Review of Economic Studies (2011) **78**, 1490–1518 doi: 10.1093/restud/rdr001

Advance access publication 3 April 2011

The Surprising Power of Age-Dependent Taxes

MATTHEW WEINZIERL
Harvard Business School and NBER

First version received June 2009*; final version accepted October* 2010 (*Eds.*)

This paper provides a new, empirically driven application of the dynamic Mirrleesian framework by studying a feasible and potentially powerful tax reform: age-dependent labour income taxation. I show analytically how age dependence improves policy on both the intratemporal and intertemporal margins. I use detailed numerical simulations, calibrated with data from the U.S. Panel Study of Income Dynamics, to generate robust policy 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. Finally, I calculate and characterize the welfare gains from age dependence. Despite its simplicity, age dependence generates a welfare gain equal to between 0·6% and 1·5% of aggregate annual consumption, and it captures more than 60% of the gain from reform to the dynamic optimal policy. The gains are due to substantial increases in both efficiency and equity. When age dependence is restricted to be Pareto improving, the welfare gain is nearly as large.

Key words: Dynamic optimal taxation, Tagging, Age dependence

JEL Codes: H20, H21

1. INTRODUCTION

The fundamental challenge for tax policy design, as first posed rigorously by Mirrlees (1971), is to use observable information to redistribute income to the low skilled and pay for public goods without discouraging work effort. The three decades following Mirrlees' initial work saw substantial progress towards meeting this challenge in his canonical model economy,[1] but a recent surge of research has taken up the challenge in more complicated settings. This research, beginning with Golosov, Kocherlakota and Tsyvinski (2003), has focused on dynamic economies in which individuals' wages may change over their lifecycles. It has shown that dynamic optimal taxation depends, other than in special cases,[2] on a taxpayer's detailed income history.

The complexity of dynamic optimal taxes has raised concerns over the practicality of the model's recommendations and has generated interest in capturing some of their gains through more limited and pragmatic policy reforms.[3] Following early, related work by Guesnerie (1977)

1. Saez (2001) is the leading recent study of the standard Mirrlees model.

2. Albanesi and Sleet (2006) shows that history dependence can be replaced with dependence on current wealth if shocks to skills are i.i.d. For general shock processes, Kocherlakota (2005) shows that history dependence is required. Grochulski and Piskorski (2006) extend the analysis to a model with human capital and confirm that history dependence is generally required for an optimum.

3. The Mirrleesian framework generally assumes a benevolent tax authority, avoiding practical complications due to self-interested politicians such as those studied by Yared (2010).

and Feldstein (1976), these proposals are collectively called "partial reforms." Recent examples include Golosov and Tsyvinski (2006), who characterize optimal policy towards permanent disability shocks; Huggett and Parra (2010), who analyse an optimal social security benefit function; and Farhi and Werning (2009), who perturb the allocation of consumption over time to satisfy an optimality condition while leaving labour allocations unaltered.[4]

In this paper, I study a feasible but powerful partial reform that responds to changes in wages over the lifecycle: age-dependent labour income taxes. Using data from the U.S. Panel Study of Income Dynamics (PSID), I find that age dependence yields a welfare gain equivalent to between 0·6% and 1·5% of aggregate annual consumption and captures more than 60% of the potential gain from full reform to the dynamic optimal policy. As an application of the dynamic Mirrleesian framework that would be simple to implement in modern tax systems, age dependence provides the potential for a policy reform that bridges the gap between theoretical tax design and practical tax policymaking.

In fact, age-dependent taxation has begun to attract attention from prominent academics and policymakers. In the recent Mirrlees Review,[5] a leading summary of the current state of tax theory as it relates to policy, Banks and Diamond (2010) argue that age dependence is one of the most promising areas for the near-term reform of developed-country taxes. During the 2008 U.S. Presidential election, then-candidate Barack Obama proposed to "eliminate all income taxation of seniors making less than $50,000 per year."[6]

This paper makes three contributions to the study of age-dependent taxation.

First, I characterize the partial reform of age-dependence analytically, deriving results along both the intratemporal and the intertemporal behavioural margins that I compare to both age-independent and dynamic optimal taxation. Age dependence tailors the schedule of marginal labour income tax rates to the distribution of wages at each age, avoiding inefficient distortions to labour effort. It also transfers resources over individuals' lifecycles, ameliorating any limits on private saving and borrowing. When agents can save and borrow, age-dependent marginal rates may also help discourage strategies, such as oversaving, that undermine the tax system.

Second, I calibrate and simulate optimal policy in both deterministic and stochastic settings using data on heterogeneous individual wage paths. While much of the dynamic optimal taxation literature has used illustrative numerical simulations, this paper's simulations are designed to bring an empirical realism such as that found in Saez's (2001) study of the static Mirrlees model to this literature. 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, I show that age dependence, despite being an easily implemented partial reform, yields a large welfare gain and that this gain represents more than 60 % of the potential gain from full reform to the dynamic optimal policy. This illustrates the potential for powerful policymaking lessons from dynamic optimal taxation analysis even if the full package of recommended

4. Note that the partial reform in Farhi and Werning (2009) is complementary to this paper's, which focuses on individuals' choices of consumption and labour at different ages over the lifecycle but leaves individuals free to transfer resources intertemporally. If I allow for linear savings taxation, the power of age-dependent taxes increases (see Section 3).

5. The Mirrlees Review is the modern counterpart to the Meade Report of 1978.

6. Intentional age dependencies in existing policy tend to be focused on tax breaks for the elderly, such as in Singapore and Australia. More common, implicit age dependencies often work at cross purposes. In the U.S., *e.g.* payroll taxes place a larger effective burden on the young (see Feldstein and Samwick, 1992), while disability insurance places a larger effective burden on the old. Such collections of age dependencies are unlikely to be optimal.

reforms is infeasible. I decompose the welfare gain from age dependence into three main components. Increased efficiency explains between 20% and 40% of the gain, while more equitable and timely allocations of consumption and required income explain between 30% and 40% each. Strikingly, the welfare gains from age dependence are nearly as large when I constrain it to be Pareto-improving. As Guesnerie (1977) noted in early work on partial reforms, a particularly strong political and normative case can be made for reforms that yield Pareto improvements.

The idea of having taxes depend on age has been around for some time and was even mentioned in passing by Mirrlees (1971). The most well-known prior treatment of the subject in the Mirrleesian framework is an elegant paper by Kremer (2002). In a classic Mirrleesian setting, Kremer demonstrates that marginal income taxes not conditioned on age are unlikely to be optimal. He also uses evidence on some key parameters to argue that marginal tax rates should be lower for young workers, a claim supported and clarified by the theoretical analysis and numerical simulations I perform below.

This paper builds on Kremer's analysis in three ways, each of which was mentioned in his paper as an important challenge for future research. First, Kremer's analysis is in a static setting. The dynamic setting of this paper allows me to derive the implications of new factors, such as the ability of individuals to move resources intertemporally, and to show that the power of age dependence is robust to them. Second, Kremer focuses on showing that first-order conditions imply age dependence is desirable, while this paper's numerical simulations yield insights on optimal age-dependent average tax rates, intertemporal distortions, and marginal tax rates. Third, I calculate the large welfare gains from age dependence relative to age-independent taxes and full reform to dynamic optimal taxes. While Kremer argued that age dependence was likely to raise welfare, this paper estimates its potential gain, making a powerful case for it as a policy reform.

More recently, two papers address components of the design of age-dependent taxes.

Blomquist and Micheletto (2003) take a theoretical approach and analyse a simple dynamic economy in which all individuals start with the same skill level when young and have one of two skill levels when older. Of particular value in their analysis is the focus on whether age dependence can be Pareto-improving. I explore that possibility in detail in Section 5 of this paper. Though tractable, their model prevents Blomquist and Micheletto from deriving results on the pattern of distortions across types at different ages and on intertemporal distortions when there are multiple types at each age. I am able to address these two important dimensions of optimal policy. I also analytically compare age dependence to not only age-independent policy but also the fully optimal policy.[7]

A second recent paper, Judd and Su (2006), takes a numerical approach. Their analysis of age dependence is part of their broader study of multidimensional screening in optimal taxation, which they show can substantially complicate the mathematical and computational analysis of the problem. Importantly, they introduce to the literature numerical methods that address those complications. They apply their methods by, among other examples, simulating age-dependent and age-independent policy for an illustrative set of deterministic wage paths. In contrast to their simulations, I use individual-level data to calibrate a model economy with a realistic set of deterministic and stochastic wage paths. This generates policy-relevant results on the design of age-dependent taxes, and it allows me to estimate and decompose the potential welfare gain from age dependence. I study the differences in age-dependent policy when individuals can and cannot save or borrow, and I compare age-dependent and dynamic optimal taxation.

7. A more recent working paper by these authors and others, Bastani *et al.* (2010), calibrates an overlapping generations model with retirement and highlights the relationship between the welfare gains from age dependence and the extent of earnings inequality.

