives.

In contrast, the Full Optimum planner's problem is unchanged from the problem without private saving, just as its Case 2 problem was unchanged from its baseline problem. Thus, the Full Optimum planner's problem in Case 4 is identical to the Case 3 Full Optimum planner's problem.

The complexity of the Static Mirrlees and Partial Reform problems makes it most convenient to study

---

[53] The Static Mirrlees and Partial Reform allocations may, in principle, be different for two individuals with different histories but the same current wage, in that these individuals could choose different (x, y) pairs. Computational considerations prevent me from allowing for this, however, and instead I restrict these policies to allocations that are identical across two such individuals. The impact of this restriction is likely to be minimal, as economic efficiency and incentive constraints require allocations to these individuals to be similar. Moreover, this restriction has no effect on the Full Optimum policy and primarily handicaps the Partial Reform policy, causing me to, if anything, underestimate the relative gain from age dependence.

their solutions numerically, so I turn to the quantitative simulations now.

4.1 Numerical results

The computational demands of the Case 4 model are substantial, so I further limit the size of the economy. I condense the lifecycle into two age periods ($T = 2$) covering the same range as before, so that $t = 1$ for ages 30-44 and $t = 2$ for ages 45-59. I also limit the number of wage levels at each age to three ($I = 3$).

The construction of the data is the same as in Case 3, though I now choose the 5th, 50th, and 95th percentiles for each age range as the representative wage levels and classify individuals into three wage quantiles. Table 15 gives the wage levels at each age and the transition matrix between the two age ranges.

Table 15: Data for simulation of Case 4

		Percentile of wage distribution		
Wage levels	*Age range*	*5th*	*50th*	*95th*
Wage for	30-44	5.48	15.92	37.83
specified	45-59	5.25	16.46	44.16
	Initial pbb	0.35	0.39	0.25

Transition matrix		Wage quantile in 44-59 age range		
		Bottom	*Middle*	*Top*
Wage quantile	*Bottom*	0.72	0.27	0.01
in 30-44 age	*Middle*	0.15	0.63	0.22
range	*Top*	0.01	0.20	0.79

Source: PSID core sample household heads 1968-2001; author's calculations.

The parameterization of the models is the same as in the previous cases, and the Pareto weights follow a similar pattern as before, with a maximum weight of 1.00 and a minimum weight of 0.83.

I examine results on intratemporal distortions, average tax rates, and social welfare.

First, consider intratemporal distortions. The distortions in the Static Mirrlees and Partial Reform policies are listed in Table 16.

Table 16: Intratemporal distortions in Case 4 model simulation

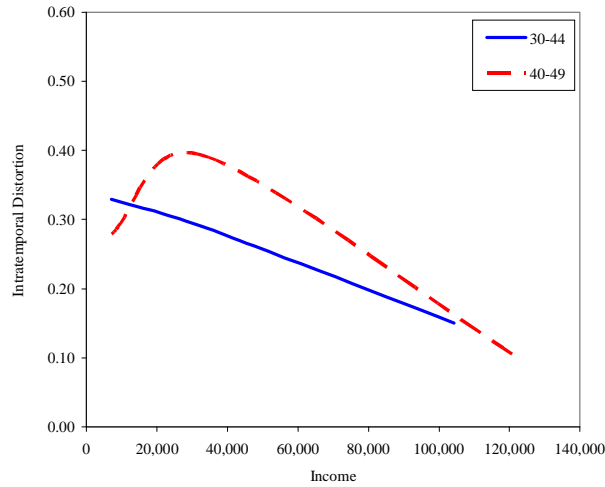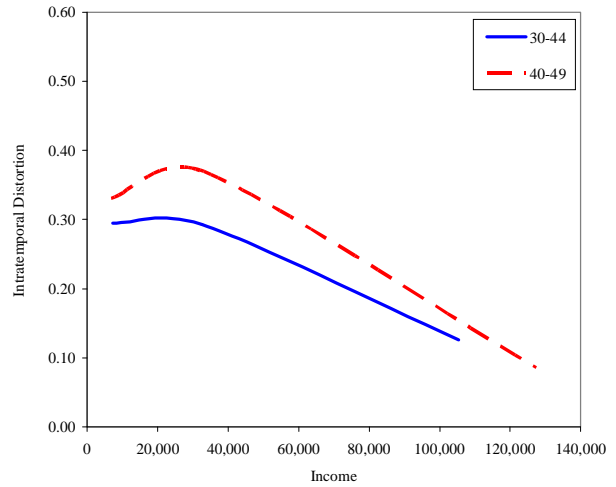| **Intratemporal distortion**		Wage quantile in 45-59 age range			
	Age range				*Simple average*
		Bottom	*Middle*	*Top*	
Status Quo	30-44	0.33	0.29	0.15	0.256
	45-59	0.28	0.39	0.10	
Partial Reform	30-44	0.29	0.29	0.13	0.250
	45-59	0.33	0.37	0.09	

