PREFERENCE HETEROGENEITY AND OPTIMAL CAPITAL INCOME TAXATION

Mikhail Golosov
Maxim Troshkin
Aleh Tsyvinski
Matthew Weinzierl

Preference Heterogeneity and Optimal Capital Income Taxation
Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl
NBER Working Paper No. 16619
December 2010
JEL No. E6,E62,H2,H21

ABSTRACT

We examine a prominent justification for capital income taxation: goods preferred by those with high ability ought to be taxed. In an environment where commodity taxes are allowed to be nonlinear functions of income and consumption, we derive an analytical expression that reveals the forces determining optimal commodity taxation. We then calibrate the model to evidence on the relationship between skills and preferences and extensively examine the quantitative case for taxes on future consumption (saving). In our baseline case of a unit intertemporal elasticity, optimal capital income tax rates are 2% on average and 4.5% on high earners. We find that the intertemporal elasticity of substitution has a substantial effect on optimal capital taxation. If the intertemporal elasticity is one-third, optimal capital income tax rates rise to 15% on average and 23% on high earners; if the intertemporal elasticity is two, optimal rates fall to 0.6% on average and 1.6% on high earners. Nevertheless, in all cases that we consider the welfare gains of using optimal capital taxes are small.

Mikhail Golosov
Department of Economics
Yale University
Box 208268
New Haven, CT 06520-8269
and NBER
m.golosov@yale.edu

Maxim Troshkin
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 South 4th Street
Minneapolis MN 55455-0462
troshkin@umn.edu

Aleh Tsyvinski
Department of Economics
Yale University
Box 208268
New Haven, CT 06520-8268
and NBER
a.tsyvinski@yale.edu

Matthew Weinzierl
Harvard Business School
277 Morgan
Soldiers Field
Boston, MA 02163
and NBER
mweinzierl@hbs.edu

Introduction

A prominent justification for positive capital income taxation is that goods preferred by high-ability individuals ought to be taxed because consumption of these goods provides a signal of individuals' otherwise unobservable ability. If individuals' abilities are positively related to preferences for saving, this argument implies that capital income should be taxed. The key exposition of this justification is Saez (2002). Saez shows that a small *linear* tax on a commodity preferred by individuals with higher ability generates a smaller efficiency loss than does an increase in the optimal nonlinear income tax that raises the same revenue from each individual. He applies this logic to capital income taxation and concludes that, assuming the discount rate is negatively correlated with skills, interest income ought to be taxed. Importantly, Banks and Diamond (2009) in the chapter on direct taxation in the *Mirrlees Review* use this justification as one of the essential arguments for why policymakers ought to tax capital. Commissioned by the Institute for Fiscal Studies, the *Mirrlees Review* is the successor to the influential *Meade Report* of 1978 and is the authoritative summary of the current state of tax theory as it relates to policy. Their chapter concludes: "With the plausible assumption that those with higher earnings abilities discount the future less (and thus save more out of any given income), then taxation of saving helps with the equity-efficiency trade-off by being a source of indirect evidence about who has higher earnings abilities and thus contributes to more efficient redistributive taxation."

We study the case for taxing goods preferred by those with high ability when commodity taxes are allowed to be *nonlinear* functions of both income and consumption. In particular, we focus on the taxation of future consumption (i.e., saving). In other words, this paper addresses the question whether taxing capital is a good or a bad idea in an environment with heterogeneous discount factors.[1] We analytically show that heterogeneity in preferences across goods adds a force calling for nonlinear taxation that discourages lower earners from consuming a good preferred by high earners. These optimal distortions encourage effort among high earners by threatening a larger distortion to their choices if they earn less. Quantitatively our main finding is that, for a plausibly calibrated model, preference heterogeneity of this type recommends capital income tax rates that are 2 percent on average, and converge to 4.5 percent for high incomes. Tax rates can be substantially higher if the intertemporal elasticity of substitution is lower. In all cases, however, the welfare gains due to these capital income taxes are small. Our work is in the line of recent research, such as Golosov and Tsyvinski (2006), Ales and Maziero (2009), Weinzierl (forthcoming), and Kocherlakota and Pistaferri (2007, 2009), that uses micro level data to evaluate predictions of dynamic optimal policy models.[2]

Our specific results are as follows. We first derive analytical expressions that determine the shape of optimal commodity taxation. We start in a two-type, two-commodity economy and demonstrate that the high ability type faces no distortion to its chosen commodity basket, while the low type faces a distortion away from consumption of the good preferred by the high type. We show that this simple example illustrates a key intuition: the distortion faced by a high type if it mimics a lower type is larger than the distortion the

---

[1] The answer of the survey by Atkeson, Chari, and Kehoe 1998 in a Ramsey environment is "Taxing Capital Income: A Bad Idea." Conesa, Kitao, and Krueger 2008 argue in a quantitative OLG model with heterogeneous agents that "Taxing Capital? Not a Bad Idea After All!". The New Dynamic Public Finance literature argues for a generally positive capital wedge (see, e.g., surveys by Golosov, Tsyvinski, and Werning 2006 and Kocherlakota 2010).

[2] See also Pavoni and Violante (2007) for an application of optimal insurance models to the design of welfare to work programs.

high type faces if it truthfully reveals its type. We then examine an economy with two goods and a continuum of types where the relative preference for one good rises with ability. As in Diamond (1998), Saez (2001), and Golosov, Troshkin, and Tsyvinski (2010a,b), we analytically study the forces driving the optimal distortions to commodity choices. Our analysis shows that the key force of optimal nonlinear commodity taxation in this setting is that it discourages the consumption of a good preferred by high earners among lower earners. The intuition is as follows. The goal of optimal tax policy (in the Mirrleesian framework) is to redistribute from high-ability workers without discouraging their work effort. The optimal use of commodity taxation then aims to increase the attractiveness of earning a high income. High-ability individuals will choose to earn more if relative marginal commodity tax rates on the goods they most value generate distortions to their consumption choices that are greater when they earn less. These distortions allow the tax authority to levy higher income taxes on high-ability individuals and redistribute more resources to those with lower ability.

We then examine the quantitative case for capital income taxation in this environment. We use data from the National Longitudinal Survey of Youth (NLSY) to calibrate the relationship between ability[3] and intertemporal discounting, i.e., preferences for future relative to current consumption. This relationship is distinct from that between *income* and intertemporal discounting, which has been the focus of most of the relevant prior literature on preference heterogeneity. One exception is Benjamin, Brown, and Shapiro (2006), who find a positive relationship between ability and the holding of positive net assets, and our results are consistent with theirs. An important paper by Cagetti (2003) analyzes a related question: the relationship between education and time preferences. His finding–that higher education groups exhibit (substantially) greater preferences for saving–is consistent with the positive relationship between ability and savings preferences that we uncover in the data. For a state of the art review of earnings, consumption and life cycle choices, including environments with informational frictions, see the Handbook chapter by Meghir and Pistaferri (forthcoming).

Our main finding is that the computed optimal capital income tax rates for empirically plausible calibrations are as follows. For the baseline example of an intertemporal elasticity of substitution equal to one, optimal rates are U-shaped in income up to a high wage and then plateau at approximately \$150,000 of annual income.[4] The optimal maximal capital income tax rate is everywhere less than 4.54%, and the population-weighted average capital income tax rate is 2.0%. Welfare gains from these optimal capital income taxes are negligible.

We show that these baseline results are robust to varying the form of the social welfare function and the elasticity of labor supply. In contrast, the intertemporal elasticity of substitution, which equals $\frac{1}{\gamma}$ in our model, has a substantial effect on optimal capital income tax rates. The baseline assumption of $\gamma = 1$ is a standard benchmark in mainstream optimal tax and macroeconomic models. The smaller the intertemporal

---

[3] We measure ability by the survey respondent's score on the cognitive ability portion of the Armed Forces Qualification Test (AFQT). While it is impossible to measure ability perfectly, the AFQT score is commonly used, such as in the study of the returns to education.

[4] Each individual faces, in equilibrium, a distortion to consumption choices smaller than that if he earned less, consistent with our analytical results. The simulations are performed with a bounded distribution, so there is a highest type. This highest type faces no distortion, as shown analytically in Section 3. Whether rates decline with income for a range of types just below the highest type depends on the specification of the population distribution.

elasticity, the larger the optimal rates. For a low intertemporal elasticity ($\gamma = 3$), optimal rates rise to 15.0% on average and 23.5% on high earners, while for a high intertemporal elasticity ($\gamma = 0.5$) they rise to only 1.6%. Even when sizeable capital income tax rates are optimal, however, they still yield small welfare gains.

As an extension we also study optimal capital taxation in a stochastic setting in which there is a relationship between ex post ability and preferences over goods consumed within a period. We show that this relationship does not affect the optimal intertemporal distortion: i.e., the inverse Euler equation as in Golosov, Kocherlakota, and Tsyvinski (2003) continues to hold. Optimal distortions within the second period are similar to the results from the static model.

The idea that goods preferred by the highly able ought to be taxed has a long history in tax research and is a favorite of tax theorists. Nearly all comprehensive treatments of modern tax policy contain a section on this result. For example, Tuomala (1991) writes "...the marginal tax rates on commodities that the more able people tend to prefer should be greater;" Salanie (2003) warns "If there is a positive correlation between the taste for fine wines and productivity, then fine wines should be taxed relatively heavily (God Forbid!);" while Kaplow (2008) argues "it tends to be optimal to impose a heavier burden on commodities preferred by the more able and a lighter burden on those preferred by the less able." No doubt the enthusiasm for this result is due to the notion that, as Mirrlees put it "This prescription is most agreeable to common sense." In other words, taxes on goods preferred by high-ability individuals contribute to progressivity and the redistribution of income. The starting point in the literature for this idea is Mirrlees (1976, 1986), who shows that goods preferred by the able ought to be taxed. His results are on the ratio of after-tax to pre-tax prices for an individual's marginal purchase of a good, so they impose no linearity or income-independence constraints on optimal taxes. His results do not, however, tell us how these taxes ought to vary with the distribution of abilities or the details of individual preferences. Perhaps in part to make progress along this dimension, subsequent work often focused on linear, income-independent commodity taxes in the presence of preference heterogeneity (such as Saez, 2002). Our analysis returns to the general Mirrleesian setting, characterizing optimal policy analytically and, for capital taxation, quantitatively.