In addition, several other recent papers suggest that age dependence may have a powerful role to play in optimal taxes in a variety of theoretical settings. Farhi and Werning (2010) consider a dynamic Mirrleesian environment and show that variation in taxation over the lifecycle is an important component of optimal taxation. In addition to characterizing optimal allocations, they simulate a linear age-dependent income tax and find that it generates substantial welfare gains over a linear age-independent policy. These results support this paper's findings on optimal non-linear age-dependent taxes. Golosov, Troshkin and Tsyvinski (2010) also provide a valuable characterization of optimal taxes in a dynamic Mirrleesian model and discuss ways in which age dependence relates to the implementation of optimal policy. Conesa, Kitao and Krueger (2009) consider a calibrated life cycle model with idiosyncratic risk. They find that capital taxation can serve as a substitute for age-dependent taxes, which by implication would play an important role in the optimal design of taxes.

This paper also adds to a recent literature revisiting and building on the insights of Akerlof (1978), who showed how "tagging," *i.e.* conditioning taxes on personal characteristics, can improve redistributive taxation. Alesina, Ichino and Karabarbounis (2010) study the benefits of gender-dependent taxation, while Mankiw and Weinzierl (2010) explore the case of height-dependent taxation. Though age dependence resembles standard tagging in some ways, it cannot be fully understood with intuition from conventional tagging analysis.[8]

A related literature beginning with Vickrey (1939) has considered the potential benefits of taxing income averaged over a span of time, even an entire lifetime. Liebman (2003) is a modern treatment of that idea. While income averaging has the benefit of taxing based on a more complete estimate of an individual's earnings, it does not tailor marginal distortions to the distribution of earnings at each age, a key benefit of age-dependent taxation. Moreover, lifetime income taxation also depends on age, so it provides little advantage in simplicity.

The paper proceeds as follows. Section 2 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 behaviour 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. The next two sections generalize the baseline model. Section 3 allows individuals to save and borrow, and Section 4 adds stochastic wage paths. Throughout the settings considered in these first three sections, the results on optimal policy and the welfare gains from age dependence are consistent. Section 5 provides discussion, and Section 6 concludes. A technical appendix available at this journal's website contains the proofs of the propositions and supplementary material mentioned in the text.

2. BASELINE ECONOMY

In this section, I analyse age-dependent labour income taxes for a baseline economy characterized by two simplifying assumptions. First, individuals cannot transfer resources across periods: *i.e.*, 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

8. While a standard tag, such as gender, provides information on an individual's expected place in the distribution of lifetime income, age reveals no information by itself. A standard 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 be useful for taxation, age must be combined with data on current income and how incomes at each age relate to lifetime income.

lifetime. These simplifying assumptions allow for clean analytical results that build intuition for the more general models considered in later sections.

2.1. *Set-up*

All individuals live and work for T periods,[9] indexed by $t = \{1, 2, \ldots, T\}$, and are members of the same generation.[10] Individuals are heterogeneous in their ability to earn income over their lifetimes. This ability comes in I types, indexed by $i = \{1, 2, \ldots, I\}$, 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 labour effort, and each individual knows its full lifetime path of wages $\{w_t^i\}_{t=1}^{T}$ at time $t = 1$.[11] Wages are not publicly observable. Labour income y is observable, and it is the product of the wage and unobservable labour effort l, so $y = wl$. There is no capital.

All individuals have the same separable[12] preferences over consumption c and labour effort l, where $l = \frac{y}{w}$. The utility U_t^i for individual i of age t is

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

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

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

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 equation (3) allows us to consider any point along the Pareto frontier.

2.2. *The 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, 2007, for a review).[13]

9. I assume an exogenous date of entry into the labour market. Age dependence could generate even larger welfare gains if it were designed to operate effectively on these incentives.

10. In the technical appendix, I discuss how the analysis generalizes largely unchanged to a setting with overlapping generations, as well as the relationship of calendar-year dependence to age dependence.

11. The extent to which wages are stochastic is the subject of substantial recent research. See Keane and Wolpin (1997), Storesletten, Telmer and Yaron (2001), and Guvenen (2007).

12. I have performed numerical simulations assuming King-Plosser-Rebelo preferences, as discussed in Kimball and Shapiro (2008), which allow for complementarity between consumption and labour. Increasing the degree of complementarity has no effect on the main lessons from the separable case other than raising the potential gains from reform to age-dependent or fully optimal taxes.

13. In the technical appendix, I show how the optimal allocations for each policy scenario derived below using a direct mechanism can be implemented using non-linear tax functions that depend on income, income and age, or income history and age, respectively. The formal approach follows Kocherlakota (2005).

The social planner maximizes social welfare equation (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.

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

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

which says that the lifetime paths of income must fund the lifetime paths of consumption across all types.[14] 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 are inequalities ensuring that each individual chooses the allocation of c_t^i and y_t^i that the planner intended.[15]

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.

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

Problem 1. *(Static Mirrlees (SM): Age-Independent)*
Choose $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$ *to maximize equation* (3) *subject to the feasibility constraint* (4) *and the incentive constraints*

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

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

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 age s.[16] 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.

The partial reform planner in the baseline model solves the following problem:

14. 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.

15. 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.

16. 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.

Problem 2. *(Partial Reform: Age-Dependent)*
Choose $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$ *to maximize equation*(3) *subject to the feasibility constraint (4) and the incentive constraints*

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

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

These incentive constraints reflect the partial reform planner's ability to restrict individuals to age-specific allocations. Now, the planner must guarantee 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, note that the right-hand side of inequality (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 the set of constraints (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 $\{\mu_t^{j|i}\}_t^{i,j}$, 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.

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

Problem 3. *(Full Optimum: Age-Dependent and History Dependent)*
Choose $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$ *to maximize equation* (3) *subject to the feasibility constraint (4) and the incentive constraints*

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

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

These incentive constraints reflect the Full Optimum planner's ability to make, and commit to, history-dependent allocations.[17] 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 inequality (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 (c_t^i, y_t^i) 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.

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 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.

17. 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.

2.3. *Analytical results*

Now, I compare policies along two margins: the intratemporal margin between consumption and leisure and the intertemporal margin between consumption in one period and the next. I evaluate average taxes in Section 2.4.

2.3.1. Intratemporal distortions.

First, I compare the distortions to individuals' choices of how much income to earn.

Definition 2.1 (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 $\tau^{\mathrm{SM}}(i,t)$, $\tau^{\mathrm{PR}}(i,t)$, and $\tau^{\mathrm{FO}}(i,t)$ 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 towards 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.

If wage paths were constant for each individual over the life cycle, $\tau(i,t)$ would be the same for individual i at all ages. Moreover, each policy would use the same set of $\tau(i,t)$ in this economy, as all ages would be identical.[18]

In reality, wage paths are far from constant, and optimal policy will vary intratemporal distortions in response if it is able to do so. As in most Mirrleesian analyses, characterizing intratemporal distortions analytically is difficult, and the general expressions for $\tau(i,t)$ are not available in closed form.[19]

The key lesson is that age-dependent taxes allow policymakers to tailor intratemporal distortions to the wage distribution at each age. In contrast, optimal age-independent taxes distortions must be based on the distribution of wages across all ages.

The clearest example of age-dependent taxes being tailored to age-specific wage distributions builds on the most well-known result from static Mirrleesian tax analysis: the top earner in the economy should face *no* intratemporal distortion.[20] The intuition for the classic result starts with the recognition that an intratemporal distortion has both a cost and a benefit. The cost is that it causes individuals directly affected by the distortion 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. At the top of income distribution, this benefit is

18. I prove these results in an earlier version of this paper, Weinzierl (2008), Proposition 1.

19. In the technical appendix, I show and discuss expressions for these distortions that include the multipliers on the incentive constraints.

20. Diamond (1998) and Saez (2001) show that this result depends on the shape of the wage distribution. With a bounded wage distribution such as that used throughout this paper, the zero top rate result always holds. In the technical appendix, I show that the results presented here have analogues in a model with a wage distribution more similar to that which Diamond and Saez use. While the top rates are not zero at any age for any policy, rates for the high-skilled vary by age, reflecting the different shapes of the skill distributions at each age.

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 2.1 (Top Marginal Distortion). *In the baseline economy,*

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

Proof. In technical appendix. ‖

The first part of this proposition states that the highest wage earner at each age faces a non-positive, possibly negative, intratemporal distortion whenever the social planner can condition taxes on age.[21] Though empirically unlikely, suppose an individual has the highest wage within its current age but also has a low lifetime income-earning potential. 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.

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. To see why it holds, consider an individual i^* with the highest wage at all ages. 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. A simple numerical illustration of this is provided in the technical appendix for the interested reader.

The second part of the proposition thereby highlights the fundamental limitation of the Static Mirrlees policy: the inability to hold individuals to age-specific tax schedules. The effects of this limitation ripple throughout the wage distribution, and we will its impact on social welfare in the numerical simulations below. Before turning to the simulations, however, I discuss a second set of analytical results, these on the intertemporal consumption margin.