Figures 11a and 11b plot these intratemporal distortions against lifetime income.

As in all previous models, the high-skilled young workers are inefficiently discouraged from working by (here, slightly) higher intratemporal distortions under an age-independent tax system.

Moreover, the use of marginal distortions is lower in the Partial Reform policy than in the Static Mirrlees policy in general, with the unweighted average distortion falling from 0.256 to 0.250 between the policies. If the distortions are weighted by income, the gap is larger, with the average distortion falling from 0.176 in the Static Mirrlees to 0.161 in the Partial Reform policy. The smaller magnitude of these results in Case 4 than in the other models is likely due, at least in part, to the compression of the data into two age groups. This compression limits the extent of cross-age incentives problems that cause the Static Mirrlees planner to use more distortionary taxation than the Partial Reform planner.

As in Case 2, optimal average tax rates are indeterminate in this setting for the Partial Reform policy. Lump-sum transfers across ages, which affect calculated average tax rates, can be reallocated across time by individuals using the same technology as the planner. This allows for the possibility that average tax rates in the Partial Reform policy follow a similar pattern as in the baseline and Case 3 models, where average tax rates are lower on young workers. An example of such a policy is shown in Table 17.

Table 17: Average Tax Rates in Case 4 model simulation

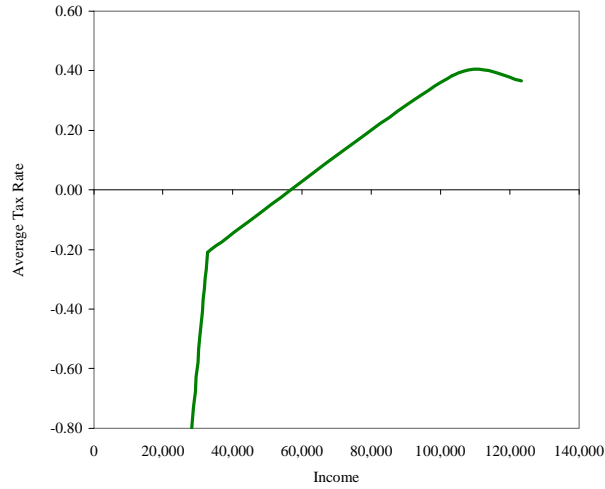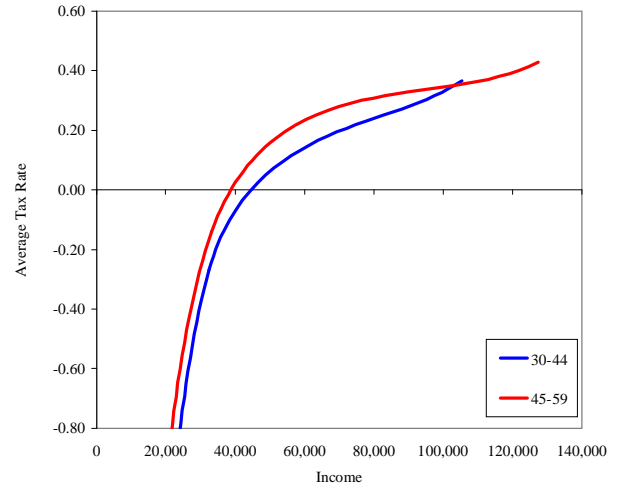| Average tax (in percent)		Wage quantile in 45-59 age range		
	Age range	Bottom	Middle	Top
Status Quo	30-44	-352.0	-20.4	38.6
	45-59	-352.0	-20.4	36.8
Partial Reform	30-44	-370.9	-26.9	36.7
	45-59	-353.4	-14.0	42.9

Figures 12a and 12b plot these example average tax rates against lifetime income

It is important to emphasize that the average tax schedules shown in Figure 12b are not the unique optimal schedules, and there exist schedules that include higher average rates for young workers that yield the same aggregate welfare. Nevertheless, a policy placing lower average taxes on the young may be preferable, given its advantages when private saving and borrowing are restricted as in the baseline and Case 3 settings.

