Technical Appendix: The Surprising Power of Age-Dependent Taxes

August 11, 2010

Contents

This Appendix contains a variety of material, grouped into six sections:

1 Implementation of the optimal allocations

In this section I show how the allocations in the three baseline policy scenarios discussed in the main paper can be implemented with nonlinear labor income taxes. Then, I show how the Partial Reform (age-dependent) optimal allocations in Case 2 can be implemented with nonlinear labor income taxes. Similar reasoning applies to Cases 3 and 4.

1.1 Baseline case

The tax authority sets labor income taxes $T(\cdot)$ to maximize social welfare. I consider three tax policies: Static Mirrlees, Partial Reform, and Full Optimum. In the Static Mirrlees policy, the labor income tax is a direct function of income only; in the Partial Reform it is a direct function of income and the taxpayer's current age; while in the Full Optimum it is a direct function of the lifetime path of incomes and the taxpayer's current age. Formally, write $T(y)$ for the Static Mirrlees policy, $T(y,t)$ for the Partial Reform policy, and $T\left(\{y(\cdot)\}_{t=1}^{T}, t\right)$ for the Full Optimum policy. Note that variation in the wage with age means that the tax $T(y)$ depends indirectly on age through income, even though the Static Mirrlees tax policy cannot directly depend on age. This is why the allocations c_t^i and y_t^i depend on age in the Static Mirrlees policy in the main paper.

Individuals maximize utility subject to these tax policies. The maximization problem of an individual of type i is

Problem 1 *(Individual Maximization)*

$$\max_{\{c,y\}_{t=1}^{T}} \sum_{t=1}^{T} \beta^{t-1} \left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right)\right)$$

subject to the individual's budget constraint for each age $t \in \{1, 2, ..., T\}$:

$$c_t^i = y_t^i - T(\cdot),$$

where the arguments of the tax function depend on the policy scenario.

What we want to show is that the solutions to the Static Mirrlees, Partial Reform, and Full Optimum planner's problems from the paper's baseline model will be endogenously chosen by individuals solving the above maximization subject to the appropriate tax function $T(\cdot)$ for each policy.

The following proposition states this formally for the Static Mirrlees policy:

Proposition 1 *For the solution* $\{c_t^{*i}, y_t^{*i}\}_{i=1,t=1}^{I,T}$ *to the Static Mirrlees planner's problem from the baseline model, there exists a function of labor income* $T^*(y)$ *such that individuals solving the Individual Maximization problem subject to* $T^*(y)$ *choose* $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$*, where* $c_t^i = c_t^{*i}$ *and* $y_t^i = y_t^{*i}$ *for all* i *and all* t*.*

This proof resembles the approach in Kocherlakota (2005). Define

$$T^*(y) = \begin{cases} y_t^{*i} - c_t^{*i}, \text{ if } y = y_t^{*i} \text{ for some } i, t \\ 2y, \qquad\qquad \text{if } y \neq y_t^{*i} \text{ for any } i, t \end{cases}$$

for each labor income y. Now, consider the individual of type i and age t who earns wage w_t^i and solves the Individual Maximization problem above. For any y not necessarily equal to y_t^{*i}, it will consume $c_t^i = y - T^*(y)$, with $T^*(y)$ defined as above. Any y not equal to some y_t^{*i} yields negative consumption and, therefore, negative infinite utility, so it is not chosen. Note that these punishing taxes on off-equilibrium incomes would be unnecessary if the income distribution were continuous–the discreteness of the wage distribution is required for computational purposes. If $y = y_t^{*i}$ for some i, t, then the individual consumes c_t^{*i} because $u'(c)$ is always positive and the individual cannot save or borrow across periods. All that remains is to ensure that the individual with wage w_t^i prefers to earn y_t^{*i} and consume c_t^{*i} than to pay some other tax $T^*(y_s^{*j})$, earn y_s^{*j} and consume c_s^{*j}. This is ensured by the incentive constraints in the Static Mirrlees problem, so $T^*(y)$ implements the optimal allocation. QED

Now we turn to the Partial Reform policy.

Proposition 2 *For the solution* $\{c_t^{*i}, y_t^{*i}\}_{i=1,t=1}^{I,T}$ *to the Partial Reform planner's problem from the baseline model, there exists a function of labor income and age* $T^*(y,t)$ *such that individuals solving the Individual Maximization problem subject to* $T^*(y,t)$ *choose* $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$*, where* $c_t^i = c_t^{*i}$ *and* $y_t^i = y_t^{*i}$ *for all* i *and all* t*.*

For this policy, define

$$T^*(y,t) = \begin{cases} y_t^{*i} - c_t^{*i}, & \text{if } y = y_t^{*i} \text{ for some } i, \text{ given } t \\ 2y, & \text{if } y \neq y_t^{*i} \text{ for any } i, \text{ given } t \end{cases}$$

for each labor income y and age t. Note that the function now conditions on age, so that two individuals earning a given income may pay different taxes if they are of different ages. Also, an individual earning an income not in the set of possible incomes for its age pays a high tax. Now, consider the individual of type i and age t who earns wage w_t^i and solves the Individual Maximization problem above. Any y not equal to some y_t^{*i}, given its age t, yields negative consumption and, therefore, negative infinite utility, so it is not chosen. If $y = y_t^{*i}$ for some i given t, then the individual consumes c_t^{*i} because $u'(c)$ is always positive and the individual cannot save or borrow across periods. The incentive constraints in the (PR) problem ensure that the individual with wage w_t^i at age t prefers to earn y_t^{*i} and consume c_t^{*i} than to pay some other tax $T^*\left(y_t^{*j}\right)$, earn y_t^{*j} and consume c_t^{*j}. QED

Finally we turn to the Full Optimum policy. Let $y^t = (y_1, y_2, ...y_t)$ denote a labor income history.

Proposition 3 *For the solution* $\{c_t^{*i}, y_t^{*i}\}_{i=1,t=1}^{I,T}$ *to the Full Optimum planner's problem from the baseline model, there exists a function of the individual history of labor incomes and the current age* $T^*(y^t,t)$ *such that individuals solving the Individual Maximization problem subject to* $T^*(y^t,t)$ *choose* $\{c_t^i, y_t^i\}_{i=1,t=1}^{I,T}$*, where* $c_t^i = c_t^{*i}$ *and* $y_t^i = y_t^{*i}$ *for all* i *and all* t*.*

For this policy, define

$$T^*\left(y^t, t\right) = \begin{cases} y_t^{*i} - c_t^{*i}, \text{ if there is an } i : y_s \in y^t : y_s = y_s^{*i} \text{ for all } s \in [1, 2, ...t] \\ 2y, \qquad\qquad\qquad \text{if there is no such } i \end{cases}$$

for each labor income history y^t and age t. Consider the individual of type i and age t who earns wage w_t^i and solves the Individual Maximization problem above. When $t = 1$, any income not equal to y_1^{*i} for some i yields negative consumption and, therefore, negative infinite utility, so it is not chosen. Suppose the individual chooses the income corresponding to some i'. Then, for $t > 1$, any income not equal to $y_t^{*i'}$ for that i' also yields negative consumption, so the individual chooses the age-specific income path corresponding to a single i' throughout its life. The only remaining step is to ensure that the individual prefers the path assigned to its true type than any other type, which is ensured by the incentive constraints in the Full Optimum planner's problem. QED

1.2 Case 2 Partial Reform

Now I show how the Case 2 Partial Reform allocations can be implemented with nonlinear labor income taxation. In principle, the private saving and borrowing that differentiates Case 2 from the baseline could complicate the translation from optimal allocations to taxes. It turns out, however, that the implementation is quite similar to that shown above.

The key result of this section is that the intratemporal distortions derived in the main paper are, in fact, the marginal taxes faced by individuals choosing among bundles of pre-tax and after-tax incomes (y and x), not pre-tax income and consumption (y and c) as in the baseline. To show this, we consider the following tax system:

$$T^*\left(y, t\right) = \begin{cases} y_t^{*i} - x_t^{*i}, \quad \text{if } y = y_t^{*i} \text{ for some } i \text{, given } t \\ T \cdot y, \qquad\qquad \text{if } y \neq y_t^{*i} \text{ for any } i \text{, given } t \end{cases}$$

By the same arguments as in the baseline case, individuals choosing subject to $T^*(y,t)$ will choose the optimal allocations as derived in the main paper. The nonlinearity of the tax system, which severely punishes choices other than those within the set of optimal allocations (the size of the punishment can be made arbitrarily large to ensure that only optimal allocations are chosen). The proper choices within that set are guaranteed by the incentive compatibility constraints from the direct mechanism problem specified in the main paper.

Unlike in the baseline case, it is not immediate that the intratemporal distortions which are implied by the optimal $\{x_t^{*i}, y_t^{*i}\}$ allocations are equal to marginal taxes on income that would implement these allocations. To find the marginal taxes that implement these allocations, we rewrite this tax system so that choosing each bundle $\{x_t^{*i}, y_t^{*i}\}$ is equivalent to choosing a lump-sum grant Ω_t^{*i} and a constant marginal tax on income y_t^{*i}, namely τ_t^{*i}. Formally, $y_t^{*i} - x_t^{*i} = y_t^{*i}\tau_t^{*i} - \Omega_t^{*i}$. We then show that, in fact, the marginal tax rates τ_t^{*i} that implement the optimal allocations are the same as the intratemporal distortions derived in the main paper.