2.3.2. Intertemporal distortions. Consider an individual's problem of maximizing lifetime utility in equation (2) given a wage path $\{w_t^i\}_{t=1}^T$ and the lifetime budget constraint $\sum_{t=1}^T R^{T-t}(y_t^i - c_t^i) = 0$. Continue to assume $\beta R = 1$. This individual's optimal choice of consumption satisfies, for each $(t, t+1)$ pair:

$$u'(c_t^i) = u'(c_{t+1}^i). \tag{9}$$

21. 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 labour supply elasticity differences can justify negative top distortions. As noted in Section 1, Judd and Su also perform an illustrative simulation of age-dependent taxation.

This is the familiar undistorted intertemporal Euler condition that sets the marginal utility from consumption equal across periods. Recall that individuals cannot save or borrow in the baseline economy. Thus, when studying this margin, we are interested in the extent to which the planner distorts allocations away from equation (9).

If wage paths were constant, then each policy (Static Mirrlees, Partial Reform, and Full Optimum) would satisfy equation (9).[22] Intuitively, if each age is a replica of the next, the allocations to each individual will be the same at each age.

In more realistic settings, with wage paths that are not constant over the lifecycle, we can show how each policy responds along the intertemporal margin.

The Full Optimum planner's allocations satisfy equation (9) for all $i \in \{1, 2, \dots, I\}$ and t, $t+1 \in \{1, 2, \dots, T\}$, a classic Atkinson and Stiglitz (1976) result. In words, the Full Optimum policy does not distort intertemporal allocations.[23] This result depends on the Full Optimum planner's ability to use history-dependent allocations.

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

$$u'(c_t^i) = \left(\frac{\pi^i \alpha^i + \sum_{j=1}^{I} \mu_{t+1}^{j|i} - \sum_{j=1}^{I} \mu_{t+1}^{i|j}}{\pi^i \alpha^i + \sum_{j=1}^{I} \mu_t^{j|i} - \sum_{j=1}^{I} \mu_t^{i|j}} \right) u'(c_{t+1}^i). \tag{10}$$

The ratio in parentheses in equation (10) 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 equation (10) 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. Nevertheless, the Partial Reform allocations do improve on the Static Mirrlees allocations, as they satisfy a "symmetric inverse Euler equation:"

Proposition 2.2 (Symmetric Inverse Euler). *Let the lifetime utility function for all $i \in \{1, 2, \dots, I\}$ be defined by equation (2). Then, the solution to the Partial Reform planner's problem satisfies:*

$$\sum_{i=1}^{I} \frac{\pi^i}{u'(c_t^i)} = \sum_{i=1}^{I} \frac{\pi^i}{u'(c_{t+1}^i)}, \tag{11}$$

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

Proof. In technical 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 an achievement of Partial Reform that the Static Mirrlees planner cannot replicate.[24] Because it cannot restrict individuals to age-specific tax schedules, the Static

22. I show that this in an earlier version of this paper, Weinzierl (2008), as Proposition 3. Werning (2007*a*) proves a similar result for the optimal dynamic policy.

23. 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 Section 4, I analyse a generalization of the baseline model where wages are stochastic.

24. I omit the Static Mirrlees planner's intertemporal result for brevity, but it is an intuitive modification of the Partial Reform result.

Mirrlees planner cannot make these efficient transfers across ages.

2.4. *Numerical results*

In this section, I calibrate the dynamic optimal tax problems specified above to data from the U.S. PSID, simulate and describe policy, and quantify the welfare impacts of reform.

2.4.1. Data and parameters. Simulating the model requires representative lifetime wage paths $i = \{1, 2, \ldots, I\}$. The construction of the data set required to calculate these paths can be divided into four steps.[25] First, I focus on 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 the heads of households. Second, I calculate reported real wages for each observation by dividing reported labour income by reported hours (a potentially noisy measure of the wage but the best one available) and inflating or deflating the data with the Consumer Price Index to put all wages in 1999 dollars.[26] Third, I remove potentially problematic observations by eliminating all those for which reported annual hours were less than 500 or greater than 5824, for which reported labour 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. Fourth, I limit the sample to the ages 25 through 55, as that is the range over which the numerical simulations below will be run. After these adjustments, the data set contains approximately 111,000 observations on just over 10,700 individuals with an average of 10·4 years observed per person.

Once the data set is assembled, the next task is to identify $i = \{1, 2, \ldots, I\}$ types of individuals and $t = \{1, 2, \ldots, T\}$ age ranges. A natural choice for types is education level. While other techniques[27] may be used, education divides the sample in a transparent and intuitive way. Moreover, the disparities in wages across education groups are well documented empirically (see, *e.g.* Card, 1991). I divide the sample into three education groups, $i = \{1, 2, 3\}$, as follows: (1) less than high school; (2) at least high school but less than college; and (3) at least college. These groups make up 23·3 %, 56·0 %, and 20·7 % of the sample.

For $t = \{1, 2, \ldots, T\}$, I choose the following age ranges to capture break points in the effect of age on wages in the data: (1) 25–32 years old; (2) 33–40 years old; and (3) 41–55 years old.[28] The resulting wage paths are shown in Table 1. These wage paths have shapes consistent with those described in Card (1991).

I assume the period utility function

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

and set $\sigma = 3$, which implies a constant consumption elasticity of labour supply of 0·5. The results described below are robust to alternative values for this elasticity and the curvature of

25. The process detailed in this paragraph follows Fullerton and Rogers (1993) and recent updates of their work such as Altig *et al.* (2001) and Diamond and Tung (2006). I thank John Diamond and Joyce Tung for providing helpful advice on the construction of the data set.

26. 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.

27. An earlier version of this paper, Weinzierl (2008), used latent class analysis to extract a set of representative paths from the data, *e.g.*

28. A regression of wages on dummy variables for each age shows a substantial break between ages 40 and 41. No break points between 25 and 40 are apparent in the data, so I divide that range at its midpoint.

TABLE 1
Wage paths by education group

		Education level		
		Low	Middle	High
Age range	25–32	11·56	14·76	19·86
	33–40	12·63	17·35	27·17
	41–55	13·51	19·10	32·80
	Population proportion	0·23	0·56	0·21

Source: PSID core sample household heads 1968–2001.
Note: Wage levels in 1999 U.S. dollars.

consumption utility.[29] I use an annual gross rate of return of 5%, a discount rate such that $\beta R = 1$,[30] and constant Pareto weights $\alpha^i = 1{\cdot}00$ for all i. These Pareto weights imply a pure Utilitarian social welfare function in which the weight on each type's utility in the social welfare function is equal to its proportion of the population (see expression (3)). More redistributive assumptions increase the power of age-dependent taxes to raise welfare.

2.4.2. Simulation results. I focus on four outputs from the numerical simulations: intratemporal distortions, average tax rates, intertemporal distortions, and welfare.

Intratemporal distortions. Intratemporal distortions are shown in Table 2.

The most striking difference between the policy scenarios is the treatment of the high-income young.[31] While both the Partial Reform and the Full Optimum policies have these intratemporal distortions decrease as education (and income) rises for young workers, the Static Mirrlees has them increase. The contrast is more subtle for workers aged 33–40, where the Static Mirrlees planner assesses roughly flat distortions vs. the more sophisticated planners who, again, have distortions decrease with income. The explanation for these patterns is that the age-independent policy of the Static Mirrlees planner cannot distinguish young, highly educated workers from older, middle-educated workers, so it must treat them as it does these lower earners.[32] These patterns are the numerical counterparts to Proposition 1.

A second insight to come from these simulations is the degree to which more sophisticated policies reduce distortions overall in the economy. The final column of Table 2 shows that the

29. With this utility function, the constant consumption elasticity is equal to the Frisch elasticity. Though evidence on this parameter varies widely, 0·5 lies between smaller, microeconometric estimates and larger, macroeconomic estimates. To check robustness, I vary the elasticity of labour supply from 0·25 to 3·0 and the curvature of consumption utility from 0·75 to 3·0. These parameters yield welfare gains of reform from the Static Mirrlees to the Partial Reform policies from 1·1% to 2·8% of total consumption. The baseline result is 1·45%.

30. The results of the simulations are insensitive to the value of R. If $R\beta < 1$, the welfare gains from Partial Reform increase. If $R\beta > 1$, the welfare gains decrease but only slightly. The lessons for the design of optimal policy are unchanged by these variations.

31. This result does not depend on the use of a bounded ability distribution for the numerical simulations. See the technical appendix for an analysis with unbounded ability distributions that vary by age. There, as here, intratemporal distortions on the high-ability young are lower with age-dependence because of the differences in the distributions of ability across ages.

32. 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.

TABLE 2
Intratemporal distortions in baseline model

	Age range	Education level Low	Middle	High	Average
Static Mirrlees	25–32	0·07	0·11	0·19	**0·113**
	33–40	0·12	0·10	0·12	
	41–55	0·25	0·10	0·00	
Partial Reform	25–32	0·09	0·05	0·00	**0·084**
	33–40	0·16	0·09	0·00	
	41–55	0·20	0·13	0·00	
Full Optimum	25–32	0·12	0·07	0·00	**0·075**
	33–40	0·14	0·08	0·00	
	41–55	0·14	0·09	0·00	

unweighted average marginal distortion across all types and ages is 0·113 in the Static Mirrlees policy, 0·084 in the Partial Reform policy, and 0·075 in the Full Optimum policy.[33] Moreover, these differences in levels are in addition to the improved pattern of distortions in the more sophisticated policies. The combination of lower average distortions and better-designed distortions encourages labour 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. The average tax schedule is necessarily the same for all age groups under the Static Mirrlees policy. In the Partial Reform and Full Optimum policies, workers face separate average tax schedules in each phase of their career. In particular, Figure 1 shows that workers face lower average taxes in their early working years than later in life under the optimal age-dependent tax policy. The figure also shows the Static Mirrlees average tax schedule, which lies in the middle of the age-dependent schedules.