Finally, age dependence continues to yield a large welfare gain and capture a substantial share of the welfare gain from reform to the Full Optimum policy. Table 18 shows social welfare and lifetime utility for individuals with three representative wage paths under the three policies.

Table 18: Welfare in Case 4 model simulation

	Social Welfare (consumption equivalents*)	Lifetime Utility (consumption equivalents**) Wage paths		
		Always lowest wage	*Always middle wage*	*Always highest wage*
Status Quo	3.629	2.03	2.06	2.22
Partial Reform	3.643	2.04	2.07	2.21
Full Optimum	3.662	2.06	2.09	2.19

Figure 13a plots social welfare under these policies.

52

Figure 13a: Welfare Comparisons, Case 4
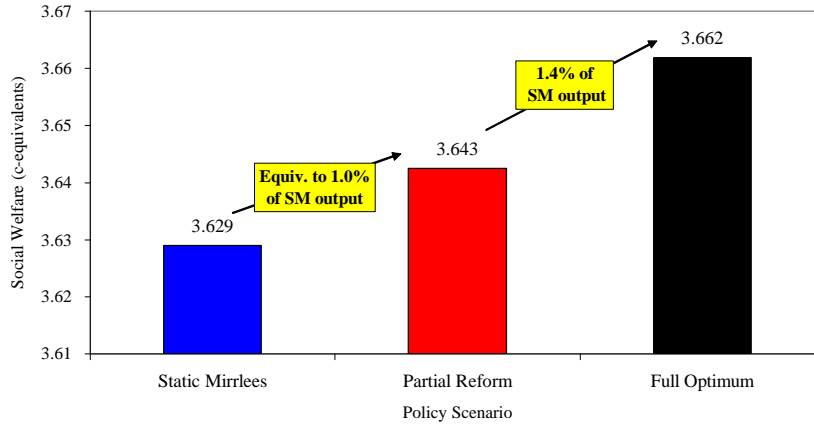
Partial Reform generates a welfare gain equivalent to 1.0 percent of aggregate consumption in the Static Mirrlees. This captures 41 percent of the welfare gain from reform to the Full Optimum. As in previous models, the low-skilled particularly benefit from reform: both the Partial Reform and Full Optimum policies provide more redistribution than the Static Mirrlees policy.

5 Discussion and Special Topics

The preceding sections show that the partial reform of age dependence yields large absolute and relative welfare gains by systematically altering optimal labor income taxation. Moreover, these effects are robust to fundamental variations in the assumed economic environment. In this section, I discuss additional topics of interest that were not addressed directly in these analyses.

5.1 Endogeneity of wage paths

As stated at the beginning of the baseline model setup, I assume throughout this paper that wage paths are exogenous to individuals. This assumption is standard in the optimal tax literature, but it is not necessarily innocuous.

If, as this paper's analysis recommends, tax schedules were to differ by age, individuals would have an incentive to tailor their career choice and employment relationships to minimize their tax bill. This could reduce the variation in wage distributions with age that gives age-dependent taxes their power and introduce additional distortions to the economy. For instance, lower average tax rates on young workers would encourage people to take jobs with flatter income profiles and to bargain with their employers to shift the timing of income.[54]

The specific results of this paper therefore require that a substantial portion of the variation of wages with age is inelastic to taxes. A few considerations suggest that this requirement's effects on the paper's results may be limited.

First, this paper's focus on age-dependent taxes between the ages of 30 and 60 limits concerns about distorting individuals' career choices. In their late teens and twenties, individuals have substantial oppor-

---

[54]It is important to be clear that while wages are assumed to be exogenous, this paper's analysis allows income to respond to taxes because individuals choose their level of labor effort.

tunities to shift the timing of higher education and job training to respond to taxes or other incentives.[55] By age thirty, however, nearly all have completed their education and begun careers, so any distortions to career choice would apply to only those individuals who were substantially forward-looking and for whom the distortion itself had relatively small costs.