To see that this is true, suppose an individual of type i when age t considers earning one extra unit of y, starting at $\{x_t^{*i}, y_t^{*i}\}$. Given the "lump-sum/linear" tax system we specified, this raises its after-tax income by $dx = (1 - \tau_t^{*i})$. Because the individual can save and borrow, it spreads dx across its lifetime. The resulting change in consumption for period $s \in \{1, 2, ...T\}$ is $R^{s-1}\frac{dx}{T}$, which changes utility by $(\beta R)^{s-1}\frac{dx}{T}u'(c_t^{*i})$, where c_t^{*i} is the optimal (constant) consumption level. Letting $\beta R = 1$ as in the main paper, the total utility change across all T periods is $dxu'(c_t^{*i})$. For the individual to choose $\{x_t^{*i}, y_t^{*i}\}$, individual maximization must therefore set $dxu'(c_t^{*i}) - v'\left(\frac{y_t^{*i}}{w_t^i}\right)\frac{1}{w_t^i} = 0$, i.e. the change in utility from earning an extra unit of y must be zero at $\{x_t^{*i}, y_t^{*i}\}$. This simplifies to $dx = \frac{v'\left(\frac{y_t^{*i}}{w_t^i}\right)}{w_t^i u'\left(c_t^{*i}\right)}$, or using the expression for dx, we obtain $\tau_t^{*i} = 1 - \frac{v'\left(\frac{y_t^{*i}}{w_t^i}\right)}{w_t^i u'\left(c_t^{*i}\right)}$. Thus, the marginal tax rate that implements the optimal allocation is the same as the intratemporal distortion derived in the main paper. (We define $\Omega_t^{*i} = x_t^{*i} - (1-\tau_t^{*i})\, y_t^{*i}$ to ensure that the rate τ_t^{*i} applied to y_t^{*i} yields x_t^{*i} in after-tax income).

2 Multiple generations

In this section, I discuss how the analysis of age dependence would change if there were multiple generations.

In the most direct extension of the analysis to multiple generations, little changes from the baseline model. Index generations with $g \in \{1, 2, ..., G\}$. Assume wages may depend on the generation and denote $\{w(i, g, t)\}_{t=1,g=0}^{T,G}$ as the wage for individual i born in year g and currently of age t. Denote the corresponding allocations of consumption and income $\{c(i, g, t)\}_{t=1,g=0}^{T,G}$ and $\{y(i, g, t)\}_{t=1,g=0}^{T,G}$. The planner's problem is:

$$\max_{c,y} \sum_{g=0}^{T} \sum_{i=1}^{I} \pi^i V(i, g)$$

I retain the generation-specific feasibility constraints, so for each g :

$$\sum_{i=1}^{I} \sum_{t=1}^{T} R^{T-t} \left[y(i, g+t, t) - c(i, g+t, t) \right] = 0$$

Generation-specific feasibility constraints capture a plausible moral restriction on policy. Relaxing them, and allowing for net transfers between generations, would introduce an additional purpose for tax policy that all three policy scenarios, if generation-dependent, could pursue. Depending on the importance of these transfers to social welfare, the Static Mirrlees may perform better relative to the Partial Reform and Full Optimum policies than in the analysis without intergenerational transfers (though the welfare ordering of the policies would be unchanged).

The incentive constraints reflect our assumption that the FO and PR planners can make taxes generation-dependent as well as age-dependent. In the Partial Reform, the incentive constraints are, for each g, t, i, i'

$$\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\left(u\left(c\left(i',g+t,t\right)\right)-v\left(\frac{y\left(i',g+t,t\right)}{w\left(i,t\right)}\right)\right),$$

which means that each individual prefers its allocation to any other individual's within its generation at each age.

For the Full Optimum, the incentive constraints are, for each g,i, i'

$$\sum_{t=1}^{T}\beta^{t-1}\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\sum_{t=1}^{T}\beta^{t-1}\left(u\left(c\left(i',g+t,t\right)\right)-v\left(\frac{y\left(i',g+t,t\right)}{w\left(i,t\right)}\right)\right),$$

which means that each individuals prefers its lifetime allocation to any other individual's within its generation.

For the Static Mirrlees policy problem, there is a question of whether the planner ought to be allowed to condition on generation or not. If it cannot, generational conditioning widens the gap between the Static Mirrlees and Partial Reform policies relative to a single generation case. If it can, the incentive constraints are, for each g, t, t', i, i'

$$\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\left(u\left(c\left(i',g+t',t'\right)\right)-v\left(\frac{y\left(i',g+t',t'\right)}{w\left(i,t\right)}\right)\right),$$

so that individuals must prefer their own allocation to any other individual's of any age within its generation.

As is clear from these formal problems, generation dependence combined with generation-specific feasibility constraints imply that these policy problems are separable into generation-specific problems that are identical (up to changes in wage paths across generations) to the single generation problem considered in the main paper. Thus, the results of the main analysis carry through unchanged, robust to the inclusion of multiple generations.

2.1 Calendar-year dependence

One interesting issue that arises once multiple generations are introduced is whether tax policy can be calendar-year dependent. In the Static Mirrlees policy, the combination of generation dependence and calendar year dependence yields age dependence, so to remain age-independent, the Static Mirrlees policy cannot be both generation-dependent and year-dependent. As shown above, the generation-dependent model closely resembles the baseline model in the paper. How would assuming calendar-year dependence instead of generation dependence affect the results of our Static Mirrlees analysis?

2.1.1 No real wage growth

I begin with the case of no real wage growth. Then, we can show that calendar-year dependence is no different from generation dependence for the Static Mirrlees policy. To see this, note that the wage distributions will be the same for each generation, so the currently old have the same wage distribution as the currently young will have when old. Formally, the Static Mirrlees planner's problem with calendar-year dependence would be identical to its problem with generation dependence except for the incentive constraints. In the year-dependent problem, the incentive constraints would be, for each g, t, t', i, i' :

$$\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\left(u\left(c\left(i',g+t,t'\right)\right)-v\left(\frac{y\left(i',g+t,t'\right)}{w\left(i,t\right)}\right)\right).$$

I want to show that the solution to the Static Mirrlees problem with generation dependence is the same as that for the problem with calendar-year dependence when the age-dependent wage distributions are the same for all generations.

The outline of the proof is as follows. Start with the generation-dependent problem. The problem can be decomposed into separate optimizations for each generation: the objective and feasibility readily separate by generation and the incentive constraints are generation-specific by definition. However, each of these component problems is identical, so they have the same solutions. If not, I could choose the welfare maximizing solution and apply it to all generations, raising welfare. Thus, allocations are constant across generations in the generation-dependent problem with no real wage growth. These allocations also satisfy the incentive constraints of the year-dependent problem. To see this, take the ICs from the generation-dependent problem:

$$\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\left(u\left(c\left(i',g+t',t'\right)\right)-v\left(\frac{y\left(i',g+t',t'\right)}{w\left(i,t\right)}\right)\right),$$

and use the result that allocations are constant across generations, so that $c\left(i',g+t',t'\right)=c\left(i',g+t,t'\right)$ and $y\left(i',g+t',t'\right)=y\left(i',g+t,t'\right)$ to substitute for the right-hand side and obtain:

$$\left(u\left(c\left(i,g+t,t\right)\right)-v\left(\frac{y\left(i,g+t,t\right)}{w\left(i,t\right)}\right)\right)\geq\left(u\left(c\left(i',g+t,t'\right)\right)-v\left(\frac{y\left(i',g+t,t'\right)}{w\left(i,t\right)}\right)\right),$$

which is the IC constraint for the year-dependent problem. Thus, the solution to the generation-dependent problem solves the year-dependent problem. Next, I show the reverse.

Start with the year-dependent Static Mirrlees problem, and recall that I assumed the age-dependent wage distributions are the same for all generations. Now, I will show that constant allocations across calendar years are optimal for this problem. Suppose not. Then, incentive-compatible allocations $\left\{c\left(i,g+s,t\right),y\left(i,g+s,t\right)\right\}_{i=1,t=1}^{I,T}$ and $\left\{c\left(i,g+s',t\right),y\left(i,g+s',t\right)\right\}_{i=1,t=1}^{I,T}$ differ for some s,s'. Assume, without loss of generality, that $s'>s$. Then, consider the generations that are of age T in year s and s', so that the first of these generations does not live to receive the allocations in year s'. These generations receive different lifetime paths of allocations. Because each generation enters symmetrically into the planner's objective function and feasibility constraint, this cannot be optimal (i.e., if replace one generation's allocations with the other's, feasibility must hold because it is by generation, but social welfare must either increase or decrease, which is not optimal). Thus, the solution to the year-dependent problem is constant allocations across calendar years, or $c\left(i',g+t',t'\right)=c\left(i',g+t,t'\right)$ and

$y(i', g + t', t') = y(i', g + t, t')$ for all t, t'. This is equivalent to constant allocations across generations, so we have shown the equivalence of the problems. In words, when the age-dependent wage distributions are the same for all generations, calendar-year dependence is equivalent to generation-dependence for the Static Mirrlees policy.