Figure 1 shows that the magnitude of the difference across ages in age-dependent average tax schedules can be substantial. For example, in the Partial Reform policy, a worker earning $44,000 faces an average tax rate early in his career that is approximately 20% points less than when he is in his peak earning years. The Full Optimum schedules (not shown) resemble the Partial Reform schedules.

Why do the more sophisticated planners use lower average taxes on individuals when they are young? The data show wages rising over these age ranges in all education 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' early working years. The Static Mirrlees planner cannot do so because it cannot target lower average taxes at an age group.

Intertemporal distortions. Next, Table 3 shows the intertemporal distortions under each policy. Here, the main advantage of history dependence is clear, as only the Full Optimum policy provides fully smoothed consumption to all workers.

33. If we weight distortions by income, the income-weighted average marginal distortion falls from 0·104 in the Static Mirrlees policy to 0·071 in the Partial Reform policy and 0·060 in the Full Optimum workers face policy.

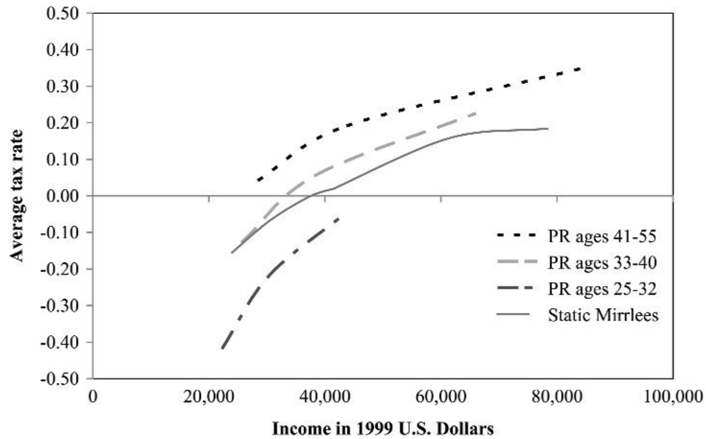

FIGURE 1
Average tax rates in baseline model

TABLE 3
Intertemporal distortions in baseline model

	Age range	Education level Low	Middle	High	Average
Static Mirrlees	25–32	1·06	1·14	1·27	**1·12**
	33–40	1·01	1·09	1·23	
	41–55	n/a	n/a	n/a	
Partial Reform	25–32	0·91	0·98	1·12	**0·99**
	33–40	0·94	0·98	1·08	
	41–55	n/a	n/a	n/a	
Full Optimum	25–32	1·00	1·00	1·00	**1·00**
	33–40	1·00	1·00	1·00	
	41–55	n/a	n/a	n/a	

Note: Cells show $u'(c_t)/u'(c_{t+1})$. A value of 1·00 indicates no distortion.

While not smoothing for each worker, the Partial Reform policy smooths in aggregate across ages. This improves on the Static Mirrlees policy, which we can see in Table 3 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. For each policy, Table 4 lists overall social welfare and lifetime utility by type. Because they are in units of welfare, the welfare and utility values in Table 4 are meaningfully ordinally, not cardinally.

Table 4 yields three main findings.

First, the increase in welfare due to age dependence alone is large, equivalent to a 1.5% increase in aggregate consumption or roughly \$150 billion in current U.S. dollars, annually. In other words, if the Static Mirrlees planner received a windfall enabling it to increase each

TABLE 4
Welfare in baseline model simulation

	Total social welfare (Utils)	Welfare gain over SM (in consumption equivalents)†	Utility for education groups (utils)		
			Low	Middle	High
Static Mirrlees	4·957	n/a	4·56	4·94	5·46
Partial Reform	4·986	1·45%	4·69	4·97	5·37
Full Optimum	4·989	1·62%	4·70	4·97	5·38

†Supplement consumption in the Static Mirrless policy by this amount to yield the welfare of the target policy.

individual's consumption by 1·5% while holding labour 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 89% of the gain from reform to the Full Optimum. The additional welfare gain from history dependence is 0·17% of aggregate consumption, so that the total gain due to reform from the Static Mirrlees to the Full Optimum is 1·62% of aggregate consumption.

Finally, this welfare gain is concentrated among the low skilled, so that the Partial Reform and Full Optimum policies achieve more egalitarian (and similar) distributions of lifetime utility across types. The Static Mirrlees policy provides greater utility to the high-education group, while the more sophisticated policies increase utility among individuals in the low-, and middle-education groups.[34]

Figure 2 shows a decomposition of the welfare gain from Partial Reform. To calculate this decomposition, I consider five separate perturbations of the Static Mirrlees allocations. These perturbations correspond to the five channels labelled in Figure 2, and they collectively exhaust the improvements from Partial Reform.[35] For each channel, I adjust the allocations of the Static Mirrlees policy in a specific way (explained below) and then calculate the windfall required to achieve the welfare obtained under the Partial Reform policy when starting from this adjusted allocation. Comparing this windfall and the windfall required to obtain Partial Reform welfare starting from the original Static Mirrlees policy allows me to measure the share of the total gain attributable to that channel.

Efficiency gains explain one-sixth of the welfare gain from age dependence. In Table 2, we saw that the average marginal distortion to labour effort is lower with age dependence. This encourages more effort, and output is 2·2 % higher under the Partial Reform policy than under the Static Mirrlees policy. To estimate the welfare impact of this increase, I raise the Static Mirrlees allocations of income and consumption by 2·2% at each age. The welfare gain of reform from this modified version of the Static Mirrlees policy to the Partial Reform policy is 1·2 % of aggregate consumption. Thus, the increased output due to efficiency gains accounts for 16% of the total welfare gain from age dependence.

34. 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.

35. When considered sequentially, the five perturbations explain 100% of the gain from Partial Reform. Figure 2 shows their impacts when considered in isolation. The sum of the gains is slightly less than 100% because the interaction of these channels generates additional gains.

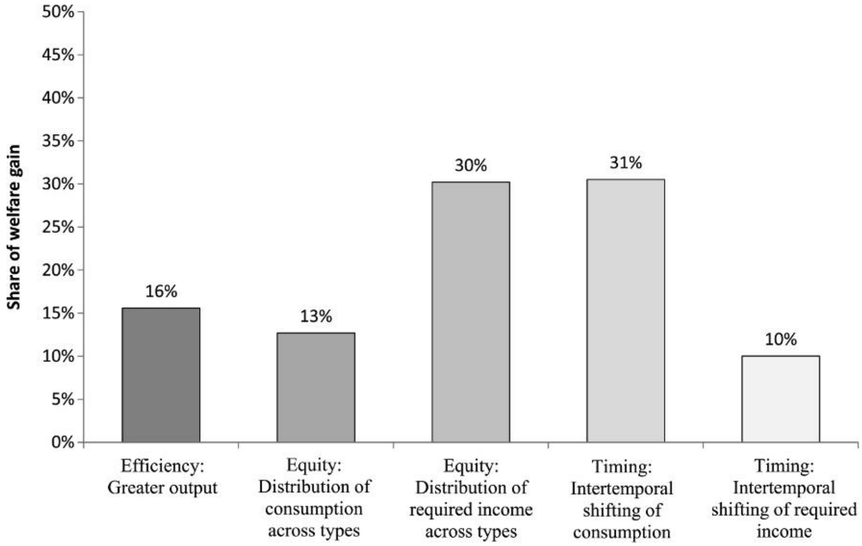

FIGURE 2
Components of welfare gain from Partial Reform in baseline model

An additional 43% of the welfare gain from age dependence is due to the Partial Reform planner allocating consumption to individuals with higher marginal utilities of consumption and requiring income from individuals with lower marginal disutilities of income. To measure the gains from the allocation of consumption, I scale each individual's consumption path in the original Static Mirrlees allocation to equal the same share of total consumption (in present value) as under the Partial Reform policy. Because utility from consumption is concave and the Partial Reform policy is more egalitarian than the Static Mirrlees policy, this reallocation of consumption across types yields 13% of the welfare gain from age dependence. To measure the gains from the allocation of required income, I scale the income required from each type to equal the share required in the Partial Reform. This accounts for 30% of the welfare gain from age dependence.

Finally, age dependence allows for intertemporal shifting of consumption and required income. The gains from smoother consumption can be calculated by allocating the present value of each individual's consumption across ages as it is in the Partial Reform policy. This change accounts for 31% of the gain from age dependence. Allocating the present value of required income for each type across ages as in the Partial Reform policy yields the final 10% of the welfare gain from Partial Reform.