Second, the temporary nature of most employer-employee relationships provides a natural barrier to shifting income across ages in response to age dependence, because shifting income is risky without long-term contracts that tie employees to employers. For instance, the taxes recommended by this paper's analysis would imply shifting income forward, so that some of a worker's earnings in her forties would be pre-paid to her in her thirties. Of course, an employer will be hesitant to do this unless the worker can commit to remaining at the firm through her forties. Such commitments are rare in modern labor markets. More generally, different theories of wage determination suggest different sensitivities of income's timing to taxation.[56] The key for this paper is that wages rise over the lifecycle because the passage of time is, for whatever reason, required for a given worker's effort to be worth more to their employers.

Finally, one piece of evidence suggests that the sensitivity to taxes of both career choice and the timing of income is limited. Currently, wage paths rise sharply with age, especially at high incomes. These upward-sloping paths exist in the context of progressive taxation that ought to encourage flat wage profiles. If wage paths were highly elastic to tax incentives, we would expect to see smoother wage profiles than we do.

Despite the potential importance of wage path endogeneity, characterizing optimal dynamic taxation (and age-dependent taxation) with endogenous wage paths is beyond the scope of this paper and is an important task for future work. Doing so will require a careful treatment of career choice and the timing of income, as discussed above. It will also require the inclusion of a model of human capital investment through education and work experience, a factor that may increase the welfare gains from age dependence as its added flexibility could be used to encourage human capital accumulation.

5.2 Elasticity of Labor Supply by Age

The analyses of the preceding sections, other than a brief discussion in Section 1, have ignored one of the most direct reasons for the differentiation of taxation by age or any other personal characteristic: variation in the elasticity of labor supply across subgroups (see, for instance, Alesina and Ichino (2007) on gender). Standard optimal tax theory implies that less elastic subgroups should face larger tax distortions, all else the same, as revenue can be raised more efficiently from them. Therefore, a potentially important determinant of age-dependent taxes absent from this paper's results is variation in the elasticity of labor supply across ages.

Unfortunately, empirical evidence on variation in the elasticity of labor supply with age is limited. Kremer (2002) argues that "The limited available evidence suggests that younger workers have more elastic labor supply than prime-age workers," citing Clark and Summers (1981), who show more variation in employment rates with the business cycle for young workers. French (2005) estimates that "labour supply elasticities rise from 0.3 at age 40 to 1.1 at age 60," but estimates for other ages are not given. Lacking more robust evidence, I have made the conservative assumption that the elasticity of labor supply is uniform across age

---

[55] Age dependence during this younger age range would be a more treacherous reform to design, though properly-designed age-dependent taxes during this range would potentially add significantly to the welfare gains calculated in this paper.

[56] For instance, if learning about the quality of matches or on-the-job training is important, individuals are unlikely to be able to shift income to earlier in their careers, while if the acquisition (but not the timing) of outside training is important, it might be shifted to an age at which taxes on income are higher.

If labor supply elasticity varies in the directions suggested by this limited evidence, the recommendations of this paper are strengthened. To illustrate this, I consider a parameterization that includes a simple difference in elasticies by age. Consider the baseline model from Section 1 where there are $I = 10$ types of individuals living for $T = 3$ periods. Suppose that $\sigma = 3$ for the second age group (workers in their forties) while $\sigma = 2$ for the workers in their thirties and fifties. Given the isoelastic disutility function (9), the constant-consumption elasticity of labor supply is $\frac{1}{\sigma-1}$, so these values imply an elasticity of 1 for the youngest and oldest groups and an elasticity of $\frac{1}{2}$ for the workers in their forties.

The results of this experiment magnify those of the main analyses. Intratemporal distortions remain too high for the high-earning young and are used more in general by the Static Mirrlees policy than by the policies with age dependence. Average tax rates for workers in their thirties are even lower relative to older workers in this Partial Reform policy than in the model with uniform elasticities. The welfare gain from Partial Reform is increased from 2.0 percent of aggregagte income in the baseline model to 2.4 percent in this model with varying elasticities, as it can target distortions at the inelastic age groups better than the Static Mirrlees. Age dependence continues to capture nearly all of the potential gain from the Full Optimum.