2.1.2 Real wage growth

The analysis is more complicated if real wage growth is positive. To take an extreme example, suppose that the wage distribution is the same across ages within a calendar year, so that real wage growth is the sole source of wage growth during a lifetime. In this situation, age- and generation-dependent policy allows for higher average taxes at older ages, transferring consumption to earlier ages and smoothing it over the lifetime (much as in the baseline analysis within a single generation), while taking advantage of higher earnings power at later ages. In contrast, the Static Mirrlees policy with calendar-year dependence cannot make such transfers, so it inefficiently allocates consumption and income across ages for each generation. This problem is worse than in the generation-dependent case because the wage distributions of the young and old in a calendar year now entirely overlap.

At the same time, this overlap of the wage distributions within a calendar year has a benefit to the Static Mirrlees planner, because it may be able to avoid some of the distortions it otherwise is forced to use. For example, if real wage growth is the sole source of wage growth, then the maximum wage for the young will also be the maximum wage for the mature in each calendar year, and neither would need to be distorted. This is a benefit of calendar-year dependence for the Static Mirrlees planner.

Generally, calendar-year dependence means that the Static Mirrlees policy acts on the cross-sectional distribution of wages rather than on the distribution of wages for a generation over time. This can generate both benefits (in terms of lower distortions) and costs (in terms of worse intertemporal allocations) for the Static Mirrlees planner. Illustrative simulations suggest that the combination of these effects can either increase or decrease the power of the

Static Mirrlees policy relative to when it is generation-dependent.

3 Hamiltonian approach to baseline model

This section works through the Partial Reform policy in the baseline model using the classic approach to static Mirrlees models. To more easily map the approach to the paper's analysis, I use discrete notation.

I simplify the analysis by assuming quasilinear utility, as in Diamond (1998):

$$u(c,l) = c - v(l)$$

The social planner specifies a tax function $T_t(wl)$, such that $c = wl - T_t(wl)$, to maximize social welfare:

$$W = \sum_i \pi^i \alpha^i \left(\sum_t \beta^{t-1} U(w_t^i) \right)$$

where all notation is as in the main paper, and period utility for individual i of age t is:

$$U(w_t^i) = u\left(w_t^i l_t^i - T_t\left(w_t^i l_t^i\right)\right) - v\left(l_t^i\right).$$

Note that the tax function $T_t(\cdot)$ can be age-specific (t-specific) in the Partial Reform policy. Using quasilinearity, we rewrite period utility as:

$$U(w_t^i) = w_t^i l_t^i - T_t\left(w_t^i l_t^i\right) - v\left(l_t^i\right).$$

The social planner's maximization is subject to two sets of constraints, feasibility and incentive constraints.

The feasibility constraint is that taxes are purely redistributive and that the (single) generation of individuals must fund its consumption with income. Given quasilinear utility, and $c = wl - T_t(wl)$, we can write the feasibility constraint as

$$\sum_t \sum_i \pi^i R^{T-t} \left[w_t^i l_t^i \left(w_t^i\right) - U\left(w_t^i\right) - v\left(l_t^i\left(w_t^i\right)\right)\right] = 0.$$

The incentive constraints are that each individual chooses labor supply to maximize utility, given the tax function. The individual solves:

$$\max_{\{l\}_t} \left\{ \sum_t \beta^{t-1} \left(u\left(w_t^i l_t^i - T_t\left(w_t^i l_t^i\right)\right) - v\left(l_t^i\right)\right) \right\}$$

which yields T first-order conditions of the form:

$$FOC_{l_t^i} : u'\left(w_t^i l_t^i - T_t\left(w_t^i l_t^i\right)\right) w_t^i \left(1 - T_t'\left(w_t^i l_t^i\right)\right) - v'\left(l_t^i\right) = 0.$$

Given quasilinear utility, these simplify to:

$$FOC_{l_t^i} : w_t^i \left(1 - T_t'\left(w_t^i l_t^i\right)\right) - v'\left(l_t^i\right) = 0$$

for each i and each t. As is conventional in what is called the "first order approach," I will replace the incentive constraints with these first order conditions when solving the planner's problem. This approach relies on the Spence-Mirrlees condition and a second order condition that income rises with wage, both of which I assume hold (the second does hold in my numerical simulations).

For the analysis below, we use these first-order conditions to simplify differential constraints on the state variables of the planner's problem. Those constraints are merely the envelope conditions of the planner's objective with respect to w_t^i, which are:

$$\frac{\partial U(w_t^i)}{\partial w_t^i} = u'\left(w_t^i l_t^i - T_t\left(w_t^i l_t^i\right)\right) l_t^i \left(1 - T_t'\left(w_t^i l_t^i\right)\right).$$

Given quasilinear utility, these simplify to

$$\frac{\partial U(w_t^i)}{\partial w_t^i} = l_t^i \left(1 - T_t'\left(w_t^i l_t^i\right)\right).$$

Substituting in the first-order conditions from above, we obtain:

$$\frac{\partial U(w_t^i)}{\partial w_t^i} = l_t^i \frac{v'\left(l_t^i\right)}{w_t^i}.$$

Following Mirrlees' (1971) original approach, I represent the Partial Reform planner's optimization problem in the baseline model with a Hamiltonian that is a function of utilities and labor effort only. Though this is an approach designed for a static model, it can be applied to this dynamic model because of time-separability, history-independence, no private saving, and Pareto-weights. In particular, these characteristics cause the planner's problem to separate into T age-specific problems, so I can write a set of Hamiltonians, one for each age t, and then combine them. For each t, I write the Hamiltonian:

$$H_t = \alpha^i \frac{\pi^i}{dw_t^i} \beta^{t-1} U\left(w_t^i\right) + \lambda \frac{\pi^i}{dw_t^i} R^{T-t}\left[w_t^i l_t^i\left(w_t^i\right) - U\left(w_t^i\right) - v\left(l_t^i\left(w_t^i\right)\right)\right] + \mu_t \frac{\partial U(w_t^i)}{\partial w_t^i},$$

or

$$H_t = \alpha^i \frac{\pi^i}{dw_t^i} \beta^{t-1} U\left(w_t^i\right) + \lambda \frac{\pi^i}{dw_t^i} R^{T-t}\left[w_t^i l_t^i\left(w_t^i\right) - U\left(w_t^i\right) - v\left(l_t^i\left(w_t^i\right)\right)\right] + \mu_t l_t^i \frac{v'\left(l_t^i\right)}{w_t^i}.$$

where dw^i is the distance between the wage level i and the next-highest wage. The multipliers on the feasibility constraint and the differential constraint are denoted λ and μ. To go from the first expression to the second, I have substituted in for the differential constraints with the first-order conditions as mentioned above. Each of these Hamiltonians is independent of any other t, so we can combine them to obtain:

$$H=\sum_t \alpha^i \frac{\pi^i}{dw_t^i}\beta^{t-1}U\left(w_t^i\right)+\lambda\sum_t \frac{\pi^i}{dw_t^i}R^{T-t}\left[w_t^i l_t^i\left(w_t^i\right)-U\left(w_t^i\right)-v\left(l_t^i\left(w_t^i\right)\right)\right]+\sum_t \mu_t l_t^i \frac{v'\left(l_t^i\right)}{w_t^i}$$

The key result is derived from the planner's first-order conditions with respect to labor effort, which are:

$$FOC_{l_t^i}:\frac{\partial H}{\partial l_t^i}=\left[\lambda\frac{\pi^i}{dw_1^i}\frac{dw_1^i}{dw_t^i}R^{T-t}\left(w_t^i-v'\left(l_t^i\left(w_t^i\right)\right)\right)+\mu_t\left(\frac{v'\left(l_t^i\right)+l_t^i v''\left(l_t^i\right)}{w_t^i}\right)\right]=0$$

Now, we want to simplify these to be in terms of the tax function only. This involves using first-order conditions and conditions on the multipliers. As for the multipliers, note that Pontryagin's Maximum Principle implies $\mu_t'\left(w_t^i\right)=-\frac{\partial H}{\partial U_t^i}$. Using the Hamiltonian, we can say:

$$\mu_t'\left(w_t^i\right)=\frac{\pi^i}{dw_t^i}\beta^{t-1}\left(\lambda R^{T-1}-\alpha^i\right),$$

The transversality conditions are

$$\mu(0)=\lim_{w_t\to\infty}\mu\left(w_t\right)=0$$

Now, integrate each of these multipliers' FOCs from w to ∞ and use the tranversality at the maximum wage to get:

$$\left[\mu_t(\infty)-\mu_t\left(w_t^i\right)\right]=-\mu_t\left(w_t^i\right)=\sum_{w_t^j=w_t^i}^{w_t=\infty}\frac{\pi^j}{dw_t^j}dw_t^j\beta^{t-1}\left(\lambda R^{T-1}-\alpha^j\right)=\sum_{w_t^j=w_t^i}^{w_t=\infty}\pi^j\beta^{t-1}\left(\lambda R^{T-1}-\alpha^j\right),$$

which yields

$$\mu_t\left(w_t^i\right)=\sum_{w_t^j=w_t^i}^{w_t^j=\infty}\pi^j\beta^{t-1}\left(\alpha^j-\lambda R^{T-1}\right).$$

Next, use these expressions and both transversality conditions to get

$$[\mu_t(\infty) - \mu_t(0)] = 0 = \sum_{w_t^j=0}^{w_t^j=\infty} \pi^j \beta^{t-1} \left(\alpha^j - \lambda R^{T-1}\right),$$

so

$$\lambda = \frac{\sum_{w_t^j=0}^{w_t^j=\infty} \pi^j \beta^{t-1} \alpha^j}{\sum_{w_t^j=0}^{w_t^j=\infty} \pi^j \beta^{t-1} R^{T-1}} = \frac{1}{R^{T-1}} \sum_{w_t^j=0}^{w_t^j=\infty} \pi^j \alpha^j,$$

which is constant across ages.