A contemporaneous analysis of this issue with a focus complementary to ours is Diamond and Spinnewijn (2009).[5] While we focus on the how preferences change with ability on average, they focus on heterogeneity of preferences among individuals with the same ability. In their model, individuals sort into occupations and the task of the tax authority is to use occupation-specific linear capital taxation to ensure that high-ability individuals of all preference types choose the high-productivity occupation. Because they assume that individuals with higher discount rates also have a lower willingness to work, Diamond and Spinnewijn find that the tax authority should levy a linear capital tax on the high earners and a linear subsidy on the low earners. This discourages the high-skilled, impatient workers from deviating to the low-productivity occupation. While important, this result depends on the absence of an intensive margin of effort and on the assumption of a positive relationship between discount rates and the willingness to work, which may be difficult to demonstrate with available empirical evidence. It also does not consider the possibility of

---

[5] Gordon and Kopczuck (2010) also find evidence to suggest including capital income in the tax function because of information it carries about individuals' wages.

nonlinear capital taxation. The approach we take in this paper allows for nonlinear capital taxation and is set in the standard Mirrleesian framework where individuals choose effort.

The paper proceeds as follows. Section 1 provides an illustrative example of our theoretical results in an economy with two ability types and heterogeneity in preferences over two goods. Section 2 derives conditions on the optimal policy in a general model of optimal taxation with a continuum of ability types and heterogeneity in preferences. In Section 3, we calibrate the model to data from the NLSY on heterogeneous time preferences and calculate optimal distortions for a baseline setting. In Section 4 we extensively examine the robustness of the baseline results to variation in the intertemporal elasticity of substitution, the labor supply elasticity, and the form of the social welfare function. We also compare the estimated relationship between ability and time preference to that which would be required for the prevailing capital income tax rates of developed economies to be optimal. Section 5 considers the dynamic, stochastic model. An Appendix contains technical details referred to in the text.

1   A simple example

In this section we introduce a simple two-type example that captures the main intuition behind the more general model. We show that, in this setting, the optimal relative commodity tax discourages the consumption by the low ability agents of the good preferred by the high ability agents. In particular, the relative marginal tax (wedge) is positive on this good for the low-ability individual, while the high-ability individual faces no distortion.

There is a continuum of measure one of two types of individuals indexed by $i = \{l, h\}$. The size of each group is equal to $1/2$. These individuals differ in wage (skill) w^i, where $w^h > w^l > 0$. The wage is private information to the agent. There are two commodities. The consumption of each commodity by an agent of type i is denoted by c_1^i and c_2^i. The utility function for an individual $i = \{l, h\}$ is given by:

$$u^i \left(c_1^i, c_2^i, \frac{y^i}{w^i} \right),$$

where y_i denotes the amount of output (income) produced by the agent. That is, the agent i provides the amount of labor $l^i \geq 0$ to produce output $y^i = w^i l^i \geq 0$. The planner observes output y^i but not the wage w^i or effort l^i. Agents' consumption of each good $\{c_t^i\}_{t=\{1,2\}}^{i=\{l,h\}} \geq 0$ is also observable. Let u_n^i be the partial derivative of $u^i(c_1, c_2, l)$ with respect to the nth argument. Note that these marginal utilities, and preferences in general, may depend on ability. We assume that $u_n^i > 0$ for $n = \{1, 2\}$ and $u_3^i < 0$.

The planner's problem is a mechanism design problem in which the mechanism assigns consumption and income allocations to each wage type reported by agents. The planner designs the mechanism to maximize a Utilitarian social welfare function.

Problem 1 *Planner's problem in two-type example:*

$$\max_{\left\{c_1^i, c_2^i, y^i\right\}_{i=l,h}} \sum_i u^i \left(c_1^i, c_2^i, \frac{y^i}{w^i} \right) \tag{1}$$

subject to the incentive compatibility constraint

$$u^h \left(c_1^h, c_2^h, \frac{y^h}{w^h} \right) \geq u^h \left(c_1^l, c_2^l, \frac{y^l}{w^h} \right), \tag{2}$$

and the feasibility constraint

$$\sum_i \left(y^i - c_1^i - p_2 c_2^i \right) \geq 0. \tag{3}$$

Constraint (2) is an incentive compatibility constraint stating that an individual of type $i = h$ prefers the consumption and income bundle intended for it by the planner, $\{c_1^h, c_2^h, y^h\}$, to a bundle $\{c_1^l, c_2^l, y^l\}$ allocated to an individual of type $i = l$.[6] Constraint (3) is feasibility, where we assume that the marginal rate of transformation of consumption commodities is equal to the price ratio $\frac{1}{p_2}$.

Consider first a benchmark environment in which the wage w^i is observable to the planner. Then the constrained efficient problem does not have the incentive compatibility constraint (2). The solution is an undistorted consumption margin for both ability types – the marginal rate of substitution across commodities is equal to the marginal rate of transformation:[7]

$$\frac{u_1^i \left(c_1^i, c_2^i, \frac{y^i}{w^i} \right)}{u_2^i \left(c_1^i, c_2^i, \frac{y^i}{w^i} \right)} = \frac{1}{p_2}. \tag{4}$$

for $i = \{l, h\}$.

Now, consider a program with unobservable wages. Let $\mu \geq 0$ be the multiplier on constraint (2). From the first order conditions for consumption, we obtain the following expressions for the marginal rate of substitution between consumption commodities for the high-wage individual, type $i = h$:

$$\frac{u_1^h \left(c_1^h, c_2^h, \frac{y^h}{w^h} \right)}{u_2^h \left(c_1^h, c_2^h, \frac{y^h}{w^h} \right)} = \frac{1}{p_2}, \tag{5}$$

and for the low-wage individual, type $i = l$:

$$\frac{u_1^l \left(c_1^l, c_2^l, \frac{y^l}{w^l} \right)}{u_2^l \left(c_1^l, c_2^l, \frac{y^l}{w^l} \right)} = \frac{1}{p_2} \frac{1 - \frac{u_2^h \left(c_1^l, c_2^l, \frac{y^l}{w^h} \right)}{u_2^l \left(c_1^l, c_2^l, \frac{y^l}{w^l} \right)} \mu}{1 - \frac{u_1^h \left(c_1^l, c_2^l, \frac{y^l}{w^h} \right)}{u_1^l \left(c_1^l, c_2^l, \frac{y^l}{w^l} \right)} \mu}. \tag{6}$$

Equation (5) shows that the consumption choices of the high-ability individual are undistorted. The marginal rate of substitution $\frac{u_1^h(\cdot)}{u_2^h(\cdot)}$ is equal to the marginal rate of transformation, which is equal to $\frac{1}{p_2}$. Equation (6) shows that if the multiplier μ on the incentive compatibility constraint is not equal to zero, then the consumption choices of the low-ability individual may be distorted.

Now, suppose we impose a condition requiring that if all individuals are given the same consumption and

---

[6] Writing this constraint we assumed that only an individual of type $i = h$ can misrepresent his type. This is easy to ensure if the ratio w^h / w^l is high enough.

[7] The consumption-labor margin is also undistorted.

income allocation, (c_1, c_2, y), the marginal utility of good 2 relative to good 1 is higher for the high-ability individual (type $i = h$) than for the low-ability individual (type $i = l$). This condition on the relative shape of indifference curves between goods for individuals of different ability levels resembles that discussed by Mirrlees (1976) in equation (37) of his treatment of this topic.[8]

Assumption 1 The utility function u satisfies:

$$\frac{u_2^h\left(c_1, c_2, \frac{y}{w^h}\right)}{u_1^h\left(c_1, c_2, \frac{y}{w^h}\right)} > \frac{u_2^l\left(c_1, c_2, \frac{y}{w^l}\right)}{u_1^l\left(c_1, c_2, \frac{y}{w^l}\right)} \tag{7}$$

for any $(c_1, c_2, y) \geq 0$.

The first order conditions (5) and (6), together with Assumption 1, imply a proposition characterizing the distortions in the optimal allocation.

Proposition 1 *Suppose that $\left\{c_1^i, c_2^i, y^i\right\}_{i=l,h}$ is an optimal allocation solving (1) through (3). Then the optimal choice of consumption for the high-ability individual ($i = h$) is not distorted. Suppose that Assumption 1 holds. Then the optimal choice of consumption for the low-ability agent ($i = l$) is distorted away from good 2 in favor of good 1:*

$$\frac{u_1^l\left(c_1^l, c_2^l, \frac{y^l}{w^l}\right)}{u_2^l\left(c_1^l, c_2^l, \frac{y^l}{w^l}\right)} < \frac{1}{p_2}.$$

This Proposition states that if good 2 is particularly enjoyed by high-ability workers, the planner should impose a distortion (i.e., a positive relative tax)[9] on the consumption of good 2 by the low-ability workers (but not on consumption of that good by high-ability workers). The intuition for this result is as follows. The planner wants to discourage a high-ability individual from deviating and claiming that he is a low type. A high-ability agent finds deviating less attractive if doing so causes him to face a positive relative tax on the good that he values highly. The cost to the planner of such a positive relative tax is a distortion in the consumption choices by the low-ability agent. Assumption 1 ensures that the costs of such distortion are smaller than the gain from relaxing the incentive compatibility constraint.