5.3 Extensive Margin

One reason that we may intuitively think the elasticity of labor supply is higher for the young and old is not captured by the previous discussion. Young and old workers may be elastic along the extensive labor supply margin (the choice whether to work or not) rather than the intensive margin (the choice of how much to work). How would an extensive margin affect this paper's results?

To add an extensive margin to the analysis, I modify the baseline model of Section 1 to include an eleventh type of individual, type $i = 0$, who never works. A worker with type $i > 0$ who chooses not to work is operating on the extensive margin. Note that, because the Partial Reform planner cannot make history-dependent allocations, workers can move across the extensive margin in a single period or any combination of periods.

To properly model this extensive margin, I must make not working qualitatively different from working less. To do so, I add a fixed cost of working, ϕ. Formally, the incentive constraints in the Partial Reform planner's problem for individual i of age t have two parts: first,

$$\beta^{t-1}\left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) - \phi\right) \geq \beta^{t-1} u\left(c_t^0\right), \tag{29}$$

for all $i \in \{1, 2, ..., I\}$ which prevents $i > 0$ from preferring not to work; and second,

$$\beta^{t-1}\left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) - \phi\right) \geq \beta^{t-1}\left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^i}\right) - \phi\right), \tag{30}$$

for all $i, j \in \{1, 2, ..., I\}$ and all t, which simplify to the same conditions as (6) from the baseline model because ϕ cancels on both sides. I simulate the model with $\phi = \ln(1.712)$, representing a fixed cost equal to 25 percent of the average wage for type $i = 1$.

The lessons from this paper's main analyses are unchanged by adding an extensive margin, though the optimal policies do respond to the extensive margin. One response of policy is that, while average tax rates have the same shape as in the baseline model, they are increased throughout the income distribution by the addition of an extensive margin. Intuitively, consumption is being provided to the $i = 0$ individuals who

do not work, so average taxes on all other types must increase. A second, more subtle response is consistent with the analysis of Saez (2002). In the simulation of policy with an extensive margin, allocations mimic the U.S. Earned Income Tax Credit, whereby low earners receive a subsidy (i.e., a negative marginal tax rate) to encourage them to work rather than claim the $i = 0$ allocation.

5.4 Pareto-improving age dependence

The main analysis in this paper assumes that social planner's problem is to maximize a Utilitarian social welfare function. This is a restrictive though standard assumption, and concerns about it have inspired research on Pareto efficient taxation such as Stiglitz (1987) and, more recently, Werning (2007b). In a similar vein, the original partial reform approach of Guesnerie (1977) stressed incremental Pareto improvements to tax policy, not incremental Utilitarian improvements. For those uncomfortable with reforms that sacrifice the welfare of some individuals for greater gains by others, the key question is whether age dependence is a Pareto-improving partial reform: that is, a reform that can raise social welfare without harming any individuals.[57]

Pareto-improving age dependence would also be more likely to succeed as a policy proposal. In particular, concerns about the impact of moving to an age-dependent system can be mitigated by using some of the surplus value generated by the Pareto improvement to compensate those who would otherwise lose in the transition.[58]

To test whether age dependence is a Pareto-improving partial reform, I simulate the baseline model with the additional restriction that no individual can be worse off under the age-dependent policy than under the Static Mirrlees policy.[59] As in the Utilitarian model, marginal distortions on high-income young workers and average taxes on all young workers are lower under the Pareto-improving age dependent tax policy than under the Static Mirrlees. More important, the welfare gain from Partial Reform is nearly as large as in the baseline, equivalent to 1.8% of aggregate consumption in the Static Mirrlees. The Pareto-improvement restriction ensures that the highest earners are left with their utility levels from the Static Mirrlees policy, while reform generates a substantial increase in welfare for lower earners. This result suggests that age dependence is a reform capable of attracting broad-based support.

6   Conclusion

In this paper, I studied a partial reform of tax policy: age-dependent labor income taxes. To do so, I used modern dynamic Mirrleesian optimal tax methods to contrast three policy scenarios: a Static Mirrlees policy restricted to age-independent taxes, a Partial Reform policy in which labor income taxes can be age-dependent, and a Full Optimum policy in which only private information constrains the design of taxes. In a baseline model, I showed how classic theoretical results on the intratemporal and intertemporal policy margins apply to age-dependent policy. I examined how age dependence affects these margins in economic environments with stochastic wages and private saving and borrowing, as well.