Now, define

$$\Pi_t\left(w_t^i\right) = \sum_{w_t^j=0}^{w_t^i} \pi^j,$$

as the proportion of t-aged individuals who have wages below w_t^i. Also, define

$$D_t\left(w_t^i\right) = \frac{1}{1-\Pi_t\left(w_t^i\right)} \sum_{w_t^j=w_t^i}^{w_t^j=\infty} \pi^j \alpha^j.$$

as the average social marginal welfare weight of t-aged individuals who have wages above w_t^i (recall that utility is quasilinear, so $u'(c) = 1$).

Thus,

$$\lambda = \frac{1}{R^{T-1}} \left(1 - \Pi_t(0)\right) D_t(0) = \frac{1}{R^{T-1}} D_t(0) \text{ for all } t.$$

Also, we can simplify the expression for $\mu_t\left(w_t^i\right)$ to obtain:

$$\mu_t\left(w_t^i\right) = \beta^{t-1}\left(1 - \Pi_t\left(w_t^i\right)\right)\left(D_t\left(w_t^i\right) - D_t(0)\right).$$

Then, define the labor supply elasticity for person i of age t as:

$$\varepsilon_t^i = \frac{w_t^i\left[1 - T_t'\left(w_t^i l_t^i\right)\right]}{l_t^i v''\left(l_t^i\right)} = \frac{v'\left(l_t^i\right)}{l_t^i v''\left(l_t^i\right)}.$$

Finally, recall the individual FOCs:

$$FOC_{l_t^i} : w_t^i \left(1 - T_t' \left(w_t^i l_t^i\right)\right) - v' \left(l_t^i\right) = 0$$

Now simplify the planner's FOCs using the conditions on λ and μ_t, the expression for the labor supply elasticity, and the individual FOCs. After some simplification, we obtain

$$FOC_{l_t^i} : \frac{\partial H}{\partial l_t^i} = \left(D_t(0) \frac{\pi^i}{dw_t^i} w_t^i T_t' \left(w_t^i l_t^i\right) + \left(1 - \Pi_t \left(w_t^i\right)\right) \left(D_t \left(w_t^i\right) - D_t(0)\right) \left(1 + \frac{1}{\varepsilon_t^i}\right) \left(1 - T_t' \left(w_t^i l_t^i\right)\right)\right) = 0$$

which implies

$$\frac{T_t' \left(w_t^i l_t^i\right)}{1 - T_t' \left(w_t^i l_t^i\right)} = \left(\frac{1 - \Pi_t \left(w_t^i\right)}{\pi^i}\right) \frac{dw_t^i}{w_t^i} \left(1 - \frac{D_t \left(w_t^i\right)}{D_t(0)}\right) \left(1 + \frac{1}{\varepsilon_t^i}\right)$$

This condition characterizes marginal labor income taxes in the Partial Reform model. It is identical to the condition derived in Kremer (2002), though the dynamic setting of this derivation provides more detail on the determinants of the distributional terms $D_t \left(w_t^i\right)$ than Kremer's static setting could. As derived in Diamond (1998), an analogous condition holds for the Static Mirrlees policy, but without the age subscripts. This seemingly small difference drives the disparities in intratemporal distortions identified in the main text.

3.1 Results on high-income marginal distortions; relation to Saez (2001)

Here, I use the same data as in the baseline numerical simulation to estimate the key components of the results from the Hamiltonian method above, which allows for the use of a much finer wage distribution.

Assuming constant elasticities across age, the key terms are $\left(\frac{1-\Pi_t\left(w_t^i\right)}{w_t^i \pi^i}\right)$ and $\left(1 - \frac{D_t\left(w_t^i\right)}{D_t(0)}\right)$. I call these the "hazard term" and "distribution term," respectively.

In Figures A1, A2, and A3, I plot these terms for my baseline calibration against income for each age group. I take a moving average for the hazard term over a five dollar wage range to smooth out noise.

Figure A1: "Hazard Term"
Moving average over five wages

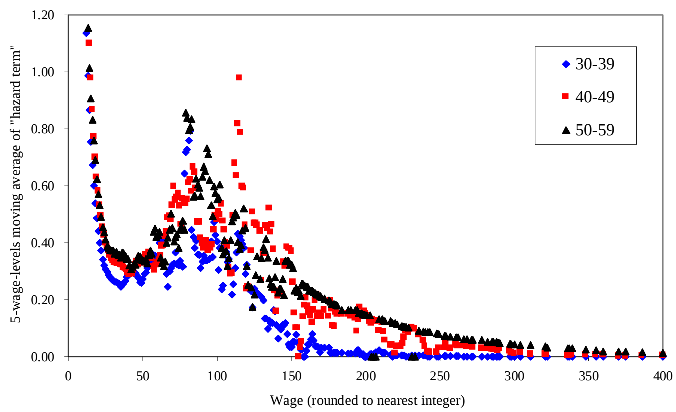


Figure A2: "Distribution Term"

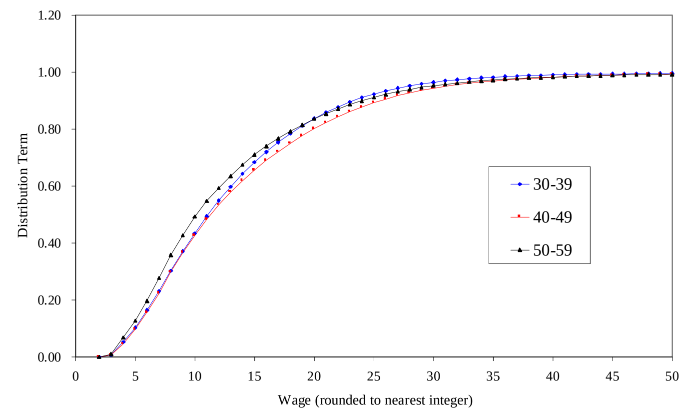
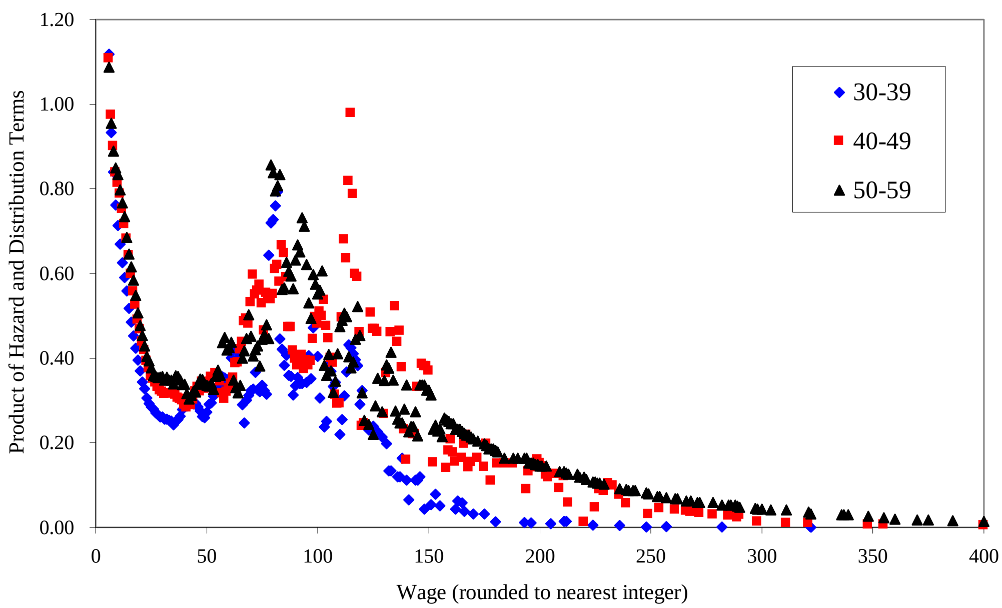


Figure A3: Product of "Hazard" and "Distribution" Terms



In Figures A1 and A2, we can see that the two work in opposite directions–the hazard term is lower for the young and the distribution term is higher.