It is important to be clear that this result depends on preferences varying by ability level, not income. In particular, it does not apply to goods with an income elasticity of demand greater than one but for which preferences are unrelated to ability. For those goods, the inequality in (7) would be an equality because each type would have the same ratio of marginal utilities given the same consumption and income bundle. Instead, the case for differential taxes requires the high-ability individuals to prefer good 2 even when at the same income level as the low-ability individuals.

Related to Proposition 1, we now derive a second result characterizing the design of optimal nonlinear commodity taxes. This result compares the distortions that individuals face under the optimal policy when

---

[8]Though Assumption 1 does not rule out complementarity of consumption and leisure, our results do not rely on that. In fact, in the more general analysis below, we specify separability to highlight that our results are driven by the effects that preferences have on the sub-utility from consumption goods. The alternative, in which some goods are more complementary to labor than others, provides a second reason to deviate from uniform commodity taxation first discussed by Corlett and Hague, 1953.

[9]In this model, equilibrium wedges correspond to taxes for each type. We interchangeably use the terms wedges and taxes to refer to these distortions.

they reveal their type and when they mimic a lower type. We call the latter the "deviator's distortion" to contrast it with the distortion faced by individuals who truthfully reveal their types.

Definition 1 The "deviator's distortion $(i'|i)$" is defined as $\tau^{i'|i}$ (for $i \neq i'$):

$$\tau^{i'|i} = \frac{u_1^i\left(c_1^{i'}, c_2^{i'}, \frac{y^{i'}}{w^i}\right)}{u_2^i\left(c_1^{i'}, c_2^{i'}, \frac{y^{i'}}{w^i}\right)}.$$

In words, this measures the distortion to the consumption choices of an individual of type i who reports being of type i' and receives the latter's allocation of consumption and income.

We now state the following corollary.

Corollary 1 *Suppose that $\left\{c_1^i, c_2^i, y^i\right\}_{i=l,h}$ is an optimal allocation solving (1) through (3). Then the optimal choice of consumption for the high-ability agent $(i = h)$ is distorted away from good 2 in favor of good 1 more strongly by the "deviator's distortion $(l|h)$" than by the distortion the high-ability agent faces if it reveals its type. Formally,*

$$\frac{u_1^h\left(c_1^l, c_2^l, \frac{y^l}{w^h}\right)}{u_2^h\left(c_1^l, c_2^l, \frac{y^l}{w^h}\right)} < \frac{u_1^h\left(c_1^h, c_2^h, \frac{y^h}{w^h}\right)}{u_2^h\left(c_1^h, c_2^h, \frac{y^h}{w^h}\right)} = \frac{1}{p_2}.$$

Proof. This result follows immediately from the previous Proposition and Assumption 1. ∎

Corollary 1 helps with understanding the role of optimal commodity taxes and shows that the planner encourages individuals to exert effort by threatening them with higher distortions to their consumption choices if they earn less. The relevant distortions these individuals would face if they earned less are not the distortions faced by lower-ability individuals who tell the truth about their type, because preferences differ with ability. Specifically, because higher ability individuals prefer good 2 in our example, a distortion away from good 2 as perceived by type i' is perceived to be more distortionary by type i with $w^i > w^{i'}$. This "deviator's distortion $(i'|i)$" adds an incentive for high ability individuals to exert effort.

2 Model

In this section, we set up a model with a continuum of ability types, as in the classic Mirrlees (1971) framework. Agents are heterogeneous in their preferences. We derive a formula for optimal relative commodity taxes that are allowed to be nonlinear in consumption and to depend on income, and we explain the novel components of this formula relative to models without preference heterogeneity.

There is a continuum of measure one of individual agents. Agents are indexed by $i \in [0, 1]$. Individuals differ in their abilities, which we measure with their wages, denoted by w^i and distributed according to the density function $f(w)$ over the interval $[w_{\min}, w_{\max}]$. Ability is private information to the agent. Each individual has the continuous and differentiable utility function:

$$U\left(w^i\right) = u\left(c_1^i, c_2^i, l^i, w^i\right). \tag{8}$$

Utility is a function of the consumption of good 1, $c_1 \geq 0$, and the consumption of good 2, $c_2 \geq 0$, as well as

of labor effort $l \geq 0$.[10] Superscripts i on consumption and labor denote the values of these variables for the individual, and the partial derivatives of utility take the following signs: $u_{c_1}(\cdot) > 0$, $u_{c_2}(\cdot) > 0$, $u_l(\cdot) < 0$. The output $y^i = w^i l^i \geq 0$. Utility is also a function of the wage w^i because we assume that preferences across consumption goods are a function of ability. This assumption simplifies the planner's problem by retaining a single dimension of heterogeneity. Two or more dimensions introduce a multiple screening problem for which a tractable analytical approach at this level of generality has not been developed.[11] Later, we will parameterize the influence of ability on preferences with the function $\alpha(w^i)$ where $\alpha(w^i) > 0$ for all w^i.

A social planner maximizes a utilitarian social welfare function. The planner's problem is given as follows.

Problem 2

$$\max_{\{c_1^i, c_2^i, y^i\}_{i \in [0,1]}} \int_{w_{\min}}^{w_{\max}} u\left(c_1^i, c_2^i, \frac{y^i}{w^i}, w^i\right) f(w^i) \, dw^i \tag{9}$$

subject to the feasibility constraint

$$\int_{w_{\min}}^{w_{\max}} \left(y^i - c_1^i - p_2 c_2^i\right) f(w^i) \, dw^i \geq 0, \tag{10}$$

and the incentive compatibility constraint

$$u\left(c_1^i, c_2^i, \frac{y^i}{w^i}, w^i\right) \geq u\left(c_1^j, c_2^j, \frac{y^j}{w^i}, w^i\right), \tag{11}$$

for all $i, j \in [0,1]$.

Constraint (11) is the incentive compatibility constraint ensuring that an individual of type i prefers the consumption and income allocation intended for it by the planner to the allocations intended for any other individual of type j. As in the previous section, the relative price of c_2 is p_2.

It is standard to rewrite the planner's problem with explicit tax functions. To characterize the form of these optimal tax functions, we follow the formal techniques of the Mirrleesian literature. We start with the statement of the problem solved by each individual, who takes the tax functions as given.

Problem 3 *Individual's Problem, $i \in [0,1]$:*

$$\max_{\{c_1^i, c_2^i, l^i\}} U(w^i) \tag{12}$$

subject to the individual's after-tax budget constraint,

$$w^i l^i - T(w^i l^i) - \left(c_1^i + t^1(w^i l^i, c_1^i)\right) - p_2\left(c_2^i + t^2(w^i l^i, c_2^i)\right) \geq 0. \tag{13}$$

The budget constraint requires careful examination. The nonlinear income tax $T(w^i l^i) : R_+ \rightarrow R$ is a continuous, differentiable function of income $y^i = w^i l^i$. The two other tax functions, $t^1(w^i l^i, c_1^i)$,

---

[10] Extending the model to more than two goods (for example, to more than two periods) is straightforward. The analytical results on optimal distortions are direct analogues of those derived below.

[11] See Kleven, Kreiner, and Saez (2009), Tarkiainen and Tuomala (2007), and Judd and Su (2008) for discussions of the approach to optimal taxation with multi-dimensional heterogeneity.

$t^2\left(w^i l^i, c_2^i\right) : R_+ \times R_+ \to R$ are commodity tax functions that we also assume to be continuous and differentiable. Importantly, note that we explicitly allow for the taxation of each commodity to be nonlinear in consumption of that good and to depend on income.[12] The budget constraint (13) has the multiplier $\mu\left(i\right) \geq 0$.

In this approach, the social planner's problem is as follows:

Problem 4 *Planner's Problem*

$$\max_{\{T(\cdot),t^1(\cdot),t^2(\cdot)\}_{i\in[0,1]}} \int_{w_{\min}}^{w_{\max}} U\left(w^i\right) f\left(w^i\right) dw^i \tag{14}$$

subject to the feasibility constraint

$$\int_{w_{\min}}^{w_{\max}} \left(T\left(w^i l^i\right) + t^1\left(w^i l^i, c_1^i\right) + p_2 t^2\left(w^i l^i, c_2^i\right)\right) f\left(w^i\right) dw^i \geq 0, \tag{15}$$

and incentive compatibility, which is that each individual $i \in [0,1]$ solves the optimization problem in (12), given tax policies $T\left(w^i l^i\right)$, $t^1\left(w^i l^i, c_1^i\right)$, and $t^2\left(w^i l^i, c_2^i\right)$.

In words, the social planner chooses a tax system to maximize utilitarian social welfare subject to two constraints. First, the budget constraint requires that total tax revenue be non-negative (we assume no government spending for simplicity). Second, each individual will respond to the tax system by choosing labor supply and a consumption bundle that maximize his or her utility.

2.1 The optimal commodity choice wedge

We now derive a formula that allows us to study the forces determining the optimal commodity wedge, i.e., the wedge distorting commodity choices. We formulate the Hamiltonian from the planner's problem (9) using the budget constraint, envelope condition, and first order condition with respect to labor l^i from the individual's utility maximization problem:

$$H\left(w^i\right) = \left(U\left(w^i\right) + \lambda\left(w^i l^i - c_1^i - p_2 c_2^i\right)\right) f\left(w^i\right) + \phi\left(u_{w^i}\left(\cdot\right) - \frac{l^i u_{l^i}\left(\cdot\right)}{w^i}\right), \tag{16}$$

where subscripts denote partial derivatives and $(\cdot)$ denotes the set of arguments of the utility function, $\left(c_1^i, c_2^i, l^i, w^i\right)$. The first term of the Hamiltonian is the utility of the individual with wage w^i. The second is the government's budget constraint multiplied by its multiplier λ. The third term is the evolution of the state variable $U\left(w^i\right)$ with respect to w^i, as derived above, and is multiplied by the co-state variable ϕ.[13]

To solve for the optimal policy, choose l and c_1^i as the control variables, with c_2^i an implicit function defined by the budget constraint. The first order condition with respect to c_1^i combined with the condition

---

[12] These tax instruments are notationally redundant, in that a single tax function of the consumption of one good and income would be sufficient to characterize the full policy. Separating taxes into these functions aids interpretation and has no effect on the analytical or quantitative results of the paper.