Why does Partial Reform capture nearly 90% 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 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, the wage paths have similar slopes, thereby reducing the benefit from history dependence. In Section 4, I consider an extension of the baseline model that incorporates more heterogeneity in wage paths by allowing wage paths to be stochastic rather than deterministic. As discussed there, the Partial Reform policy nevertheless continues to capture a large majority of the gain from the Full Optimum.

3. 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.[36] This affects the marginal trade-offs facing individuals at each age and, therefore, the optimal policy towards 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, as 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% 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.[37]

3.1. *The 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, these bundles are of pre-tax income and after-tax income, not consumption. For brevity, I relegate the formal statement of the planner's problems to the technical appendix and discuss their key components here.[38]

Private saving and borrowing complicate the incentive constraints facing the planner in the two policies that lack history dependence. In the Static Mirrlees policy, 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 by saving and borrowing. As in the baseline model, the Partial Reform planner has the advantage of conditioning taxes on age, simplifying its problem, but individuals now have the ability to smooth consumption on their own. Meanwhile, the planner's problem for the Full Optimum scenario is unchanged from the

36. 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.

37. If the Partial Reform planner were allowed to choose a flat rate on capital income, it would choose a substantially higher rate to avoid the dynamic incentive problems that private saving creates. The welfare gains from Partial Reform would therefore increase.

38. Also in the technical appendix, I show that the allocations derived below using a direct mechanism are implementable through a non-linear tax system, as in the Baseline case. I also show that the intratemporal distortions can be mapped to equivalent marginal taxes on earned income. The key to this result is that individuals smooth any extra after-tax earned income, so the utility value of a dollar of after-tax income is equal to the marginal utility of consumption for the individual, and the standard expression for the intratemporal distortion is thus equal to the marginal tax rate.

baseline model. Because it can link allocations across ages, the Full Optimum planner spreads the after-tax income received by an individual over its lifetime optimally, leaving the individual's optimal choice undistorted.

How do these more complicated incentive constraints affect policy? In the baseline model, I analysed policy along two margins: the intratemporal and intertemporal. In this model, with private saving and borrowing, the second of these margins goes undistorted in all three policy scenarios.[39] Therefore, I focus my analysis 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 $w_1^{\mathrm{H}} > w_1^{\mathrm{L}}$ and $w_2^{\mathrm{H}} > w_2^{\mathrm{L}}$. Finally, assume that the utility function takes the simple form used in the numerical simulations of the baseline model:

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

The Full Optimum policy's treatment of the high-skilled worker is unchanged by private saving and borrowing. Thus, we know from Proposition 2.1 *(Top Marginal Distortion)* that the high-skilled worker in this example will face a zero intratemporal distortion at both ages in the Full Optimum policy.

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

$$\tau^{\mathrm{PR}}(y_1^{\mathrm{H}})=\frac{\mu^{\mathrm{HL|HH}}}{\pi^{\mathrm{H}}\alpha^{\mathrm{H}}+\mu^{\mathrm{LH|HH}}+\mu^{\mathrm{LL|HH}}}\left(\frac{c_1^{\mathrm{HH}}}{c_1^{\mathrm{HL}}}-1\right), \tag{12}$$

where I use $\mu^{ij|kk}$ to denote the Lagrange multiplier on the incentive constraint preventing type k from claiming the series of wages $\{w_1^i,w_2^j\}$, and all other notation is as in the baseline model. Now, the distortion is positive if $\mu^{\mathrm{HL|HH}}>0$ and $\frac{c_1^{\mathrm{HH}}}{c_1^{\mathrm{HL}}}>1$. These conditions are that the incentive constraint preventing the high-skilled worker from claiming the $\{w_1^{\mathrm{H}},w_2^{\mathrm{L}}\}$ wage path binds, and that a high-skilled worker claiming the $\{w_1^{\mathrm{H}},w_2^{\mathrm{L}}\}$ wage path subsidizes its consumption in the second period with savings from the first. Intuitively, the Partial Reform policy distorts the young, high-skilled intertemporal margin to discourage him or her from earning the extra income that funds oversaving.[40]

The logic works in the other direction as well. If old, high-skilled workers are tempted to overborrow and work less when young, the planner can make such cheating more costly by distorting the high-skilled worker's intratemporal margin when old. The condition for the intratemporal distortion on the high-skilled old worker is analogous:

$$\tau^{\mathrm{PR}}(y_2^{\mathrm{H}})=\frac{\mu^{\mathrm{LH|HH}}}{\alpha^{\mathrm{H}}\pi^{\mathrm{H}}+\mu^{\mathrm{HL|HH}}+\mu^{\mathrm{LL|HH}}}\left(\frac{c_2^{\mathrm{HH}}}{c_2^{\mathrm{LH}}}-1\right)$$

so that if the worker is tempted to borrow and mimic the low-skilled worker when young, the planner levies a positive intratemporal distortion on the older worker to raise its marginal utility

39. The Static Mirrlees and Partial Reform policies cannot distort the intertemporal margin by assumption, as discussed above. The Full Optimum chooses not to, as proven in Section 2.

40. Recall the caveat noted before that, in multidimensional screening problems such as this, the multipliers may not be uniquely valued. This was shown in Judd and Su (2006). Nevertheless, all that is required for this result is that the relevant incentive constraint bind.

TABLE 5
Intratemporal distortions in Case 2 model

	Age range	Education level: Low	Education level: Middle	Education level: High	Average
Static Mirrlees	25–32	0·09	0·10	0·11	
	33–40	0·14	0·03	0·10	**0·096**
	41–55	0·07	0·16	0·04	
Partial Reform	25–32	0·11	0·07	0·06	
	33–40	0·13	0·09	0·05	**0·087**
	41–55	0·14	0·10	0·05	
Full Optimum	25–32	0·12	0·07	0·00	
	33–40	0·14	0·08	0·00	**0·075**
	41–55	0·14	0·09	0·00	

of consumption and discourage that deviation. Note that it is possible to have positive distortions on the high-skilled worker in both periods.[41]

3.2. *Numerical results*

I use the same data and parameters as in the baseline model to simulate this model.

The intratemporal distortions in each policy are shown in Table 5.

As in the baseline model, the young with middle and high education levels continue to face larger distortions to labour supply in the age-independent Static Mirrlees than in the more sophisticated policies. Unlike in the baseline model, however, the highest-earning workers at each age face a positive intratemporal distortion in the Partial Reform. As discussed above, this reflects the Partial Reform policy's use of intratemporal distortions to discourage the deviation strategy in which individuals plan to use income earned at one stage of their career to supplement consumption at another in which they work less and claim a more generous tax treatment.

Also as in the baseline model, the use of intratemporal distortions in general decreases as the sophistication of policy increases. The final column of Table 6 shows that the unweighted average marginal distortion is 0·096 in the Static Mirrlees policy, 0·087 in the Partial Reform policy, and 0·075 in the Full Optimum policy.[42] As in the baseline, the combination of lower average distortions and a better-designed pattern of distortions encourages labour effort under the more sophisticated policies, raising total output and the efficiency of the economy.

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 is equivalent to another that transfers resources lump-sum across ages. Thus, adding private saving and borrowing does not contradict the recommendation from the baseline case to have the young face a lower average tax schedule.[43]

41. The Inverse Euler Equation, first noted in Rogerson (1985) and explored by Golosov, Kocherlakota and Tsyvinski (2003), suggests that optimal policy distorts the after-tax return to saving to counteract the temptation people face in the presence of skill shocks to oversave and falsely claim a low skill level. Though, by assumption, the Partial Reform planner has no way to tax savings, it uses intratemporal distortions to try to achieve the same results.

42. When we weight distortions by income these are 0·092 in the Static Mirrlees policy, 0·080 in the Partial Reform policy, and 0·060 in the Full Optimum policy.

43. I am grateful to Ivan Werning for suggesting this result.

TABLE 6
Welfare in Case 2 model simulation

	Total social welfare (Utils)	Welfare gain over SM (in consumption equivalents)†	Utility for education groups (utils)		
			Low	Middle	High
Static Mirrlees	4·972	n/a	4·55	4·95	5·50
Partial Reform	4·984	0·58%	4·65	4·96	5·43
Full Optimum	4·989	0·87%	4·70	4·97	5·38

†Supplement consumption in the Static Mirrlees policy by this amount to yield the welfare of the target policy.

Finally, the Partial Reform continues to capture a large absolute welfare gain and a substantial share of the potential gains from more comprehensive reform. Table 6 shows social welfare and the lifetime utility for each type under the three policies.

Reform from the Static Mirrlees policy to the Partial Reform policy yields a gain of 0·58% of aggregate consumption, less than half as much in the baseline model but still a sizable gain. Because the Full Optimum planner is better able to respond to the new incentive problems introduced by private saving and borrowing, the gain from Partial Reform makes up a smaller share of the 0·87% potential gain from full reform. Nevertheless, it captures more than two-thirds (67%) of the gains from full reform. More sophisticated capital taxation would magnify the power of age dependence.[44]

As in the baseline model, the Partial Reform's greater overall welfare is also shared more equally among the individuals in the population. Table 6 shows that the Partial Reform policy produces a more egalitarian distribution of utility than does the Static Mirrlees, though not as egalitarian as the Full Optimum.