Then, I used data from the U.S. Panel Study of Income Dynamics to calibrate and simulate the three

---

[57]Blomquist and Micheletto (2003) illustrate the theoretical potential for age dependence to be Pareto-improving.

[58]Another option to avoid transition concerns is to make age dependence apply only to generations born after the date of the policy being approved.

[59]Recall that there is no mortality risk in the model economy. In reality, individuals with shorter lives may be relatively disadvantaged, as taxes are likely to be lower on the old. As the welfare benefits calculated above are based solely on the ages between 30 and 59, however, the affected population is small.

policy scenarios. This quantitative analysis yielded two specific policy recommendations that were largely robust across settings. First, marginal income taxes ought to be lower for high-earning young workers in an optimal age-dependent policy than in an age-independent policy. These individuals are near the top of their age-specific wage distribution, so the efficiency costs of distorting their labor effort are substantial. In an age-dependent tax system, the benefit from such a distortion (increasing tax revenue from higher earners) is relatively small, whereas the benefit appears much larger in an age-independent system that cannot recognize the position of these individuals within their age's distribution. This specific example illustrates a more general finding that age dependence avoids using marginal distortions that age-independent policy cannot, raising the efficiency of the tax system. Second, younger workers ought to face a lower average tax schedule than middle-aged workers if private saving and borrowing are restricted, as differential average taxes by age substitute for private borrowing in the presence of rising wage paths. In models with private saving and borrowing, a variety of average tax schedules can implement the optimum, including policies that have lower average taxes on the young.

Finally, the calibrated policy simulations allowed me to quantify the welfare gain from age dependence and understand its components. Age dependence yields a large welfare gain equal to between one and three percent of aggregate annual consumption. Moreover it captures a substantial portion of the gain from reform to the optimal dynamic policy, ranging from above 40 percent to approximately 95 percent of the potential gain depending on assumptions about the economic environment. Age dependence provides especially large welfare gains for the low-skilled, but most people obtain higher utility than they would under an age-independent policy. In fact, a simulation with the added constraint that age dependence be Pareto-improving yields nearly as large a social welfare gain as does the standard, Utilitarian-optimal age-dependent policy.

These findings show that that age dependence, which requires only a simple change to current tax policies, is nevertheless a potentially powerful reform. Future work on age dependence ought to extend this analysis in a few directions outside the scope of this paper.

First, the quantitative analysis of this paper focuses on individuals between 30 and 60 years of age. Some of the largest gains to age dependence may come from individuals outside this range, as wage distributions for people in their twenties and sixties are substantially different from those in the range studied here. This paper neglected those age ranges to avoid large uncertainties about how to treat distortions to the acquisition of human capital early in life (i.e., education) and to the retirement margin in the presence of Social Security, and the age-dependent taxes derived above would have little effect on individuals' choices on these margins. If these margins were properly modeled, however, the benefits of age dependence may be substantially increased.

Second, as mentioned in Section 5, this paper has assumed that the elasticity of labor supply is constant across age groups. While this assumption is almost certainly false, solid evidence on variation in labor supply elasticity with age is surprisingly rare. Any variation in this elasticity will raise the value of age dependence, so identifying it should be a high priority for future work.

Finally, an important next step toward taking advantage of this policy opportunity is to use the results of this paper to design and study specific changes to existing taxes. The findings of this paper suggest that such an exercise would identify relatively simple ways to increase the efficiency and equity of current tax policy, yielding substantial welfare gains.

References

[1] Akerlof, George, (1978). "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," American Economic Review, 68(1), March, pp. 8-19.

[2] Albanesi, Stefania and Christopher Sleet, (2006). "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies* 73, pp. 1-30.

[3] Alesina, Alberto and Andrea Ichino (2007). "Gender-based Taxation," Working Paper, March.

[4] Altig, David, Alan Auerbach, Laurence Kotilkoff, Kent Smetters, and Jan Walliser (2001). "Simulating Fundamental Tax Reform in the United States," *American Economic Review* 91, pp. 574-595

[5] Atkinson, Anthony and Joseph E. Stiglitz, (1976). "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics* 6, pp. 55-75.

[6] Banks, James, Peter Diamond, and James Mirrlees, (2007). "Mirrlees Report Chapter on the Base for Direct Taxation." Manuscript.