Figure A3 shows that the hazard term is the more powerful factor and that the product of the two terms, and thus the recommended marginal tax, is substantially lower for the young over a relatively wide range of high incomes. This is the source of the footnote in the main paper.

Saez (2002) stressed that skills appear to follow a Pareto distribution. For the purposes of this paper, the shape of the distribution is less important than that the product of the key terms is lower for the young than the old at high incomes, meaning lower distortions on the high-income young.

4 Proofs of Propositions

4.1 Intratemporal Benchmark: formerly Proposition 1 from Weinzierl (2008)

We want to show that the Full Optimum (FO) planner chooses constant allocations. If that holds, then each other planner will do so as well, if feasible, because they can do no better than the FO

Consider the first-order conditions of the Full Optimum planner. The intratemporal distortion derived from these FOCs is

$$1-\frac{v'\left(l_t^i\right)}{w_t^i u'\left(c_t^i\right)}=\frac{\sum_j\left(1-\left(\frac{w_t^i}{w_t^j}\right)^\sigma\right)\mu^{i|j}}{\alpha^i\pi^i+\sum_j\mu^{j|i}-\sum_j\left(\frac{w_t^i}{w_t^j}\right)^\sigma\mu^{i|j}}.$$

If the wage distribution is fixed across periods, all variables on the RHS are constant over time, so the distortion is identical across periods. Now consider the intertemporal first-order condition for consumption:

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

which is the classic Atkinson-Stiglitz result for a deterministic economy. Thus, the FO planner chooses a constant consumption allocation for each individual i. With the intratemporal distortion, this implies a constant income allocation, as well. Thus, the Full

Optimum planner assigns a constant $\{c, y\}$ allocation for each i over t.

Both the Partial Reform (PR) and Static Mirrlees (SM) planner can do no better than the FO planner, as their constraint sets are subsets of its constraint set. Thus, if the FO allocation is feasible and incentive compatible, they will choose it. It is feasible for both because all three scenarios use the same feasibility condition. It is incentive compatible for the PR planner because the FO planner's ICs with constant allocations can be rewritten as:

$$\left(1+\beta+...\beta^{T-1}\right)\left(u\left(c^{i}\right)-v\left(\frac{y^{i}}{w^{i}}\right)\right) \geq \left(1+\beta+...\beta^{T-1}\right)\left(u\left(c^{j}\right)-v\left(\frac{y^{j}}{w^{i}}\right)\right).$$

or

$$\left(u\left(c^{i}\right)-v\left(\frac{y^{i}}{w^{i}}\right)\right) \geq \left(u\left(c^{j}\right)-v\left(\frac{y^{j}}{w^{i}}\right)\right).$$

which holds, therefore, for each t and thus for the PR's set of IC constraints. For the SM, cross-age deviations are replicas of intra-age deviations, given constant allocations and wages, so its IC set is identical to the PR's, and the constant allocations are incentive compatible for it as well. Thus, all three scenarios choose constant allocations.

4.2 Proposition 1 (Top Marginal Distortion)

We begin with the first implication. Consider the expression for the Partial Reform (PR) distortion:

$$1-\frac{v'\left(l_t^i\right)}{w_t^i u'\left(c_t^i\right)} = \frac{\sum_j \left(1-\left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\right)\mu_t^{i|j}}{\alpha^i\pi^i+\sum_j \mu_t^{j|i}-\sum_j \left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\mu_t^{i|j}}.$$

By the supposition, $\frac{w_t^i}{w_t^j}>1$ for all j. We show that if $\mu_t^{i|j}>0$ for any j, the distortion must be negative. If, instead, $\mu_t^{i|j}=0$ for all j, the distortion is zero (the classic "no distortion on the top" case). The IC constraint for $\mu_t^{i|j}$ is

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

We can construct examples in which this IC binds, namely if redistributive preferences of the planner raise c_t^i high enough to offset higher y_t^i. Thus, it is possible that $\mu_t^{i|j} > 0$ for some j. If that is the case, then the numerator in the PR intratemporal distortion expression is negative. To prove the implication, the denominator must be positive. Any $\mu_t^{j|i} > 0$ will make this more likely, so suppose that $\mu_t^{j|i} = 0$ for all j to be conservative. Thus, if $\alpha^i \pi^i > \sum_j \left(\frac{w_t^i}{w_t^j}\right)^\sigma \mu_t^{i|j}$, we have proved the result. The first-order condition for the PR planner with respect to l_t^i is:

$$v'\left(l_t^i\right) \beta^{t-1} \left(\alpha^i \pi^i - \sum_j \left(\frac{w_t^i}{w_t^j}\right)^\sigma \mu_t^{i|j}\right) = R^{T-t} \lambda \pi^i$$

The multiplier on the feasibility constraint must be positive, i.e., $\lambda > 0$,since marginal utility is always positive. We assume that disutility of labor is such that $v'\left(l_t^i\right) > 0$. Thus, the first-order condition implies $\alpha^i \pi^i > \sum_j \left(\frac{w_t^i}{w_t^j}\right)^\sigma \mu_t^{i|j}$, and we have proved the result. This proves that top distortions within each age are nonpositive for the PR scenario.

A similar proof holds for the FO scenario, where the intratemporal distortion is

$$1 - \frac{v'\left(l_t^i\right)}{w_t^i u'\left(c_t^i\right)} = \frac{\sum_j \left(1 - \left(\frac{w_t^i}{w_t^j}\right)^\sigma\right) \mu^{i|j}}{\alpha^i \pi^i + \sum_j \mu^{j|i} - \sum_j \left(\frac{w_t^i}{w_t^j}\right)^\sigma \mu^{i|j}}$$

and the relevant IC is

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

If $\mu^{i|j} > 0$ for some j and $\frac{w_t^i}{w_t^j} > 1$ for all j at t (by assumption) then the numerator of the FO distortion expression is negative. Again, the relevant FOC is

$$\frac{1}{w_t^i} v'\left(\frac{y_t^i}{w_t^i}\right) \beta^{t-1} \left(\alpha^i \pi^i + \sum_j \mu^{j|i} - \sum_j \left(\frac{w_t^i}{w_t^j}\right)^\sigma \mu^{i|j}\right) = R^{T-t} \pi^i \lambda$$

so that the denominator must be positive and the distortion negative when $\mu^{i|j} > 0$ for some

j. If $\mu^{i|j} = 0$ for all j, the distortion is zero. This completes the proof.

We now consider the second implication.

We want to show that no IC constraint corresponding to the multipliers $\mu_t^{i|j}$ will bind. I use variational arguments assuming non-increasing Pareto weights. I show the case for $I = 2$, but the same arguments can be used when $I > 2$. Suppose, contrary to the proposition, that $\mu_t^{i|j} > 0$. The reverse IC (multiplied by $\mu_t^{j|i}$) does not bind because $\frac{w_t^i}{w_t^j} > 1$. The incentive constraints imply that if $\mu_t^{i|j} > 0$, then either (1) $c_t^i > c_t^j$ or (2) $y_t^i < y_t^j$ or (3) both $c_t^i > c_t^j$ and $y_t^i < y_t^j$. First, if $c_t^i > c_t^j$ and $y_t^i \geq y_t^j$, transfer ε units of consumption from i to j at age t. These transfers are feasible and raise social welfare due to diminishing marginal utility of consumption and non-increasing Pareto weights. If these transfers proceed until $c_t^i = c_t^j$, then $\mu_t^{i|j} = 0$ because $y_t^i \geq y_t^j$ and $\frac{w_t^i}{w_t^j} > 1$. If, however, these transfers cause $\mu_t^{j|i} > 0$ with $c_t^i > c_t^j$, then $\mu_t^{i|j} = 0$ because i and j cannot both be indifferent between the two allocations when $\frac{w_t^i}{w_t^j} > 1$. Second, if $c_t^i \leq c_t^j$ and $y_t^i < y_t^j$, increase y_t^i and decrease y_t^j by δ. This is feasible and raises social welfare due to increasing marginal disutility of labor (note $l_t^i < l_t^j$) and non-increasing Pareto weights. If this adjustment proceeds until $\frac{1}{w_t^i}\frac{y_t^i}{w_t^i} = l_t^i = l_t^j = \frac{1}{w_t^j}\frac{y_t^j}{w_t^j}$, which implies $y_t^i > y_t^j$, then $\mu_t^{i|j} = 0$ because $c_t^i \leq c_t^j$. If this adjustment causes $\mu_t^{j|i} > 0$ with $l_t^i < l_t^j$, then $\mu_t^{i|j} = 0$ because i and j cannot both be indifferent between the two allocations when $\frac{w_t^i}{w_t^j} > 1$ Finally, if $c_t^i > c_t^j$ and $y_t^i < y_t^j$, then use both the transfers of consumption and adjustments of income. If both proceed until $c_t^i = c_t^j$ and $l_t^i = l_t^j$, then $\mu_t^{i|j} = 0$ because $y_t^i > y_t^j$. If not, then $\mu_t^{j|i} > 0$, implying $\mu_t^{i|j} = 0$. This shows that the incentive constraint preventing j from claiming i's allocations cannot bind at the optimal policy. If $I > 2$, different j's may be relevant at different ages. As long as $\frac{w_t^i}{w_t^j} > 1$ for all j and all t, as assumed in the proposition, the same procedures as above can be applied.