Figure 3 shows a decomposition of the welfare gain from Partial Reform into the same components as in the baseline case: improvements in efficiency, equity, and intertemporal shifting.

The most noticeable difference between Figure 2 from the baseline case and Figure 3 is that intertemporal consumption smoothing has no effect in the latter because individuals can perfectly smooth consumption privately. Private saving and borrowing also explains the increased role for the distribution of consumption in this model. When individuals can smooth consumption privately, the level of consumption each individual controls over his or her lifetime becomes a key determinant of welfare. Therefore, the Partial Reform policy's ability to provide higher levels of consumption to the low skilled generates a large share of the gains from reform.

4. CASE 3: MODEL WITH STOCHASTIC WAGE PATHS

In this section, I relax the second main restriction in the baseline model by allowing wage paths to be stochastic rather than deterministic. Thus, there is now substantially more heterogeneity in wage paths than in the previous models, and 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' labour supply and saving and borrowing behavior, with important implications for dynamic optimal policy. As in Case 2, individuals can save and borrow in this model.

44. In a simulation with a linear capital tax rate of 15%, roughly that of the U.S. on some forms of capital income, the welfare gain of Partial Reform is 0·60% of aggregate consumption, a slight increase over the gain without a capital tax.

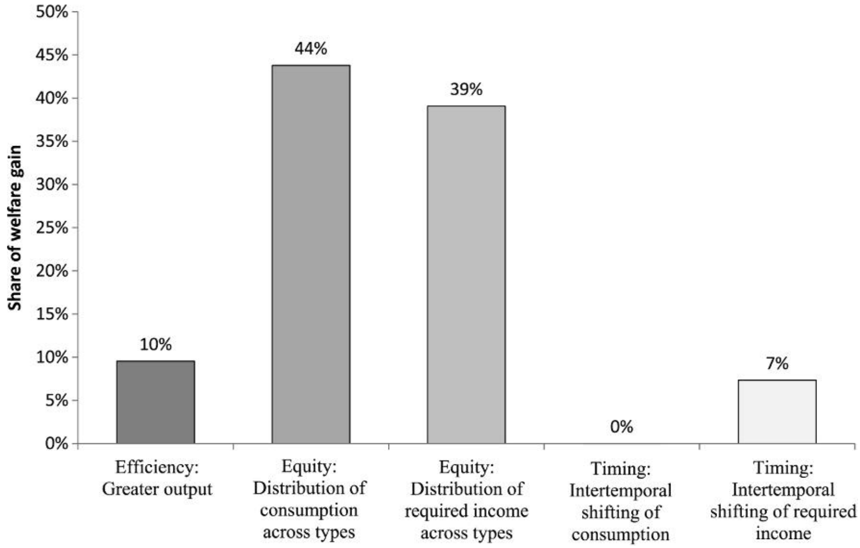

FIGURE 3
Components of welfare gain from Partial Reform in Case 2

4.1. *The planner's problem in three policy scenarios*

Continuing the approach of the previous sections, I consider a social planning problem for each policy. As in Case 2, each planner specifies a menu of pre-tax income and after-tax income pairs to maximize social welfare subject to feasibility and incentive constraints.[45] The Static Mirrlees and Partial Reform planners can neither tax nor subsidize private saving or borrowing, while the Full Optimum policy can be history-dependent.[46]

The novel element of this section's analysis is the stochastic nature of wages. 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.

The combination of private access to capital markets and wage stochasticity makes it difficult to characterize policy analytically. After describing a few key characteristics of policy,

45. With stochastic wages, the social welfare weights α^j are assigned to different wage paths rather than individuals. In the numerical simulations below, however, I assume that weights are constant across all paths (*i.e.* the planner is a pure Utilitarian). See the technical appendix for further discussion.

46. 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.

therefore, this section will move directly to numerical analysis of the policies. Full statements of the planner's problems can be found in the technical appendix.[47]

By itself, adding stochasticity in wages to the baseline case (Section 2) would change nothing for the Static Mirrlees and Partial Reform planners. For them, incentives must be met age-by-age rather than over a lifetime, so their policies are always designed as if wages were stochastic. For individuals, the addition of stochasticity has no effect on their incentives if they cannot transfer resources between periods and utility is time separable.

In this model, however, when individuals can save and borrow, stochasticity handicaps the history-independent planners. For example, if the Static Mirrlees or Partial Reform planners wanted to provide support to an older worker with a suddenly low-ability level, they would not be able to condition that support on past wages. Providing unconditional support would tempt high-ability workers to oversave when young and pretend to receive a bad shock when old, so the history-independent policies must provide less support to all low-ability older workers.

Regardless of whether individuals can save and borrow, the Full Optimum planner's problem is affected by stochasticity because its incentive constraints must guarantee that individuals would rather reveal their true wage path *age-by-age* rather than any other path, taking into account that individuals know the true transition matrices. To provide these incentives, the Full Optimum planner distorts the intertemporal margin. 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.

We turn to numerical simulations to characterize optimal policy in this setting and to see whether stochasticity undermines or reinforces the case for age dependence.

4.2. *Numerical results*

The simulations require a set of wage nodes for each age and transition matrices between nodes. To maximize comparability with the previous sections' results, I use the same wages and age groups as in Sections 2 and 3. That is, the average wage paths by education level used in the deterministic models above provide the nodes of the stochastic wage distributions in this model.[48] To calculate transition matrices, I start by assigning each individual to the wage node closest to their reported wage in each age range. Then, I calculate the empirical transition probabilities for individuals assigned to each node. Wage levels and transition matrices are shown in Table 7. While the most common movement from one age to the next is between the same node, about half of middle wage earners and 30% of low- and high-wage earners move across nodes in each transition.

The simulation results for the Case 3 planner's problems reinforce the lessons of the baseline and Case 2 simulations. I provide results on optimal intratemporal distortions and the welfare implications of reform. As in Case 2, optimal average tax rates are indeterminate in this setting for the Static Mirrlees and Partial Reform policies.[49]

47. A technical note: the planner's problems incorporate all incentive constraints explicitly and operate through backwards induction starting at age T. The constraints guarantee that individuals prefer their chosen allocation to any other allocation regardless of their reports when younger.

48. For reference, these wages correspond to the 38th, 56th, and 78th percentile wages within the first age group, the 35th, 55th, and 84th percentiles within the second age group, and the 37th, 58th, and 88th percentiles within the third age group.

49. In an earlier version of this paper, Weinzierl (2008), I simulate a model with stochastic wages but no private saving or borrowing. Average tax rates in that case are lower on younger workers, as in Case 1.

TABLE 7
Data for simulation of Case 3

Wage levels ($1999)		Wage node in each age range		
		Low	Middle	High
Age range	25–32	11·56	14·76	19·86
	33–40	12·63	17·35	27·17
	41–55	13·51	19·10	32·80
	Initial pbb	0·52	0·21	0·27
Transition matrices		Wage node in 33–40 age range		
		Low	Middle	High
Wage node in 25–32 age range	Low	0·71	0·21	0·08
	Middle	0·36	0·45	0·19
	High	0·10	0·23	0·67
		Wage node in 41–55 age range		
		Low	Middle	High
Wage node in 33–40 age range	Low	0·72	0·21	0·07
	Middle	0·29	0·50	0·21
	High	0·08	0·22	0·70

Source: PSID core sample household heads 1968–2001.

First, intratemporal distortions in the three policies are listed in Table 8.

As in the previous models, high-skilled young workers are inefficiently discouraged from working by higher-intratemporal distortions under an age-independent tax system. As in Case 2, both history-independent policies distort the highest-skilled workers at all ages to discourage cheating along the intertemporal margin. Also as before, the use of marginal distortions decreases with the sophistication of the policy. In the first age range, the population-weighted average distortion falls from 0·12 under the Static Mirrlees policy to 0·10 under the Partial Reform and 0·06 in the Full Optimum. A similar pattern holds for the middle-age range, though the oldest sees the Full Optimum assess slightly higher distortions, on average, than the Partial Reform.

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 9 shows social welfare (in utils) and the welfare gain (in consumption equivalents) of reform from the Static Mirrlees to the Partial Reform and Full Optimum policies. The welfare gain from age dependence is equivalent to approximately 1·0% of aggregate output in this model. This captures 63% of the gain from reform from the Static Mirrlees to the Full Optimum. Table 9 also shows the lifetime utility (in utils) for individuals with three sample wage paths under the three policies. As before, more sophisticated policies benefit especially those with low ability, while the high skilled obtain lower utility when the planner's tools are stronger.

As in previous analyses, we can decompose this welfare gain along efficiency and equity dimensions, as shown in Figure 4. In this model economy, stochasticity blurs the distinction between providing more consumption and smoother consumption to an individual. An individual that knows it will command fewer resources if it suffers a negative skill shock will save more out of current income, and thereby smooth consumption less, than an individual facing less risk. Similarly, individuals will be less likely to delay work until prime working age if they face greater consumption risk in the event of a future negative skill shock. Therefore, in contrast to the previous welfare decompositions, this figure combines the welfare gains from adjusting the timing of consumption and income with the gains from their distribution across types.