[13] The above procedure uses the so-called first order approach, where the first-order conditions of the individual's problem are assumed to be sufficient, not just necessary, conditions for a maximum. We check that these are sufficient in all numerical simulations we perform in Section 3.

that individuals will set the ratio of marginal utilities from the consumption goods equal to the price ratio multiplied by the marginal tax ratio, yields the following expression for the distortion to individual $i's$ consumption basket:

$$\frac{1}{p_2}\frac{1+t^1_{c_1}\left(w^i l^i, c^i_1\right)}{1+t^2_{c_2}\left(w^i l^i, c^i_2\right)} = \frac{1}{p_2}\frac{\lambda f\left(w^i\right) - \phi\left(w^i\right)\left(u_{w^i c^i_1}(\cdot) - \frac{l^i u_{l^i c^i_1}(\cdot)}{w^i}\right)}{\lambda f\left(w^i\right) - \frac{1}{p_2}\phi\left(w^i\right)\left(u_{w^i c^i_2}(\cdot) - \frac{l^i u_{l^i c^i_2}(\cdot)}{w^i}\right)}. \tag{17}$$

To further characterize the optimal distortion to commodity purchases given by (17), we solve for the multipliers λ and $\phi\left(w^i\right)$ under the following assumption:

Assumption 2 Utility function u in (8) is separable in consumption and labor:

$$u_{l^i c^i_1}\left(c^i_1, c^i_2, l^i, w^i\right) = u_{l^i c^i_2}\left(c^i_1, c^i_2, l^i, w^i\right) = 0 \tag{18}$$

The following proposition derives an expression for optimal commodity taxes.

Proposition 2 *Given Assumption 2 on the individual utility function, the solution to the Planner's Problem* (14) *satisfies:*

$$\frac{1}{p_2}\frac{1+t^1_{c_1}\left(w^i l^i, c^i_1\right)}{1+t^2_{c_2}\left(w^i l^i, c^i_2\right)} = \frac{A_1\left(w^i\right) + \frac{C(w^i)}{B(w^i)}}{A_2\left(w^i\right) + p_2\frac{C(w^i)}{B(w^i)}}, \tag{19}$$

where

$$A_1\left(w^i\right) = u_{w^i c^i_1}, \quad A_2\left(w^i\right) = u_{w^i c^i_2} \tag{20}$$

$$B\left(w^i\right) = p_2\left(\int_{w^j=w^i}^{w^j=w_{\max}}\frac{1}{u_{c^j_2}}f\left(w^j\right)dw^j - \frac{1-F\left(w^i\right)}{1-F\left(w_{\min}\right)}\int_{w^j=w_{\min}}^{w^j=w_{\max}}\frac{1}{u_{c^j_2}}f\left(w^j\right)dw^j\right) \tag{21}$$

$$C\left(w^i\right) = f\left(w^i\right) \tag{22}$$

Proof. In the Appendix, we derive the following expressions for λ and $\phi\left(w^i\right)$:

$$\lambda = \frac{1}{\int_{w^j=w_{\min}}^{w^j=w_{\max}}\frac{p_2}{u_{c^j_2}}f\left(w^j\right)dw^j} \tag{23}$$

$$\phi\left(w^i\right) = \left(1-F\left(w^i\right)\right)\left(1 - \frac{\frac{1}{(1-F(w^i))}\int_{w^j=w^i}^{w^j=w_{\max}}\frac{1}{u_{c^j_2}}f\left(w^j\right)dw^j}{\frac{1}{(1-F(w_{\min}))}\int_{w^j=w_{\min}}^{w^j=w_{\max}}\frac{1}{u_{c^j_2}}f\left(w^j\right)dw^j}\right). \tag{24}$$

Using these results in expression (17), we obtain (19). ∎

As with the conditions for optimal marginal income tax rates from, e.g., Diamond (1998), Saez (2001), and Golosov, Troshkin, and Tsyvinski (2010b), expression (19) is not a fully closed-form solution as it depends on optimal utility and consumption levels. Instead, it is a representation of the first order conditions of the optimal problem allowing us to examine the forces affecting optimal taxes.

We identify three important forces at play. Two are familiar from previous results in Mirrleesian optimal

taxation, for instance from the formulas for the income tax in Diamond (1998) and Saez (2001). However, they have no impact in our model without the existence of an additional, novel, force.

The novel force affecting distortions in result (19) is the disparity between $A_1\left(w^i\right)$ and $A_2\left(w^i\right)$, which are the derivatives of the marginal utility of consumption of goods 1 and 2 with respect to the wage. This disparity determines whether policy discourages consumption of good 1 or good 2. If $A_1\left(w^i\right)$ and $A_2\left(w^i\right)$ are equal (for instance, if they are both zero), there is no distortion to consumption choices in the optimal policy. If, instead, higher-ability workers relatively prefer good 2, then $u_{w^i c_1^i} < 0$ while $u_{w^i c_2^i} > 0$, so $A_1\left(w^i\right) < 0$ and $A_2\left(w^i\right) > 0$. In that case, because both $B\left(w^i\right)$ and $C\left(w^i\right)$ are non-negative, the ratio on the right-hand side of (19) is less than $\frac{1}{p_2}$, and the optimal distortion discourages marginal consumption of good 2. In other words, preferences over goods that vary with ability introduce a reason for using differentiated marginal commodity taxes to provide incentives for high-ability individuals to exert work effort.

Whether the distortion to consumption choices increases or decreases with wages depends on the behavior of $A_1\left(w^i\right)$ and $A_2\left(w^i\right)$ as the wage level increases. Though a full characterization depends on the specific form of the utility function, the lower marginal utilities of consumption that come with higher w^i will push $A_1\left(w^i\right)$ and $A_2\left(w^i\right)$ toward zero as the wage level increases, reducing the size of this distortion at higher wages. Intuitively, marginal commodity tax rates that decline with income on the good more valued by high-ability individuals will encourage them to earn more, allowing the tax authority to levy higher income taxes on them and redistribute more resources to those with lower ability.

The two forces familiar from previous optimal tax analyses generate the ratio $\frac{C\left(w^i\right)}{B\left(w^i\right)}$ in (19). This ratio can be interpreted as the cost-benefit ratio of the distortion, so a higher value for it reduces the optimal distortion by offsetting the disparity between $A_1\left(w^i\right)$ and $A_2\left(w^i\right)$.

First, $B\left(w^i\right)$ measures the redistributive benefit of a distortion at wage w^i. That distortion allows the planner to shift income from those with wages above w^i to the population as a whole, raising total welfare. Formally, consider a two-part perturbation in the planner's allocations made possible by this distortion. First, the planner lowers utility by 1 unit for each individual above w^i by extracting consumption from them while preserving incentive compatibility.[14] The planner extracts $p_2 \int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j$ in resources from this action, and it lowers social welfare by $\left(1 - F\left(w^i\right)\right)$ units. Second, the planner raises utility by $\left(1 - F\left(w^i\right)\right)$ units for each individual in the population by granting them additional consumption while preserving incentive compatibility. The cost to the planner of this action is $\frac{p_2}{1-F(w_{\min})} \int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1-F\left(w^i\right)}{u_{c_2^j}} f\left(w^j\right) dw^j$, and it raises social welfare by $\left(1 - F\left(w^i\right)\right)$ units. The net change in social welfare from these two actions is zero, while the net resources raised by the planner is:

$$p_2 \int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j - \frac{p_2}{1 - F\left(w_{\min}\right)} \int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1 - F\left(w^i\right)}{u_{c_2^j}} f\left(w^j\right) dw^j.$$

Rearranging this result, these two actions yield excess resources if:

$$B\left(w^i\right) > 0,$$

---

[14] Note that this action is possible only because of the distortion at w^i. Otherwise, individuals above w^i would respond by earning less.

so that the planner can raise social welfare through redistribution whenever $B\left(w^i\right)$ is positive. Moreover, the greater is $B\left(w^i\right)$), the more valuable is this distortion to the planner. Intuitively, higher-ability workers have lower marginal utilities of consumption, and the more concave is utility in good 2 above wage w^i, the more valuable is the redistribution made possible (i.e., incentive compatible) by the commodity choice distortion at w^i.

Second, $C\left(w^i\right)$ measures the cost of the distortion at wage w^i because it is the share of the population whose choices are directly affected by a commodity tax at w^i. When this share is low, the optimal consumption distortion (if non-zero) is larger, as the planner wants to concentrate distortions on small sub-populations all else the same. The ratio $\frac{C\left(w^i\right)}{B\left(w^i\right)}$ is multiplied by p_2 in the denominator because if $A_1\left(w^i\right) = A_2\left(w^i\right) = 0$, the undistorted marginal rate of transformation equals $\frac{1}{p_2}$.