4.3 Intertemporal Benchmark: formerly Proposition 3 in Weinzierl (2008)

This proposition is a direct consequence of the proof of Proposition 1, as it implies that all individuals face constant allocations over time.

4.4 Proposition 2 (Symmetric Inverse Euler)

The proof manipulates the first order conditions from the planner's problem in the baseline economy, Partial Reform scenario. These FOCs are, for consumption in periods t and $t+1$ for individual i:

$$u'\left(c_t^i\right)\beta^{t-1}\left(\pi^i\alpha^i+\sum_j \mu_t^{j|i}-\sum_j \mu_t^{i|j}\right)=\lambda R^{T-t}\pi^i,$$

$$u'\left(c_{t+1}^i\right)\beta^{t}\left(\pi^i\alpha^i+\sum_j \mu_{t+1}^{j|i}-\sum_j \mu_{t+1}^{i|j}\right)=\lambda R^{T-t-1}\pi^i,$$

implying

$$\beta^{t-1}\left(\pi^i\alpha^i+\sum_j \mu_t^{j|i}-\sum_j \mu_t^{i|j}\right)=\frac{\lambda R^{T-t}\pi^i}{u'\left(c_t^i\right)},$$

$$\beta^{t}\left(\pi^i\alpha^i+\sum_j \mu_{t+1}^{j|i}-\sum_j \mu_{t+1}^{i|j}\right)=\frac{\lambda R^{T-t-t}\pi^i}{u'\left(c_{t+1}^i\right)},$$

Now, sum across i on each side:

$$\beta^{t-1}\sum_{i=1}^{I}\left[\pi^i\alpha^i+\sum_j \mu_t^{j|i}-\sum_j \mu_t^{i|j}\right]=\sum_{i=1}^{I}\left[\frac{\lambda R^{T-t}\pi^i}{u'\left(c_t^i\right)}\right],$$

$$\beta^{t}\sum_{i=1}^{I}\left[\pi^i\alpha^i+\sum_j \mu_{t+1}^{j|i}-\sum_j \mu_{t+1}^{i|j}\right]=\sum_{i=1}^{I}\left[\frac{\lambda R^{T-t-1}\pi^i}{u'\left(c_{t+1}^i\right)}\right],$$

which simplify (due to constraint multipliers that cancel and $\beta R=1$) to:

$$\beta^{T-1}\sum_{i=1}^{I}\pi^i\alpha^i=\sum_{i=1}^{I}\frac{\lambda\pi^i}{u'\left(c_t^i\right)},$$

$$\beta^{T-1}\sum_{i=1}^{I}\pi^i\alpha^i=\sum_{i=1}^{I}\frac{\lambda\pi^i}{u'\left(c_{t+1}^i\right)},$$

This implies the symmetric IEE

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

The SIEE can also be proven through a variational argument. In the Partial Reform policy, baseline scenario, lower utility for each individual at age t by δ through providing lower consumption c. This preserves incentive compatibility.within age t and raises "revenue" of $\frac{\delta}{u'\left(c_t^i\right)}$ from each individual, for a total of $\sum_{i=1}^{I}\frac{\pi^i\delta}{u'\left(c_t^i\right)}$. Then, raise utility for each individual at age $t+1$ by the same δ through c. This leaves total utility same for each individual, since $\beta R=1$. It also preserves incentive compatibility.within age $t+1$. The "cost" is $\frac{\delta}{u'\left(c_t^i\right)}$ for each individual, for a total of $\sum_{i=1}^{I}\frac{\pi^i\delta}{u'\left(c_{t+1}^i\right)}$. If "revenue" exceeds "cost", total utility can be increased with these operations. Since that's not optimal, it must be that the optimal allocation satisfies: $\sum_{i=1}^{I}\frac{\pi^i}{u'\left(c_t^i\right)}=\sum_{i=1}^{I}\frac{\pi^i}{u'\left(c_{t+1}^i\right)}$, the symmetric inverse Euler equation.

4.5 Proposition 3 (Baseline and Case 3 Equivalence)

The proof of this proposition is to show that the objective function, feasibility constraint, and incentive compatibility constraints are the same in each model for the Static Mirrlees and Partial Reform policies. Intuitively, the same paths of wages exist in each model ex post, so I assign each ex post wage path in the Case 3 model to a deterministic path in the baseline model. Because neither the planner nor the individuals can link one age to another in these two models, the problems are therefore identical.

I show the proof for the Static Mirrlees; the same approach applies to the Partial Reform.

The Case 3 feasibility condition is:

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

Replace j with k to avoid confusion

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

Then, use $\pi_t^k = \sum_{i(t):w_t^{i_t}=w_t^k} \pi^{i(t)}$ to rewrite this as:

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

or, noting the summation over k,

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

Now, use the proposition's condition that there is a unique j such that $w_t^{i_t} = w_t^j$ for all t and $\pi^{i(t)} = \pi^j$.

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

which is the same as the baseline feasibility constraint, up to a change from j to i :

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

For the incentive constraints, the Case 3 model's constraints are of the form:

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

The proposition's conditions allow us to replace each i_t in the Case 3 constraint with j, yielding

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

exactly the baseline condition, up to a change from j to i and k to j :

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

Finally, for the objective function, the Case 3 objective is

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

Use $\pi_t^j=\sum_{i(t):w_t^{i_t}=w_t^j}\pi^{i(t)}$ and $\alpha_t^j=\frac{\sum_{i(t):w_t^{i_t}=w_t^j}\pi^{i(t)}\alpha\left(W_T^{i(t)}\right)}{\sum_{i(t):w_t^{i_t}=w_t^j}\pi^{i(t)}}$ to rewrite the Case 3 objective as

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

Now, to avoid confusion, change j to k,

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

Noting the summation over k, this is also

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

Then apply the proposition's condition that there is a unique j in the baseline model such that $w_t^{i_t}=w_t^j$ for all t and $\pi^{i(t)}=\pi^j$, and note that $\alpha\left(W_T^j\right)=\alpha^j$ in the baseline (deterministic) model, so that this condition can be written

$$\max_{\{c,y\}}\sum_t\beta^{t-1}\sum_j\pi^j\alpha^j\left(u\left(c_t^j\right)-v\left(\frac{y_t^j}{w_t^j}\right)\right).$$

which is the same as the baseline objective function, up to a reordering of the summations and a change from j to i:

$$\max_{\{c,y\}} \sum_{i=1}^{I} \pi^i \alpha^i \sum_{t=1}^{T} \beta^{t-1} \left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right)\right).$$

Thus, the Static Mirrlees planner solves the same problems in the baseline and Case 3 models. The same procedure can be used to prove the result for the Partial Reform planner.

5 Specific expressions omitted from text

5.1 Baseline case: Expressions for intratemporal distortions

To simplify the results, I assume disutility takes the isoelastic form

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

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

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

$$\tau^{SM}(i,t) = \frac{\sum_{s=1}^{T}\sum_{j=1}^{I}\left(1-\left(\frac{w_t^i}{w_s^j}\right)^{\sigma}\right)\beta^{s-t}\mu_{t|s}^{i|j}}{\alpha^i\pi^i + \sum_{s=1}^{T}\sum_{j=1}^{I}\mu_{s|t}^{j|i} - \sum_{s=1}^{T}\sum_{j=1}^{I}\left(\frac{w_t^i}{w_s^j}\right)^{\sigma}\beta^{s-t}\mu_{t|s}^{i|j}}. \tag{2}$$

where $\mu_{s|t}^{j|i}$ is the multiplier on the incentive constraint preventing individual i of age t from claiming the allocation of any other individual j of age s.[1]

In the Partial Reform scenario, it is:

$$\tau^{PR}(i,t) = \frac{\sum_{j=1}^{I}\left(1-\left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\right)\mu_t^{i|j}}{\alpha^i\pi^i + \sum_{j=1}^{I}\mu_t^{j|i} - \sum_{j=1}^{I}\left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\mu_t^{i|j}}. \tag{3}$$

[1] As Judd and Su (2006) point out, it is possible that optimal policies in these settings include pooling multiple types of individuals at the same allocation. In that case, the values of the incentive constraints' multipliers may not be uniquely defined. These expressions implicilty assume that some values for these has been chosen, but the discusion of the expressions does not presume anything about those values.