TABLE 8
Intratemporal distortions in Case 3 model simulation

Static Mirrlees

Distortion in first age range

	Wage node in 25–32 age range: Low	Middle	High	Average
	0·04	0·22	0·20	0·12

Distortion in second age range

Wage node in in 25–32 age range	Wage node in 33–40 age range: Low	Middle	High	Average
Low	0·12	0·23	0·14	
Middle	0·11	0·22	0·13	0·15
High	0·11	0·22	0·13	

Distortion in third age range

Wage node in 41–55 age range	Wage node in 25–32 age range	Wage node in 33–40 age range: Low	Middle	High	Average
Low	Low	0·23	0·22	0·17	
	Middle	0·22	0·20	0·15	
	High	0·22	0·21	0·16	
Middle	Low	0·27	0·26	0·22	
	Middle	0·26	0·25	0·21	0·18
	High	0·26	0·25	0·21	
High	Low	0·06	0·05	0·02	
	Middle	0·05	0·04	0·02	
	High	0·05	0·04	0·02	

Partial Reform

Distortion in first age range

	Wage node in 25–32 age range: Low	Middle	High	Average
	0·07	0·18	0·10	0·10

Distortion in second age range

Wage node in 25–32 age range	Wage node in 33–40 age range: Low	Middle	High	Average
Low	0·12	0·24	0·07	
Middle	0·10	0·22	0·06	0·13
High	0·09	0·22	0·05	

Distortion in third age range

Wage node in 41–55 age range	Wage node in 25–32 age range Low	Wage node in 33–40 age range: Middle	High	Average	
Low	Low	0·15	0·13	0·03	
	Middle	0·13	0·11	0·02	
	High	0·12	0·10	0·00	
Middle	Low	0·32	0·31	0·24	
	Middle	0·31	0·29	0·23	0·15
	High	0·30	0·29	0·22	
High	Low	0·09	0·08	0·02	
	Middle	0·08	0·07	0·01	
	High	0·07	0·06	0·01	

Full Optimum

Distortion in first age range

	Wage node in 25–32 age range: Low	Middle	High	Average
	0·06	0·15	0·00	0·06

Distortion in second age range

Wage node in 25–32 age range	Wage node in 33–40 age range: Low	Middle	High	Average
Low	0·09	0·15	0·00	
Middle	0·22	0·26	0·00	0·11
High	0·22	0·24	0·00	

Distortion in third age range

Wage node in 41–55 age range	Wage node in 25–32 age range	Wage node in 33–40 age range: Low	Middle	High	Average
Low	Low	0·13	0·27	0·28	
	Middle	0·21	0·33	0·28	
	High	0·21	0·31	0·30	
Middle	Low	0·20	0·28	0·32	
	Middle	0·29	0·38	0·32	0·16
	High	0·29	0·36	0·32	
High	Low	0·00	0·00	0·00	
	Middle	0·00	0·00	0·00	
	High	0·00	0·00	0·00	

TABLE 9

Welfare in Case 3 model simulation

	Total social welfare (Utils)	Welfare gain over SM (in consumption equivalents)†	Utility for sample wage paths (utils)		
			Always low	Always middle	Always high
Static Mirrlees	4·895	n/a	4·57	4·94	5·43
Partial Reform	4·914	0·98%	4·64	4·95	5·35
Full Optimum	4·926	1·56%	4·68	4·97	5·30

†Supplement consumption in the initial policy by this amount to yield the welfare of the target policy.

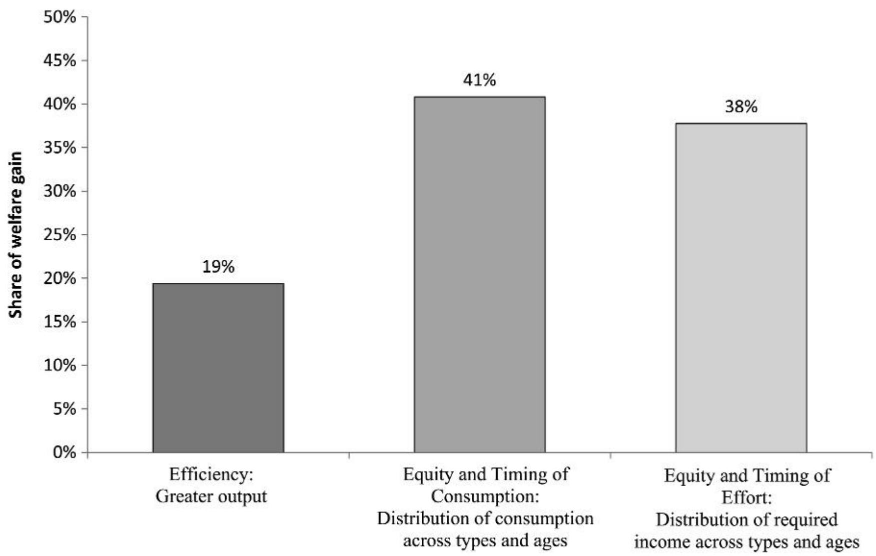

FIGURE 4

Components of welfare gain from Partial Reform in Case 3

Specifically, the second channel in Figure 4 is computed by scaling the consumption allocation for each type at each age under the Static Mirrlees policy to equal the same share of total consumption as under the Partial Reform policy. This adjustment includes two channels separately calculated in the baseline and Case 2 models: the allocation of consumption across types and the smoothing of consumption across ages. The third channel follows the same approach for required income. In both cases, the adjusted allocations yield greater welfare than in the Static Mirrlees, so that a smaller windfall would be required to raise welfare to the level achieved by the Partial Reform policy. The results of this decomposition closely resembles that of Case 1, where efficiency explains less than one-fifth of the gain and the allocations of consumption and required income each explain approximately 40%.

5. DISCUSSION

In this section, I discuss additional topics of interest that were not addressed directly above.

5.1. *Comparison to existing tax policy*

The analyses above have taken the Static Mirrlees policy as the starting point for tax reform. In reality, tax policies may differ substantially from the Static Mirrlees optimal policy. We may be interested in knowing how large the gains from reform would be relative to an existing tax policy and what share of the gain from full reform can be captured by age dependence.

To estimate these gains from reform, I apply the statutory income tax schedule of the U.S. in the mid-2000's augmented with a lump-sum grant to all individuals that balances the government's budget. Though this simplified version lacks the true tax code's numerous phase-outs, exemptions, and credits, it is a transparent representation of the existing system that captures its key components of progressive marginal tax rates and redistribution through transfers. Using the specifications of the deterministic and stochastic wage processes from Cases 1, 2, and 3 above, I have each individual choose an optimal allocation of labour effort and consumption given the tax code. The resulting allocations are used to calculate social welfare and the potential value of reform.

In Case 3, the welfare gains starting at this existing policy are 2·77% of consumption for Partial Reform and 3·45% for full reform. Thus, Partial Reform captures 80% of the gain from full reform.

5.2. *Endogeneity of wage paths*

I assume throughout this paper that wage paths are exogenous to individuals. Some recent work has begun to address this issue, such as Grochulski and Piskorski (2006) and Kapicka (2006), but the assumption of exogeneity is standard in the optimal tax literature. This does not mean it is 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.[50]

The results of this paper therefore require that a substantial portion of the variation of wages with age used in the simulations is inelastic to taxes. At least two considerations suggest that this may be the case. First, this paper's focus on age-dependent taxes between the ages of 25 and 55 limits concerns about distorting individuals' career choices. 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.

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, the timing of income, and human capital investment through education and work experience.

5.3. *Elasticity of labour supply by age*

One of the most direct reasons for the differentiation of taxation by age would be variation in the elasticity of labour supply with age. Unfortunately, empirical evidence is sparse. Kremer

50. 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 labour effort.

(2002) argues that "the limited available evidence suggests that younger workers have more elastic labour supply than prime-age workers," citing Clark and Summers (1981), who show more variation in employment rates with the business cycle for young workers.[51] 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 labour supply is uniform across age. If labour supply elasticity varies in the directions suggested by this limited evidence, the recommendations of this paper are strengthened.

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 (2007*b*). 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: *i.e.* a reform that can raise social welfare without harming any individuals.

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 reform to compensate those who would otherwise lose in the transition.[52]

To test whether age dependence is a Pareto-improving partial reform, I simulate the Case 3 models with the additional requirement that no individual of a given initial ability level can be worse off under the age-dependent policy than under the Static Mirrlees policy. As in the Utilitarian model, marginal distortions overall and especially on high-income young workers are lower under the Pareto-improving age-dependent tax policy than under the Static Mirrlees. More surprising, the welfare gain from Partial Reform is equivalent to 0·83% of aggregate consumption, nearly as large as in when using the Utilitarian welfare criterion. This gain equals 85% of the welfare gain from the unrestricted Partial Reform, and it captures 53% of the gain from reform to the Full Optimum policy.[53] The Pareto improvement requirement 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. These results suggests that age dependence is a reform capable of attracting broad-based support.