We can derive several specific results that characterize the optimum and aid intuition. First, for the top type in a bounded ability distribution, $\left(1 - F\left(w_{\max}\right)\right)$ is zero, and the result (19) reduces to:

$$\frac{1}{p_2} \frac{1 + t^1_{c_1}\left(w_{\max}\right)}{1 + t^2_{c_2}\left(w_{\max}\right)} = \frac{1}{p_2}.$$

so the commodity distortion is zero on the highest ability worker.[15] Second, the distortion is also zero on the lowest ability worker, as $B\left(w_{\min}\right) = 0$. Third, if we restrict attention to commodity distortions that are a linear function of the consumption of the good, an argument similar to Saez (2002) and Salanie (2003) shows that goods preferred by the highly able ought to be taxed.

As in the two-type model of Section 1, we can clarify the way in which the optimal policy provides incentives for the high-ability individuals to exert effort. First, the "deviator's distortion $(i'|i)$" in this model equals:

$$\tau^{i'|i} = \frac{u_{c_1^{i'}}\left(c_1^{i'}, c_2^{i'}, \frac{y^{i'}}{w^i}, w^i\right)}{u_{c_2^{i'}}\left(c_1^{i'}, c_2^{i'}, \frac{y^{i'}}{w^i}, w^i\right)}.$$

We will show in numerical simulations in Section 3 that the "deviator's distortion $(i'|i)$" $(i' < i)$ discourages the consumption of commodity 2 more than does the truth-telling distortion for individual i, or that

$$\tau^{i'|i} < \frac{1}{p_2} \frac{1 + t^1_{c_1}\left(w^i l^i, c_1^i\right)}{1 + t^2_{c_2}\left(w^i l^i, c_2^i\right)}.$$

One may also be interested in the pattern of marginal tax rates on income that are the focus of the conventional Mirrleesian optimal tax literature. In this paper's multiple-commodity setting, the marginal tax rate on income can be calculated relative to the distortion to consumption of one of the commodities. In the Appendix, we derive an expression analogous to (19) that describes these relative marginal income taxes. In addition, the numerical results of the next section can be used to calculate the optimal marginal income taxes (relative to commodity consumption) implied by the data.

---

[15] If the ability distribution is unbounded, as argued by Saez (2001), the pattern of rates near the top of the distribution depends on the specification of preferences. Formally, if A_1 and A_2 decrease quickly enough with w^i, the optimal distortion falls with wages as well.

3 Optimal capital income taxes in a calibrated model

The results of Sections 1 and 2 show the forces affecting optimal commodity taxation when preferences over goods vary with ability. We now turn to a quantitative study of this topic when the commodities in the utility function are current and future consumption (savings).

We begin our quantitative analysis of optimal capital income taxation by discussing the existing literature on the relationship between time preferences and income. That relationship is distinct from the relationship that matters for this paper: that between time preferences and *ability*. We provide a calibration of time preferences by ability level (and thus wages) using data from the National Longitudinal Survey of Youth. We then simulate the optimal capital income taxes justified by these estimates and relate our results to the analytical expression (19) from the previous section.

3.1 Calibrating the model

In this section, we calibrate the model of optimal commodity taxation from Section 2. In particular, we estimate the relationship between time discounting and cognitive ability using panel data from the National Longitudinal Survey of Youth (NLSY).

A sizeable literature exists on measuring and explaining differences in saving behavior across income groups.[16] Dynan, Skinner, and Zeldes (2004) find a "strong, positive relationship between saving rates and lifetime income," using data from the PSID, but they argue that preference differences cannot explain their findings (at least, without a strong bequest motive). Lawrance (1991) calculates annual time preference rates using data on food consumption and finds that implied discount factors rise with income, but Dynan (1993) shows that Lawrance's results are sensitive to the inclusion of controls.

Less research exists on whether saving preferences are related to innate *ability*, the relationship of interest for our analysis. One exception is Benjamin, Brown, and Shapiro (2006), who use the same dataset we use and find results consistent with ours, though they focus on a different measure of the relationship between preferences and ability. An intermediate case is an important paper by Cagetti (2003), who estimates discount factors by education group using detailed micro data on income and wealth from the PSID and SCF. As with income, educational attainment is likely to be affected by individuals' time preferences, so Cagetti's estimates cannot directly be used to determine the relationship of interest in this paper. Nevertheless, his finding that higher education groups exhibit (substantially) greater preferences for saving is consistent with the positive relationship between ability and savings preferences that we uncover in the data.

With the goal of calibrating our model, we provide evidence on the relationship between saving preferences and ability. In brief, our approach is to use data on income and net worth from the National Longitudinal Survey of Youth (NLSY) and a standard model of an individual's intertemporal utility maximization problem to compute a discount factor for each individual in the sample. Next, we regress these discount factors on the log of ability and other personal characteristics observed by the NLSY, where we measure ability with individuals' scores on a widely-used aptitude test. The coefficient on ability in this regression allows us to

---

[16]For example, see Hubbard, Skinner, and Zeldes (1995). A related literature focuses on the consequences of these differences for the extent of self-insurance against shocks. See Blundell, Pistaferri, and Preston (2008), for example.

predict, holding fixed other personal characteristics, a discount factor for each level of ability. Using NLSY data on wages by ability level, we are then able to estimate a functional relationship between discount factors and wages, the key inputs to our policy simulations. To summarize, we estimate an elasticity of the annual discount factor β to the wage w of 0.0036. For example, a change in the wage from \$20 to \$24 per hour, a 20 percent increase, corresponds to a change in the annual discount factor from 0.9604 to 0.9610, a 0.07 percent increase.

Throughout the numerical analysis, we use the following utility function. For consistency with the previous section, we consider a utility function that is separable in consumption and labor. Preferences over goods are normalized so that they do not mechanically affect labor effort, as detailed in the Appendix. In addition, we assume that utility from consumption is constant relative risk aversion (CRRA) and the disutility from labor effort is isoelastic:

$$U = \left(\frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)}\right)^\gamma \frac{\left(c_1^i\right)^{1-\gamma}-1}{1-\gamma} + \left(\frac{1}{1+\alpha\left(w^i\right)}\right)^\gamma (p_2)^{1-\gamma}\frac{\left(c_2^i\right)^{1-\gamma}-1}{1-\gamma} - \frac{1}{\sigma}\left(l^i\right)^\sigma. \tag{25}$$

As a baseline case, we assume $\gamma = 1$ and $\sigma = 3$. With $\gamma = 1$, this utility function simplifies to

$$U = \frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)}\ln c_1^i + \frac{1}{1+\alpha\left(w^i\right)}\ln c_2^i - \frac{1}{\sigma}\left(l^i\right)^\sigma \tag{26}$$

We now provide some more details on our calibration, beginning with the data. The NLSY consists of a nationally representative sample of individuals born between 1957 and 1964, first interviewed in 1979, and interviewed annually or biannually since. The NLSY contains data on individuals' net worth and income over time, allowing us to roughly estimate saving rates as described below.

The key advantage of the NLSY for our purposes is that it includes a direct measure of ability. This allows us to relate a measure of ability, not income, to time preferences. In 1980, the NLSY administered the Armed Forces Qualification Test (AFQT) to 94 percent of its participants. This test measured individuals' aptitudes in a wide range of areas, including some mechanical skills relevant to military service.

We use an aggregation of scores in some of the areas covered by the AFQT as the indicator of ability for each head of household whose family income and net worth we will measure.[17] This aggregation, the AFQT89, is calculated by the Center for Human Resource Research at Ohio State University, as follows:

> Creation of this revised percentile score, called AFQT89, involves (1) computing a verbal composite score by summing word knowledge and paragraph comprehension raw scores; (2) converting subtest raw scores for verbal, math knowledge, and arithmetic reasoning; (3) multiplying the verbal standard score by two; (4) summing the standard scores for verbal, math knowledge, and arithmetic reasoning; and (5) converting the summed standard score to a percentile.

Our measure of preferences is based on the discount factor δ implied by using NLSY data on individuals'

---

[17]The AFQT most likely measures some combination of innate ability and accumulated achievement. To the extent that more innately patient individuals invest more in human capital and thereby have higher AFQT scores because of achievement, not ability, our analysis will be biased toward finding a stronger relationship between ability and time preferences than that which truly holds.

household income paths and net worth in a simple model of optimization described in the Appendix. Intuitively, the higher is final net worth relative to the cumulative value of income, the greater the estimated δ.[18] To give a sense for the data, in Table 1 we show the mean and standard deviations of δ by AFQT quintile.

Table 1. Summary of δ by AFQT quintile					
	AFQT quintile				
	Bottom	2	3	4	Top
Mean δ	0.336	0.374	0.394	0.418	0.466
Std. dev. of δ	0.156	0.176	0.180	0.215	0.252
Implied $\alpha\left(w^i\right)$	1.0486	1.0437	1.0413	1.0387	1.0338
Implied $\beta\left(w^i\right)$	0.9536	0.9581	0.9603	0.9628	0.9673
Mean w^i	12.35	16.29	18.98	21.67	25.39

Table 1 also shows the implied values of $\alpha\left(w^i\right)$, the parameter of interest from the model of Section 2, and $\beta\left(w^i\right)$, the standard annual discount factor.[19] The variation in δ within AFQT quintiles is large relative to the variation across wage levels. These results have their limitations for use in calibrating our model. The data are likely to be very noisy, and our inference of δ is based on a simplified model. Moreover, simple AFQT quintile means of δ are likely to be misleading, as they fail to control for variables correlated with both ability and saving behavior.

Table 2 shows the results of a regression of $\ln\left(\delta\right)$ on ability as well as other observable characteristics. In particular, we control for the cumulative value of family income over the head of household's working life, the head's age, age squared, and gender.[20] Formally, we estimate:

$$\ln \delta = \beta_1 age + \beta_2 age^2 + \beta_3 gender + \beta_4 \ln\left(income\right) + \beta_5 \ln\left(AFQT\right).$$

where the calculation of "$income$" is described in the Appendix.