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

Finally, in the Full Optimum scenario, it is:

$$\tau^{FO}(i,t) = \frac{\sum_{j=1}^{I}\left(1-\left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\right)\mu^{i|j}}{\alpha^i\pi^i + \sum_{j=1}^{I}\mu^{j|i} - \sum_{j=1}^{I}\left(\frac{w_t^i}{w_t^j}\right)^{\sigma}\mu^{i|j}}. \tag{4}$$

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

By comparing these expressions we can learn a key lesson: 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. To see this contrast formally, note that expression (3) depends on wage ratios within age t, while expression (2) depends on wage ratios across ages t and s. The multipliers $\mu_{t|s}^{i|j}$, $\mu_t^{i|j}$, and $\mu^{i|j}$ introduce some complexity into these expressions, but the distinction remains.

5.2 Case 2: Planner's problems

I distinguish between after-tax income, denoted x_t^i for individual i of age t, and consumption, denoted c_t^i as before.

The planner assigns after-tax income to each pre-tax income to maximize social welfare from the baseline model. The feasibility constraint is similar to the baseline model's, though I replace consumption with after-tax income:

$$\sum_{i=1}^{I}\pi^i\sum_{t=1}^{T}R^{T-t}\left(y_t^i - x_t^i\right) = 0 \tag{5}$$

For the incentive constraints, we need notation that reflects the individual's ability to claim a wage different from its true wage at any age, be assigned another individual's allocations, and transfer resources across ages.[2] Let $W_T^{j(s_t)} = \left\{ w_{s_1}^{j_{s_1}}, w_{s_2}^{j_{s_2}}, ..., w_{s_T}^{j_{s_T}} \right\}$ denote a T-period path of wages corresponding to individuals of type j_{s_t} and age s_t, where s_t can vary across t. Thus, $W_T^i = \{w_1^i, w_2^i, ..., w_T^i\}$ denotes the true path of wages for individual i. Then, $\left\{ y\left(w_s^{j_{s_t}}\right) \right\}_{t=1}^T$ and $\left\{ x\left(w_s^{j_{s_t}}\right) \right\}_{t=1}^T$ are the sequence of pre-tax income and after-tax income allocations assigned to an individual who claims the wage sequence $W_T^{j(s_t)}$.

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

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

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

subject to the feasibility constraint (5) and the incentive constraints

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

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

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

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

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

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

other path $W_T^{j(s_t)}$.[3]

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

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

Problem 3 *(Case 2 Partial Reform: Age-dependent)*

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

subject to the feasibility constraint: (5) and the incentive constraints

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

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

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

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

As in the baseline model, the Partial Reform planner's incentive constraints are easier

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

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

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

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

Problem 4 *(Case 2 Full Optimum: Age-Dependent and History-Dependent)*

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

subject to the feasibility constraint and the incentive constraints from the baseline model.

5.3 Case 2: Intratemporal distortions example: Partial Reform policy

As with the results shown in the main paper, we can derive conditions for the marginal distortion on the low-skilled worker. For example, the intratemporal distortion in the first period on this worker is

$$1 - \frac{v'\left(l_1^L\right)}{w_1^L u'\left(c_1^{LL}\right)} = \frac{\left(1 - \left(\frac{w_1^L}{w_1^H}\right)^{\sigma}\right) \mu^{LL|HH} + \left(\frac{c_1^{LL}}{c_1^{LH}} - \left(\frac{w_1^L}{w_1^H}\right)^{\sigma}\right) \mu^{LH|HH}}{\pi^L \alpha^L - \left(\frac{w_1^L}{w_1^H}\right)^{\sigma} \mu^{LL|HH} - \left(\frac{w_1^L}{w_1^H}\right)^{\sigma} \mu^{LH|HH}}.$$

5.4 Case 3: More complicated Pareto weights

Notation for the stochastic wage case is as follows. Denote an individual's true path of wages as $W_T^{i(t)} = \left\{w_1^{i_1}, w_2^{i_2}, ...w_t^{i_t}, ...w_T^{i_T}\right\}$ and let $\pi^{i(t)}$ denote the population proportion represented by this individual. Using these probabilities, let $\pi_t^j = \sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)}$ denote the probability of wage level w_t^j at age t. It is the sum of the population proportions of the individuals whose wage paths equal w_t^j at age t. Note that $\sum_{j=1}^{I} \pi_t^j = 1$ for all t. Denote the transition matrix between ages t and $t+1$ as $P_{t,t+1}$, whose element (m,n) is $P_{t,t+1}(m,n) = \Pr(w_{t+1}^n | w_t^m)$. In words, $P_{t,t+1}(m,n)$ is the probability that an individual with wage w_t^m will have wage w_{t+1}^n. Thus, the population proportion of an individual with wage path $W_T^{i(t)}$ can also be written $\pi^{i(t)} = \prod_{t=1}^{T-1} \pi_1^i P_{t,t+1}(i_t, i_{t+1})$.

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

$$\alpha_t^j = \frac{\sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)} \alpha\left(W_T^{i(t)}\right)}{\sum_{i(t):w_t^{i_t}=w_t^j} \pi^{i(t)}}, \tag{6}$$

as the expected Pareto weight on an individual of age t with wage j. Expression (6) is the probability-weighted average of the Pareto weights on individuals with wage paths that include w_t^j when they are of age t. For instance, individuals with the first-period wage w_1^j will go on to have a variety of wage paths. The weight α_1^j captures the probability-weighted average of their eventual Pareto weights.

5.5 Case 3: Planners' problems and intratemporal distortions

When individuals can transfer resources across periods, the planners in the Static Mirrlees and Partial Reform scenarios do not control consumption directly. Thus, these planners specify pre-tax income and after-tax income bundles in Case 3, just as in Case 2. The objective function for these two policies is

$$\max_{\{x,y\}} \left\{ \sum_{i(t)} \pi^{i(t)} \alpha \left(W_T^{i(t)} \right) \sum_{t=1}^{T} \beta^{t-1} \left(u\left(c_t^{i_t}\right) - v\left(\frac{y_t^{i_t}}{w_t^{i_t}}\right)\right) \right\}, \tag{7}$$

and the feasibility constraint is:

$$\sum_{i(t)} \pi^{i(t)} \sum_{t=1}^{T} R^{T-t} \left(y_t^{i_t} - x_t^{i_t}\right) = 0. \tag{8}$$

The incentive constraints in Case 3 for the Static Mirrlees and Partial Reform scenarios are complex. In words, they reflect that an individual can choose a separate deviation strategy, including saving and borrowing, for each possible true path of wages. So, the Static Mirrlees incentive constraints must ensure that each individual prefers its allocation to any of the other allocation streams it might claim. If there are T periods and I wage levels, the number of these other streams is $\left[(IT)^{1+I(T-1)} - 1\right]$. That is, in the first period, an individual can claim any of IT wages, including his own. When planning for the second period, for each of the I possible second period wages, he can claim any of the IT wages again. The Partial Reform incentive constraints are, as usual, a subset of the Static Mirrlees scenario's because the planner can make age-dependent allocations. Each individual in the Partial Reform must be prevented from claiming allocation streams other than her own that number only $\left[I^{1+I(T-1)} - 1\right]$, as she can claim I, not IT, wages for each wage level she receives.

For the Static Mirrlees, let $W_T^{j(s_t)}\left(W_T^i\right) = \left\{ w_{s_1}^{j_{s_1}}\left(w_1^i\right), w_{s_2}^{j_{s_2}}\left(w_2^i\right), ..., w_{s_t}^{j_{s_t}}\left(w_t^i\right), ...w_{s_T}^{j_{s_T}}\left(w_T^i\right) \right\}$ be a (potentially false) claimed path of wages that depends on the true path W_T^i. Using this notation, the incentive constraints for the Static Mirrlees planner are that, for all

$i, j \in \{1, 2, ...I\}$ and all $W_T^{j(s_t)}(W_T^i)$, the expression

$$\left[u\left(c_1\left(W_T^i\right)\right) - v\left(\frac{y\left(w_1^{i_t}\right)}{w_1^{i_t}}\right) + \sum_{t=1}^{T} \beta^{t-1} \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right)\left(u\left(c_{t+1}\left(W_T^i\right)\right) - v\left(\frac{y\left(w_{t+1}^{i_{t+1}}\right)}{w_{t+1}^{i_{t+1}}}\right)\right)\right]$$

must be greater than or equal to

$$\left[u\left(c_1\left(W_T^{j(s_t)}\left(W_T^i\right)\right)\right) - v\left(\frac{y\left(w_{s_1}^{j_{s_1}}\left(w_1^i\right)\right)}{w_1^i}\right)\right. \tag{9}$$
$$\left. + \sum_{t=1}^{T} \beta^{t-1} \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right)\left(u\left(c_{t+1}\left(W_T^{j(s_t)}\left(W_T^i\right)\right)\right) - v\left(\frac{y\left(w_{s_{t+1}}^{j_{s_{t+1}}}\left(w_{t+1}^{i_{t+1}}\right)\right)}{w_{t+1}^{i_{t+1}}}\right)\right)\right]$$

where

$$\left\{c_t\left(W_T^{j(s_t)}\left(W_T^i\right)\right)\right\}_{t=1}^{T} = \arg\max_{\{c_t\}} \left\{ \begin{array}{c} \text{expression (9)} \\ s.t. \sum_t R^{T-t}\left(x\left(w_{s_t}^{j_{s_t}}\left(w_t^{i_t}\right)\right) - c_t\right) = 0 \end{array} \right\}$$

Note that the consumption path chosen by the individual is subject to its own feasibility constraint, as in Case 2, but that stochasticity means consumption may not be smooth. In fact, individuals will choose to satisfy a version of the intertemporal Euler equation that takes into account stochasticity:

$$u'\left(c_t\left(W_T^{j(s_t)}\left(W_T^i\right)\right)\right) = \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right)\left[u'\left(c_{t+1}\left(W_T^{j(s_t)}\left(W_T^i\right)\right)\right)\right].$$

This Euler equation implies that consumption may not be smooth because, for example, an individual who saves in order to insure against a low future skill shock may have extra resources available late in life if it receives high skill shocks instead.