6. CONCLUSION

This paper uses theoretical analysis and numerical simulations based on micro-data to contrast three tax policy scenarios: a Static Mirrlees policy restricted to age-independent taxes, a Partial Reform policy in which labour income taxes can be age dependent, and a Full Optimum policy in which only private information constrains the design of taxes.

51. This evidence suggests that the extensive margin may drive variation in the elasticity with age. In Weinzierl (2008), I include an extensive margin by adding an additional, non-working type and fixed costs of working. The results of the main simulations are unaffected, though average tax rates rise overall to fund consumption for the non-working type.

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

53. A similar analysis shows that, in the baseline case the Pareto-improving Partial Reform generates a welfare gain equivalent to 1·2 % of aggregate consumption, 83% of the gain from unrestricted Partial Reform and 75% of the gain from Full Reform.

This analysis yields two specific policy recommendations that are largely robust across settings. First, marginal income taxes are lower for high-earning young workers in an optimal age-dependent policy. These individuals are near the top of their age-specific wage distribution, so the efficiency costs of distorting their labour effort are substantial, while the benefit from such a distortion (increasing tax revenue from higher earners of the same age) is small. This specific example illustrates a more general finding that age dependence reduces marginal distortions to effort, 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 allow me to quantify the welfare gain from age dependence and understand its components. Age dependence yields a large welfare gain equal to between 0·6% and 1·5% of aggregate annual consumption, and it captures more than 60% of the gain from reform to the optimal dynamic policy. Consistent through all three model economies considered in this paper, each of three main channels of welfare improvement has played a substantial role. Increased efficiency explains between 10% and 20% of the gain, while more equitable and timely allocations of consumption and required income explain approximately 40% each.

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.

Acknowledgment. Harvard Business School and NBER; mweinzierl@hbs.edu. I am grateful to Stefania Albanesi, Robert Barro, Michael Boskin, V. V. Chari, David Cutler, Emmanuel Farhi, Martin Feldstein, Caroline Hoxby, Oleg Itskhoki, Larry Jones, Emir Kamenica, Larry Katz, Patrick Kehoe, Narayana Kocherlakota, Erzo F.P. Luttmer, Greg Mankiw, Ellen McGrattan, Thomas Mertens, Brent Neiman, Chris Phelan, Jim Poterba, Robert Shiller, Kjetil Storesletten (the editor), Larry Summers, Aleh Tsyvinski, Ivan Werning, and two anonymous referees for their comments and suggestions.

REFERENCES

AKERLOF, G. (1978), "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning", *American Economic Review*, **68**, 8–9.

ALBANESI, S. and SLEET, C. (2006), "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, **73**, 1–30.

ALESINA, A., ICHINO, A. and KARABARBOUNIS, L. (2010), "Gender-based Taxation and the Division of Family Chores" (Working Paper), forthcoming American Economic Journal: Economic Policy.

ALTIG, D., AUERBACH, A., KOTLIKOFF, L., SMETTERS, K. and WALLISER, J. (2001), "Simulating Fundamental Tax Reform in the United States", *American Economic Review*, **91**, 574–595.

ATKINSON, A. and STIGLITZ, J. E. (1976), "The Design of Tax Structure: Direct Versus Indirect Taxation", *Journal of Public Economics*, **6**, 55–75.

BANKS, J. and DIAMOND, P. (2010), "The Base for Direct Taxation", in Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. and Poterba, J. (eds) *Dimensions of Tax Design: The Mirrlees Review*, Chapter 6, Oxford University Press.

BASTANI, S., BLOMQUIST, S., MICHELETTO, L. and LINDVALL, L. (2010), "The Welfare Gains of Age Related Optimal Income Taxation" (Working Paper 2010:12, Department of Economics, Uppsala University).

BISIN, A. and RAMPINI, A. (2005), "Markets as Beneficial Constraints on the Government", *Journal of Public Economics*.

BLOMQUIST, S. and MICHELETTO, L. (2003), "Age Related Optimal Income Taxation", *Scandinavian Journal of Economics*, **110**, 45–71.

CARD, D. (1991), "Intertemporal Labor Supply: An Assessment" (NBER Working Paper, 3602).

CLARK, K. B. and SUMMERS, L. H. (1981), "Demographic Differences in Cyclical Employment Variation", *Journal of Human Resources*, **16**, 61–77.

CONESA, J. C., KITAO, S. and KRUEGER, D. (2009), "Taxing Capital? Not a Bad Idea After All!", *American Economic Review*, **99**, 25–48.

DIAMOND, P. (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, **88**, 83–95.

DIAMOND, J. W. and TUNG, J. (2006), "Lifecycle Wealth Holdings by Lifetime Income" (Working Paper, Rice University).

FARHI, E. and WERNING, I. (2009), "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation" (Working Paper, MIT).

FARHI, E. and WERNING, I. (2010), "Insurance and Taxation over the Life Cycle" (Working Paper).

FELDSTEIN, M. (1976), "On the Theory of Tax Reform", *Journal of Public Economics*, **6**, 77–104.

FELDSTEIN, M. and SAMWICK, A. (1992), "Social Security Rules and Marginal Tax Rates", *National Tax Journal*, **45**, 1–22.

FRENCH, E. (2005), "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour", *Review of Economic Studies*, **72**, 395–427.

FULLERTON, D. and ROGERS, D. L. (1993), *Who Bears the Lifetime Tax Burden?* (Washington, DC: Brookings Institution).

GOLOSOV, M., KOCHERLAKOTA, N. and TSYVINSKI, A. (2003), "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, **70**, 569–587.

GOLOSOV, M., TROSHKIN, M. and TSYVINSKI, A. (2010), "Optimal Taxation: Merging Micro and Macro Approaches", forthcoming Journal of Money, Credit, and Banking.

GOLOSOV, M. and TSYVINSKI, A. (2006), "Designing Optimal Disability Insurance: A Case for Asset Testing", *Journal of Political Economy*, **114**, 257–279.

GOLOSOV, M., TSYVINSKI, A., and WERNING, I. (2007), "New Dynamic Public Finance: A User's Guide", *NBER Macroannual* 2006, 317–387.

GROCHULSKI, B. and PISKORSKI, T. (2006), "Risky Human Capital and Deferred Capital Income Taxation" (Working Paper No. 06–13, FRB of Richmond).

GUESNERIE, R. (1977), "On the Direction of Tax Reform", *Journal of Public Economics*, **7**, 179–202.

GUVENEN, F. (2007), "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?", *American Economic Review*, **97**, 687–712.

HUGGETT, M. and PARRA, J. C. (2010), "How Well Does the U.S. Social Insurance System Provide Social Insurance?", *Journal of Political Economy*, **118**, 76–112.

JUDD, K. and SU, C.-L. (2006), "Optimal Income Taxation with Multidimensional Taxpayer Types" (Working Paper). Earliest available draft April, 2006; more recent draft November 22, 2006.

KAPICKA, M. (2006), "Optimal Income Taxation with Human Capital Accumulation and Limited Recordkeeping", *Review of Economic Dynamics*, **9**, 612–639.

KEANE, M. P. and WOLPIN, K. I. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473–522.

KIMBALL, M.S. and SHAPIRO, M.D. (2008), "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" (NBER Working Paper. 14208).

KOCHERLAKOTA, N. (2005), "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", *Econometrica*, **73**, 1587–1621.

KREMER, M. (2002), "Should Taxes be Independent of Age" (Working Paper, Harvard University).

LIEBMAN, J. B. (2003), "Should Taxes Be Based on Lifetime Income? Vickrey Taxation Revisited" (Working Paper, Harvard Kennedy School).

MANKIW, N. G. and WEINZIERL, M. (2010), "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution", *American Economic Journal: Economic Policy*, **2**, 155–176.

MEADE, J. (1978), *The Structure and Reform of Direct Taxation* (London: Allen and Unwin).

MIRRLEES, J. A. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, **38**, 175–208.

ROGERSON, W. (1985), "Repeated Moral Hazard", *Econometrica*, **53**, 69–76.

SAEZ, E. (2001), "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies*, **68**, 205–229.

SAEZ, E. (2002), "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses", *Quarterly Journal of Economics*, **117**, 1039–1073.

STIGLITZ, J. E. (1987), "Pareto Efficient and Optimal Taxation and the New New Welfare Economics", in Auerbach, A. and Feldstein, M. (eds) *Handbook on Public Economics* (North Holland: Elsevier Science Publishers) 991–1042.

STORESLETTEN, K., TELMER, C. and YARON, A. (2001), "How Important Are Idiosyncratic Shocks? Evidence from Labor Supply", *American Economic Review*, **91**, 413–417.

VICKREY, W. (1939), "Averaging of Income for Income Tax Purposes", *Journal of Political Economy*, **47**, 379–397.

WEINZIERL, M. (2008), "The Surprising Power of Age-Dependent Taxation" (Working Paper, Harvard University).

WERNING, I. (2007*a*), "Optimal Fiscal Policy with Redistribution," *Quarterly Journal of Economics*, 925–967.

WERNING, I. (2007*b*), "Pareto Efficient Income Taxation," (Working Paper, MIT).

YARED, P. (2010), "Politicians, Debt, and Taxes", *Review of Economic Studies*, 77, 806–840.