---

[18] We lack data on families' expected future income flows from sources such as Social Security and bequests. To the extent that these flows are greater relative to past income for low earners, we are underestimating the true δ for low earners and thereby overestimating the strength of the relationship between ability and savings preferences. If these flows are greater for high earners, we are underestimating the relationship's strength. Similarly, we do not take into account the existing tax system when estimating δ. If capital income tax rates are progressive, this will cause us to underestimate the positive relationship between ability and savings preferences.

[19] Note that, because the model with which we estimate δ uses 23-year periods, δ is the discount factor across these periods. The Appendix describes how to convert δ into the preference parameter $\alpha\left(w^i\right)$ from the utility function (25).

[20] The estimate of the coefficient on $\ln(AFQT)$ is 2.71E-02 (4.45E-03) if we do not control for age, age squared, or gender.

Table 2. Regression results			
Variable	Coefficient	Std. Err.	t-statistic
age	-2.62E-02	2.97E-02	-0.88
age^2	8.80E-04	8.36E-04	1.05
gender	1.16E-02	8.15E-03	1.42
$\ln(income)^{**}$	1.69E-01	7.61E-03	22.15
$\ln(AFQT)^{**}$	2.60E-02	4.46E-03	5.82

Note: ** indicates significance at the 1% level

Observations: 7008

F-statistic: 203.98

Adjusted R-squared: 0.127

This regression yields a highly significant estimate for β_5 of 0.026 (standard error of 0.004).[21,22] In words, this coefficient implies an elasticity of 0.026 for the discount factor δ with respect to ability as measured by the AFQT. For example, if ability increases by 10 percentile points from 50 to 60 (a twenty percent increase), the discount factor δ would increase from 0.394 to 0.396 (i.e., by approximately 0.47 percent). These findings are consistent with the findings of the literature cited above that relates saving to income and with Benjamin, Brown, and Shapiro (2006), who find a "strong, statistically significant, and positive relationship between AFQT score and the propensity to have positive net assets" in the NLSY. Those authors, using a different measure of time preference, report "an additional 10 percentile points of AFQT is associated with an increase of about 1.5 percentage points in the propensity to have positive net assets."

The estimate of β_5 allows us to derive a value of δ for each ability level holding fixed an individual's age, gender, and cumulative income. In particular, we use

$$\delta = 0.356 \left(AFQT\right)^{0.026}, \tag{27}$$

where the constant 0.356 is pinned down by matching the value of δ for the middle AFQT quintile from Table 1 (0.394) with the mean AFQT score in that quintile (49.26). Expression (27) allows us to calculate, from the average AFQT score by quintile, a "regression-based δ" for each quintile that can be compared to the simple means in Table 1. The results are shown in Table 3, along with the implied values of $\alpha\left(w^i\right)$ and $\beta\left(w^i\right)$.

---

[21] We also have run simulations controlling for the slope of income during the 1979-2004 period and over the past ten years for each individual. These controls reduce the coefficient on AFQT to 0.021 and 0.015, but it remains significant at the 1% level. Note that these results imply a weaker relationship between ability and preferences.

[22] Measurement error likely affects both our estimates of ability and discounting, though bias would be introduced only by error in the former. While AFQT is an imperfect measure of ability, its retest reliability is very high. Moreover, if AFQT mismeasures ability, it is unclear whether that biases our results down or up. It may be that AFQT measures those parts of ability that are particularly highly correlated with preferences (i.e., ability to delay gratification, cognitive alacrity), and a more accurate measure of ability would show less relationship with preferences.

Table 3. Regression-based δ by AFQT quintile					
	AFQT quintile				
	Bottom	2	3	4	Top
Mean δ	0.378	0.389	0.394	0.398	0.400
Implied $\alpha\left(w^i\right)$	1.0433	1.0419	1.0413	1.0409	1.0406
Implied $\beta\left(w^i\right)$	0.9585	0.9598	0.9603	0.9607	0.9610
Mean w^i	12.35	16.29	18.98	21.67	25.39

The final step is to relate these discount factors to wages, as wage rates are the measure of ability in the model from Section 2 that we will use to simulate optimal policy. The NLSY provides data on individuals' reported wages, and we report the average of these wages by AFQT quintile in Table 3.[23] Assuming the same functional form as in expression (27), the values of $\alpha\left(w^i\right)$ and w^i in Table 3 imply the following relationship between discounting and wages:

$$\alpha\left(w^i\right) = 1.0526\left(w^i\right)^{-0.0036}. \tag{28}$$

or

$$\beta\left(w^i\right) = 0.9500\left(w^i\right)^{0.0036}. \tag{29}$$

Expression (28) allows us to derive $\alpha\left(w^i\right)$ and $\beta\left(w^i\right)$ for a wide range of wages.

To simulate optimal capital income taxes using the estimated form for $\alpha\left(w^i\right)$ in expression (28), we specify a wage $\left(w^i\right)$ distribution, calculate the implied values for $\alpha\left(w^i\right)$, and numerically simulate the planner's problem in (9). We also simulate an augmented planner's problem that limits the planner to no capital income taxation. This enables us to calculate welfare gains from optimal capital taxation.

We use a wage distribution that starts at \$4 and increases in equally-sized discrete bins. Based on Saez (2001), we assume that the distribution of the population across these wages is lognormal up to \$62.50 and Pareto with a parameter value of 2.68 (following Golosov, Troshkin, and Tsyvinski, 2010b) for higher wages. We calibrate the lognormal distribution with the 2007 wage distribution for full-time workers in the United States as reported in the Current Population Survey.

To measure the intertemporal wedge we use the expression:

$$\tau\left(\cdot\right) = 1 - \frac{\frac{u_{c_1^i}\left(\cdot\right)}{u_{c_2^i}\left(\cdot\right)} - 1}{r} \tag{30}$$

where r is the annual rate of return to savings.[24] The variable $\tau\left(\cdot\right)$ measures the relative distortion toward good 1 and away from good 2 at a given income level. Under the capital income tax interpretation, $\tau\left(\cdot\right)$ is

---

[23] We compute wages from the total wage and salary income divided by the total hours worked in 1992, as reported in 1993. We calculate mean wages by AFQT quintile limiting the sample to workers who reported more than 1,000 hours worked. Using all workers does not change the pattern, but all wage levels rise because some workers with low reported hours have high imputed hourly wages.

[24] In the notation of the model of Section 3, $p_2 = \frac{1}{1+r}$, as we assume that savings earn a net rate of return r. We set $1 + r = 1 - \ln \int \beta\left(w^i\right) f\left(w^i\right)$, where we calculate the average instantaneous discount rate, $\rho = -\ln \int \beta\left(w^i\right) f\left(w^i\right)$, and set $r = \rho$. The implicit tax τ is on net capital income, i.e., the implicit after-tax return to saving (marginal rate of transformation) is $\left(1 + r\left(1 - \tau\right)\right)$.

the implicit tax on the interest income earned on good 2, i.e., capital. If this expression is positive, the tax policy is discouraging future consumption relative to current consumption. More informally, it is taxing the return to saving, so we will refer to it as the implied capital income tax.

3.2 Optimal capital income taxes

Figure 1 shows optimal nonlinear capital income tax rates in the baseline case ($\gamma = 1$ and $\sigma = 3$).
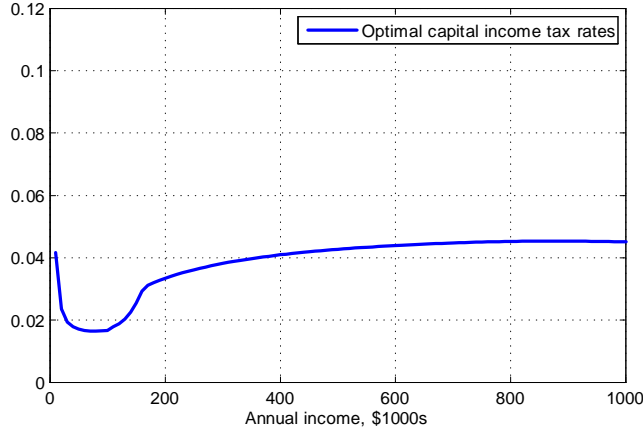

Figure 1: Optimal capital income tax rates in the baseline model

Optimal capital income tax rates are U-shaped (as in Diamond 1998 and Saez 2001). They rise from $100,000 in annual income, corresponding to a wage of $40 per hour, through the point at which the Pareto tail of the wage distribution begins, at an income of around $150,000. Above that income level optimal rates plateau at around 4.5%.