[7] Bisin, Alberto and Adriano Rampini (2005). "Markets as Beneficial Constraints on the Government," *Journal of Public Economics.*

[8] Blomquist, Soren and Luca Micheletto, (2003). "Age Related Optimal Income Taxation", Working Paper 2003:7, Department of Economics, Uppsala University. (downloadable at the website http://www.nek.uu.se/pdf/wp2003_7.pdf). Forthcoming *Scandinavian Journal of Economics.*

[9] Clark, Kim B. and Lawrence H. Summers (1981). "Demographic Differences in Cyclical Employment Variation," *Journal of Human Resources* 16, pp. 61-77.

[10] Diamond, John W. and Joyce Tung (2006). "Lifecycle Wealth Holdings by Lifetime Income," Rice University Working Paper.

[11] Diamond, Peter (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review* 88(1), March.

[12] Erosa, Andres and Martin Gervais (2002). "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory* 105, pp. 338-369.

[13] Farhi, Emmanuel and Ivan Werning (2007). "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation." MIT working paper, April.

[14] Feldstein, Martin (1976). "On the Theory of Tax Reform," *Journal of Public Economics* 6, pp. 177-204.

[15] French, Eric (2005). "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies* 72, pp. 395-427

[16] Fullerton, Don, and Diane Lim Rogers (1993). *Who Bears the Lifetime Tax Burden?* Washington, D.C.: Brookings Institution.

[17] Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning (2006). "New Dynamic Public Finance: A User's Guide," *NBER Macroannual* 2006.

[18] Golosov, Mikhail and Aleh Tsyvinski (2006). "Designing optimal disability insurance: A case for asset testing," *Journal of Political Economy.*

[19] Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski (2003). "Optimal Indirect and Capital Taxation," *Review of Economic Studies* 70, pp. 569-587.

[20] Guesnerie, Roger (1977). "On the Direction of Tax Reform," *Journal of Public Economics* 7, pp. 179-202.

[21] Guvenen, Fatih (2007). "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97(3), June.

[22] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante (2005). "Two Views of Inequality Over the Life-Cycle," *Journal of the European Economic Association* (Papers and Proceedings), 3(2-3), pp. 543-52

[23] Judd, Kenneth and Che-Lin Su (2006). "Optimal Income Taxation with Multidimensional Taxpayer Types," Working Paper, November 22.

[24] Keane, Michael P. and Kenneth I. Wolpin (1997). "The Career Decisions of Young Men," *Journal of Political Economy* 105(3), pp. 473-522.

[25] Kocherlakota, Narayana (2005). "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica* 73(5), September.

[26] Kremer, Michael (2002). "Should Taxes be Independent of Age," Harvard University Working Paper.

[27] Meade, James (1978) *The Structure and Reform of Direct Taxation.* Allen and Unwin.

[28] Mirrlees, J.A., (1971). "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies* 38, 175-208.

[29] Rogerson, William, (1985). "Repeated Moral Hazard," *Econometrica* 53, pp. 69-76.

[30] Saez, Emmanuel, (2001). "Using Elasticities to Derive Optimal Income tax Rates," *Review of Economic Studies* 68, pp. 205-229.

[31] Saez, Emmanuel, (2002). "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics,* August, pp. 1039-1073.

[32] Salanie, Bernard (2003). *The Economics of Taxation.* MIT Press.

[33] Stiglitz, Joseph E., (1987). "Pareto efficient and optimal taxation and the new new welfare economics," in: Auerbach, Alan and Martin Feldstein (eds.), *Handbook on Public Economics.* North Holland: Elsevier.Science Publishers, pp. 991-1042.

[34] Storesletten, Kjetil, Chris Telmer, and Amir Yaron, (2001). "How Important Are Idiosyncratic Shocks? Evidence from Labor Supply," *American Economic Review* 91(2), pp. 413-417, May.

[35] Vickrey, W., (1945). "Measuring Marginal Utility by Reactions to Risk," *Econometrica* 13, 319-333.

[36] Werning, Ivan (2007a). "Optimal Fiscal Policy with Redistribution," *Quarterly Journal of Economics,* pp. 925-967, August.

[37] Werning, Ivan (2007b). "Pareto Efficient Income Taxation," Working paper, April.