For the PR scenario, let $W_T^{j(t)}(W_T^i) = \left\{w_1^{j_1}(w_1^i), w_2^{j_2}(w_2^i), ..., w_t^{j_t}(w_t^i), ...w_T^{j_T}(w_T^i)\right\}$ be a (potentially false) claimed path of wages that depends on the true path W_T^i. Note that, in contrast to the Static Mirrlees, each wage claim must be of the same age as the true wage, reflecting the planner's ability to make age-dependent allocations. Using this notation, the incentive constraints for the Static Mirrlees planner are that, for all $i, j \in \{1, 2, ...I\}$ and all $W_T^{j(t)}(W_T^i)$, the expression

$$\left[u\left(c_1\left(W_T^i\right)\right) - v\left(\frac{y\left(w_1^{i_t}\right)}{w_1^{i_t}}\right) + \sum_{t=1}^{T} \beta^{t-1} \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right) \left(u\left(c_{t+1}\left(W_T^i\right)\right) - v\left(\frac{y\left(w_{t+1}^{i_{t+1}}\right)}{w_{t+1}^{i_{t+1}}}\right)\right)\right]$$

must be greater than or equal to

$$\begin{aligned} &\left[u\left(c_1\left(W_T^{j(t)}\left(W_T^i\right)\right)\right) - v\left(\frac{y\left(w_1^{j_1}\left(w_1^i\right)\right)}{w_1^i}\right)\right. \\ &\left. + \sum_{t=1}^{T} \beta^{t-1} \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right) \left(u\left(c_{t+1}\left(W_T^{j(t)}\left(W_T^i\right)\right)\right) - v\left(\frac{y\left(w_{t+1}^{j_{t+1}}\left(w_{t+1}^{i_{t+1}}\right)\right)}{w_{t+1}^{i_{t+1}}}\right)\right)\right] \end{aligned} \tag{10}$$

where

$$\left\{c_t\left(W_T^{j(t)}\left(W_T^i\right)\right)\right\}_{t=1}^{T} = \arg\max_{\{c_t\}} \left\{ \begin{array}{c} \text{expression (10)} \\ s.t. \sum_t R^{T-t}\left(x\left(w_t^{j_t}\left(w_t^{i_t}\right)\right) - c_t\right) = 0 \end{array} \right\}$$

Again, individuals will choose to satisfy a version of the intertemporal Euler equation that takes into account stochasticity:

$$u'\left(c_t\left(W_T^{j(t)}\left(W_T^i\right)\right)\right) = \sum_{i_{t+1}} P_{t,t+1}\left(i_t, i_{t+1}\right)\left[u'\left(c_{t+1}\left(W_T^{j(t)}\left(W_T^i\right)\right)\right)\right].$$

6 Numerical illustrations omitted from the main text

6.1 Baseline economy: Proposition 2 (Top Marginal Distortion)

A simple numerical example may help build intuition for the differences between scenarios. Consider an economy with only two individuals and two time periods, so that $i = \{t, s\}$ and

$t = \{1, 2\}$. Consider the following wage paths in this two-by-two economy:

$$\begin{bmatrix} & (t) & (s) \\ (1) & 8 & 12 \\ (2) & 12 & 16 \end{bmatrix},$$

Assuming the main paper's standard parameterization, the intratemporal distortions in the policy scenarios are:

$$\begin{bmatrix} \text{Static Mirrlees} & & \\ & (t) & (s) \\ (1) & 0.27 & 0.10 \\ (2) & 0.10 & 0.00 \end{bmatrix}; \begin{bmatrix} \text{Partial Reform} & & \\ & (t) & (s) \\ (1) & 0.17 & 0.00 \\ (2) & 0.15 & 0.00 \end{bmatrix}; \begin{bmatrix} \text{Full Optimum} & & \\ & (t) & (s) \\ (1) & 0.18 & 0.00 \\ (2) & 0.15 & 0.00 \end{bmatrix}.$$

The Static Mirrlees distortions differ dramatically from the Partial Reform. The Static Mirrlees distorts the young high earner, unlike either of the more sophisticated models, because it must prevent the older, low earner from mimicking that type. This new distortion forces the planner to distort the young low earner more as well, ensuring that the young high type will not mimic it. Finally, the old low earner faces a larger distortion than in the more sophisticated models because that is the only way for the planner to direct resources to it without tempting the old high earner.

6.2 Case 2: Discussion and illustrative example of intratemporal distortions

The cost of private saving to the Partial Reform planner is that it magnifies the incentive problems of redistribution to low-income individuals. Recall that any redistribution of after-tax income to low-income individuals within an age group will tempt high-earners of the same age to falsely claim the lower-income allocation. This temptation is worse with private saving and borrowing, because these high-earners can now transfer resources across ages to maintain an elevated level of consumption when they falsely claim a lower-income allocation.

At the same time, private saving and borrowing has a potentially substantial benefit for the planner. Recall, from the baseline model, that the Partial Reform planner cannot offer smooth consumption paths to individuals because it cannot keep track of their identities across ages. Private saving and borrowing allows agents to accomplish that smoothing on their own.

To make the planner's choices more concrete, consider a two-by-two example economy, where the wage paths are:

$$\begin{bmatrix} w_t^i & (t) & (s) \\ (1) & 10 & 12 \\ (2) & 12 & 18 \end{bmatrix},$$

The intratemporal distortions chosen by the Partial Reform planner in the baseline model and in the Case 2 model with private saving and borrowing are:

$$\begin{bmatrix} & \text{Baseline} & \\ & (t) & (s) \\ (1) & 0.03 & 0.00 \\ (2) & 0.14 & 0.00 \end{bmatrix}, \text{ and } \begin{bmatrix} & \text{Case 2} & \\ & (t) & (s) \\ (1) & 0.02 & 0.04 \\ (2) & 0.09 & 0.003 \end{bmatrix}$$

The planner in Case 2, where individuals can privately save and borrow, distorts the high-earner in both periods. This combats the temptation for type s to oversave or overborrow.

6.3 Elasticity varies with age

I consider a parameterization that includes a simple difference in elasticies by age. Consider the baseline model from Section 1 where there are $I = 10$ types of individuals living for $T = 3$ periods. Suppose that $\sigma = 3$ for the second age group (workers in their forties) while

$\sigma = 2$ for the workers in their thirties and fifties. Given the isoelastic disutility function (1), the constant-consumption elasticity of labor supply is $\frac{1}{\sigma-1}$, so these values imply an elasticity of 1 for the youngest and oldest groups and an elasticity of $\frac{1}{2}$ for the workers in their forties.

The results of this experiment are similar to those of the main analyses. Intratemporal distortions remain too high for the high-earning young and are used more in general by the Static Mirrlees policy than by the policies with age dependence. Average tax rates for workers in their thirties are even lower relative to older workers in this Partial Reform policy than in the model with uniform elasticities. The welfare gain from Partial Reform is unchanged at 1.8 percent of aggregate income, and age dependence captures over two-thirds of the potential gain from the Full Optimum.

6.4 Extensive Margin

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

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

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

$$\beta^{t-1}\left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) - \phi\right) \geq \beta^{t-1}u\left(c_t^0\right), \tag{11}$$

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

$$\beta^{t-1}\left(u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) - \phi\right) \geq \beta^{t-1}\left(u\left(c_t^j\right) - v\left(\frac{y_t^j}{w_t^i}\right) - \phi\right), \tag{12}$$

for all $i, j \in \{1, 2, ..., I\}$ and all t, which simplify to the same conditions as from the baseline model because ϕ cancels on both sides. I simulate the model with $\phi = \ln(1.712)$, representing a fixed cost equal to approximately 30 percent of the log average wage for type $i = 1$.

The lessons from this paper's main analyses are unchanged by adding an extensive margin, though the optimal policies do respond to the extensive margin. One response of policy is that, while average tax rates have the same shape as in the baseline model, they are increased throughout the income distribution by the addition of an extensive margin. Intuitively, consumption is being provided to the $i = 0$ individuals who do not work, so average taxes on all other types must increase. A second, more subtle response is consistent with the analysis of Saez (2002). In the simulation of policy with an extensive margin, allocations mimic the U.S. Earned Income Tax Credit, whereby low earners receive a subsidy (i.e., a negative marginal tax rate) to encourage them to work rather than claim the $i = 0$ allocation.