The pattern of optimal rates in Figure 1 can be better understood by examining the components of the analytical result describing optimal distortions from Section 3: expression (19). In Figure 2A and Figure 2B, we show the evolution of $A_2 \left(w^i \right) - A_1 \left(w^i \right)$ and the ratio $\frac{C \left(w^i \right)}{B \left(w^i \right)}$ under the optimal policy over the income distribution, which we split at $300,000 to enable easier examination.
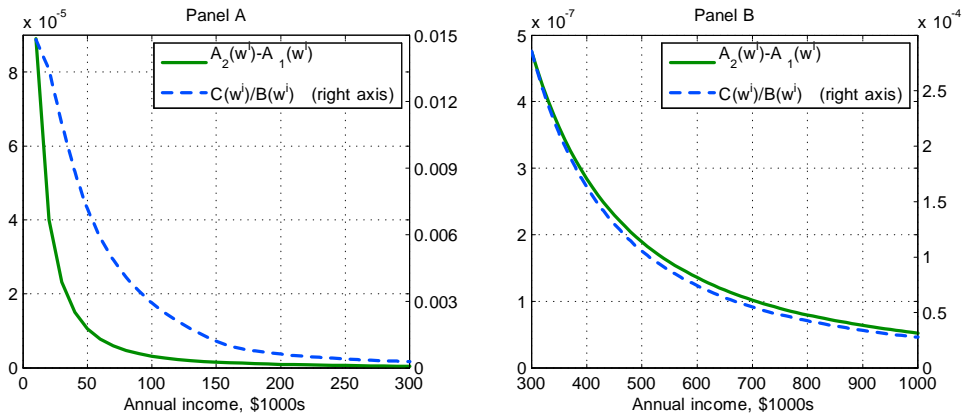

Figure 2: Components of the analytical expression for optimal distortions

These figures show that, as anticipated in Section 3, the difference between the cross-partial derivatives of the marginal utilities of consumption for each good with respect to the wage, $A_2 \left(w^i \right) - A_1 \left(w^i \right)$, falls as

19

wages increase. The cost-benefit ratio of the distortion, represented by $\frac{C(w^i)}{B(w^i)}$ also diminishes with income. Figure 2A shows that the U-shaped pattern of optimal distortions in Figure 1 is due to the rapid fall and then earlier stabilization of the $A_1(w^i) - A_2(w^i)$ term, so that the optimal distortion starts out large, diminishes quickly as the high population density causes the cost-benefit ratio to be relatively larger, and then rebounds as the rate of decline in $\frac{C(w^i)}{B(w^i)}$ exceeds that of $A_2(w^i) - A_1(w^i)$ around \$100,000 of income. Figure 2B shows that these two components decline at a similar rate at higher incomes. This pattern explains why optimal distortions plateau and are essentially constant at high incomes.

The increasing size of the distortions for most of the wage distribution in Figure 1 may seem to contradict the intuition discussed above that distorting savings among lower earners will enable more efficient redistribution from higher earners. However, the equilibrium distortions shown in these figures are not the relevant distortions for an individual claiming an allocation intended for a different type (e.g., type i claiming to be type $i-1$). Such an individual has a lower $\alpha(w^i)$ than the type whose allocation he claims, and this magnifies the effective distortion to his intertemporal optimization if he chooses to mimic the lower type. What matters for individual i's incentives, then, is that the "deviator's distortion" (as defined in Section 1) he faces if he claims to be type $i-1$ is higher than the distortion he faces if he tells the truth. Figure 3 shows the two relevant series: the "deviator's distortion $(i'|i)$" and the truth-telling distortion to type i.
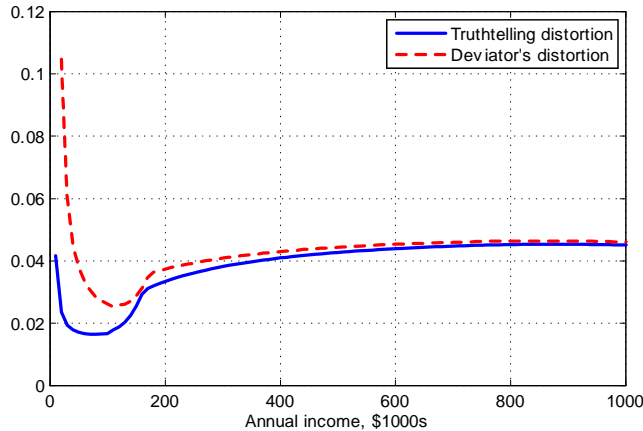

Figure 3: Deviator's distortion in the baseline model

The deviator's distortion always exceeds the truth-telling distortion, consistent with the analytical results above. Optimal nonlinear capital income taxation thereby discourages high-skilled individuals, who value saving, from earning less and claiming a more generous tax treatment.

The welfare gain from optimal capital income taxation given the calibrated $\alpha(w^i)$ is negligible. To measure the welfare gain, we first simulate the optimal policy when capital wedges are constrained to be zero. The planner designs bundles of total consumption and labor income, rather than of consumption in each period and labor income, among which individuals choose. Each individual is then free to allocate his chosen total consumption across periods according to his preferences, with no distortion. This allows us to calculate the factor by which consumption of all agents in both periods would have to be increased in the model without capital taxes to yield the same level of social welfare as in the model with the optimal taxes shown in Figure 1. This factor is 0.00002% of aggregate consumption. The welfare gain is concentrated

20

among low earners.

4 Robustness of baseline results

In this section we extensively examine the robustness of the baseline numerical results. We start by considering social welfare functions other than the Utilitarian function assumed throughout the analysis thus far; this turns out to have little impact on our results. Next, we vary the two key parameters of the utility function: the elasticity of labor supply $\left(\frac{1}{\sigma-1}\right)$ and the intertemporal elasticity of substitution $\left(\frac{1}{\gamma}\right)$. We find that the former matters very little while the latter substantially affects the magnitude of optimal capital income tax rates but has little effect on the welfare gains from optimal policy. Finally, we compare the degree of preference heterogeneity we observe in the data to that needed to justify a range of average capital tax rate levels.[25]

4.1 Alternative social welfare functions

The Utilitarian social welfare function, in which individual types are valued by the social planner according to their proportions of the total population, is a natural choice. As Vickrey (1945) and Harsanyi (1953, 1955) argued, a Utilitarian social welfare function is equivalent to the expected utility function of an individual in an *ex ante* state when he is uncertain over his type. It is also a key benchmark in modern optimal tax studies.

Nevertheless, we may be interested in social welfare functions that are more redistributive than the Utilitarian benchmark. Social welfare functions that are concave in individual utilities are a common variant of the Utilitarian assumption in optimal tax research. Denoting social welfare with W, we can write

$$W = \int_{w_{\min}}^{w_{\max}} \left(U\left(w^i\right) \right)^\rho f\left(w^i\right) dw^i \tag{31}$$

where ρ parameterizes the concavity of social welfare and where $\rho = 1$ for a Utilitarian social welfare function. We consider two more concave versions of expression (31), where $\rho = 0.5$ and $\rho = 0.25$.

The baseline results for optimal capital income taxes turn out to be robust to these different assumptions on social preferences. Figure 4 shows optimal rates for these three social welfare functions under our baseline parameter assumptions of $\gamma = 1$ and $\sigma = 3$.

---

[25] We have also checked the robustness of our results to the number of periods in the model. Numerical simulations that allow for more than two periods, with one consumption good per period, show that optimal distortions are nearly constant across time.
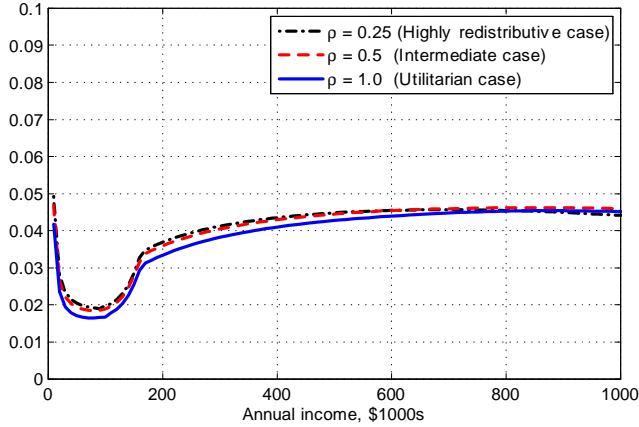
Figure 4: Robustness to varying the social welfare function

The gaps between the optimal rate schedules in Figure 4 are small over the entire income distribution. The rates for high earners plateau at 4.6%, 4.6%, and 4.5%. The differences are slightly larger at lower wage levels, as the planner maximizing a more concave social welfare function uses larger distortions on low earners' consumption choices to enable greater incentive-compatible transfers to them.

4.2 Elasticity of labor supply

The Frisch elasticity of labor supply equals $\frac{1}{\sigma-1}$ in our model. The baseline assumption of $\sigma = 3.0$ implies an elasticity of 0.5, consistent with the evidence in Chetty (2010). Figure 5 shows optimal capital income tax rates for this baseline value and two alternative values: $\sigma = 1.5$ implies an elasticity of 3.0, while $\sigma = 6.0$ implies an elasticity of only 0.2.
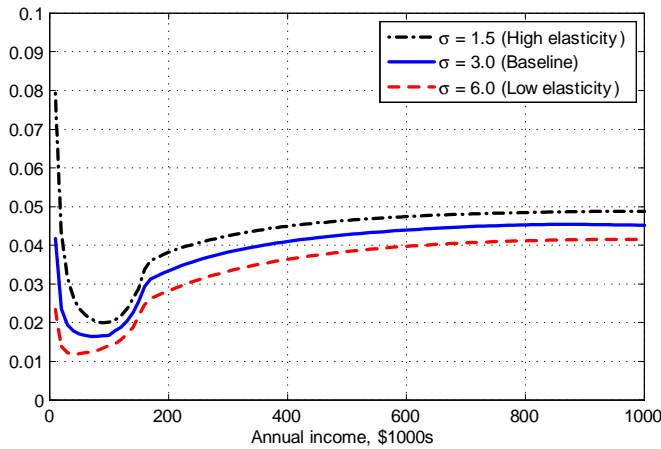
Figure 5: Robustness to varying the elasticity of labor supply

Despite the wide variation in labor supply elasticities covered by Figure 5, there are only minor differences in optimal capital income tax rates. At high income, the optimal rates plateau at similar rates, and there is a steep increase beginning around $100,000 of annual income. The only sizeable difference is for the lowest skilled, who face high rates when the labor supply elasticity is high and low rates when it is low. The

explanation for this pattern lies in the planner's use of intertemporal distortions as a substitute for marginal labor income taxes. When the labor supply elasticity is low, labor income taxes are less distortionary, so the planner does not need to distort the intertemporal margin to provide incentives for the high skilled to exert effort. When the elasticity of labor supply is high, capital income taxes serve a more important role in encouraging work.

4.3   Intertemporal elasticity of substitution

The intertemporal elasticity of substitution equals $\frac{1}{\gamma}$ in our model. The baseline assumption of $\gamma = 1$ is a standard benchmark in mainstream optimal tax and macroeconomic models. But there is substantial debate over the true value of this parameter, so we explore its effects on our baseline results by considering three alternative values: $\gamma = 0.5$, $\gamma = 2$, and $\gamma = 3$. For each case, we compute the implied $\alpha\left(w^i\right)$ following the same procedure described in Section 3.1. Figure 6 shows optimal rates under these different assumptions on γ.
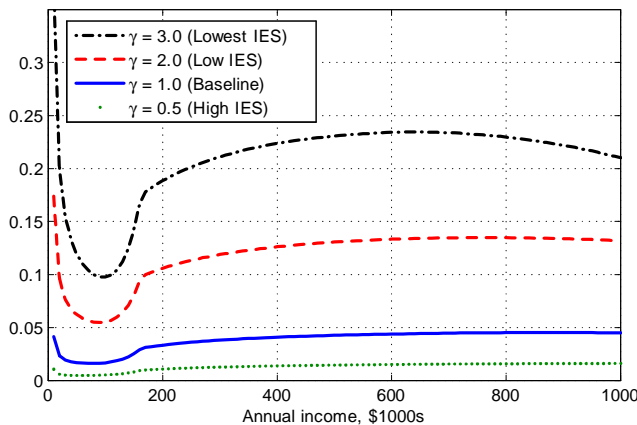

Figure 6: Robustness to varying the intertemporal elasticity of substitution

Figure 6 shows that varying the intertemporal elasticity of substitution has substantial effects on optimal capital income tax rates. For a low intertemporal elasticity ($\gamma = 3$), optimal rates rise to 23.5%, while for a high intertemporal elasticity ($\gamma = 0.5$) they rise to only 1.6%. The baseline case plateaus at 4.5%.

For the planner considering the use of optimal capital taxes, a low intertemporal elasticity of substitution means that individuals' intertemporal allocations will change little in response to distortions. Moreover, the incentive effects of these distortions will be strong, as individuals are eager to avoid allocations that distort them away from their preferred allocations. These factors explain the high optimal capital income tax rates when $\gamma = 3$, and similar reasoning explains the low rates when $\gamma = 0.5$.

Though a low intertemporal elasticity can generate substantially higher optimal tax rates, the welfare gains of moving from no capital taxation to the optimum remain negligible regardless of γ.

Further robustness checks in which we vary the elasticity of labor supply, the intertemporal elasticity of substitution, and the social welfare function together reinforce the lesson that optimal capital income tax rates are substantially larger than in the baseline case only when the intertemporal elasticity of substitution ($\frac{1}{\gamma}$) is small.

4.4 Comparing optimal to existing capital income taxes

Finally, we explore how sensitive our results are to the form of $\alpha\left(w^i\right)$. In particular, we compare our estimate of the empirical relationship between time preferences and ability to that which would be required to justify a given level of capital income taxes. This examines the robustness of our results to the strength of the relationship between preferences and ability.

We calculate the $\alpha\left(w^i\right)$ functions that yield population-weighted average optimal intertemporal wedges corresponding to a range of capital income tax rates. To do so, we continue to model (as in expression 28) the function $\alpha\left(w^i\right)$ as a two-parameter power function

$$\alpha\left(w^i\right) = \psi\left(w^i\right)^\varepsilon, \tag{32}$$

where ψ and ε are scalars. We fix $\alpha\left(w^i\right)$ at its value for $w^i = \$28$ to ensure comparability of these preferences to our empirical estimates. Then, we use the wage $\left(w^i\right)$ distribution and utility function (25) from Section 3 with $\gamma = 1$ and $\sigma = 3$, and we vary the values of ψ and ε in (32) while simulating the planner's problem in (9), (10), and (11).

Leading studies find that tax rates on capital income in developed economies today are over 40% .[26] Figure 7 plots the $\alpha\left(w^i\right)$ required for the population-weighted average optimal intertemporal wedge to imply capital income tax rates of 10%, 20%, and 40% as well as the values for $\alpha\left(w^i\right)$ from our baseline analysis of the NLSY data.
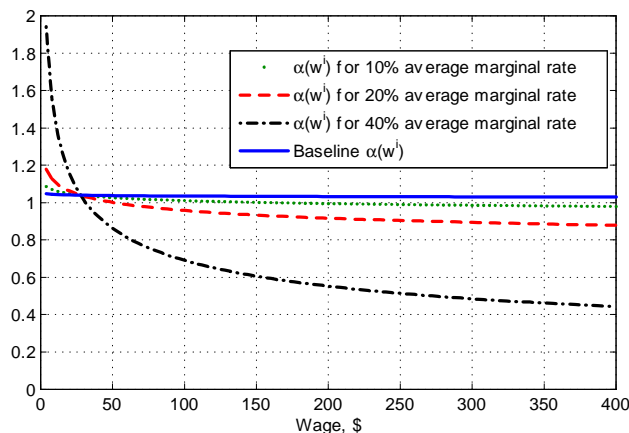
Figure 7: Preferences $\alpha\left(w^i\right)$ required to justify average capital tax rates

To aid intuition, Figure 8 plots the conventional annual discount factor $\beta\left(w^i\right)$ implied by these $\alpha\left(w^i\right)$.

---

[26] The Organization for Economic Cooperation and Development (OECD 2008) reports average combined corporate and personal statutory rates on distributed corporate profits of 42.4 percent in 2007, down from 50 percent in 2000. An alternative measure is the "tax ratio" of capital income tax revenue to total capital income. Carey and Rabesona (2004) calculate the tax ratio for capital income across sixteen OECD countries in 2000 to be 46.3.
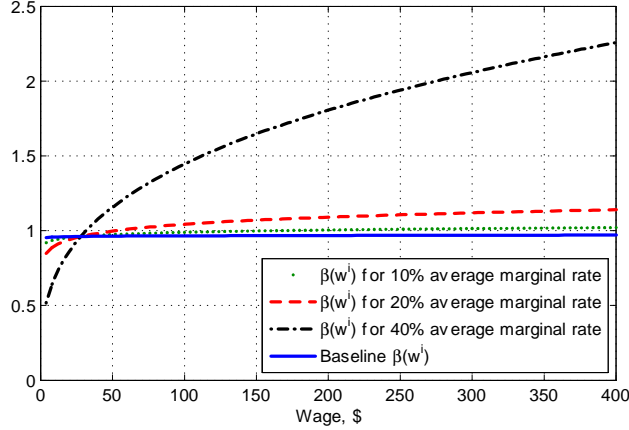
Figure 8: Preferences $\beta\left(w^i\right)$ required to justify average capital tax rates

As these figures make clear, the empirical relationship between time preferences and ability is far weaker than that which would justify the capital income tax rates prevailing in developed economies today, given our baseline calibration with $\gamma = 1$. For example, to justify a 20% capital income tax rate, the discount rate[27] would need to be more than 200% larger for an individual at the twentieth percentile of the ability distribution than for an individual at the eightieth percentile. The NLSY data implies only a 12% gap between these two individuals.

5 Extension: Optimal Capital Taxation when Stochastic Abilities are Related to Preferences

In this section, we extend our analysis of optimal capital taxation when preferences vary with ability to a stochastic setting in a simple environment. The environment below parallels the dynamic Mirrlees model similar to, for example, Golosov, Kocherlakota, and Tsyvinski (2003), Albanesi and Sleet (2006), and Golosov, Tsyvinski and Werning (2006). While different from the model analyzed in previous sections, the model here addresses an additional aspect of how optimal taxes ought to respond to a relationship between preferences and ability.

In period $t = 1$, agents have a common ability level $w = 1$, consume a good x, and produce income y. In period $t = 2$, agents have abilities w^i that take one of two values ($i = \{l, h\}$ for low and high) with probability $\pi^i : \sum_i \pi^i = 1$, consume two goods c_1 and c_2, and produce income y^i. Let $w^h > w^l > 0$. Abilities are private information to the agent. Importantly, agents with the high second-period ability have a relative preference for c_2 over c_1 measured by $\alpha\left(w^i\right)$ where $\alpha'\left(w^i\right) < 0$, just as in the previous sections.

An individual's problem in this setting with no policy is, for $i \in \{l, h\}$:

$$
\max_{x, y, \left\{c_1^i, c_2^i, y^i\right\}_{i=1,2}} \left[u\left(x\right) - v\left(y\right) + \sum_{i=l,h} \pi^i \beta \left[\frac{\alpha\left(w^i\right)}{1 + \alpha\left(w^i\right)} \ln c_1^i + \frac{1}{1 + \alpha\left(w^i\right)} \ln c_2^i - v\left(\frac{y^i}{w_2^i}\right) \right] \right]
$$

---

[27] That is, $\rho\left(w^i\right)$ where $\rho\left(w^i\right) = -\ln\left(\beta\left(w^i\right)\right)$.

subject to the budget constraint

$$(1+r)(y-x) + \left(y^i - c_1^i - c_2^i\right) \geq 0.$$

where $(1+r)$ is the marginal rate of transformation of goods across periods.

The individual allocations satisfy standard stochastic Euler equations:[28]

$$u'(x) = \beta(1+r)\sum \pi^i \frac{\alpha(w^i)}{1+\alpha(w^i)} u'(c_1^i) = \beta(1+r)\sum \pi^i \frac{1}{1+\alpha(w^i)} u'(c_2^i)$$

The social planner's problem is similar to the static problem in Section 2:

Problem 5

$$\max_{x,y,\left\{c_1^i,c_2^i,y^i\right\}_{i=1,2}} \left[u(x) - v(y) + \sum_{i=l,h} \pi^i \beta \left[\frac{\alpha(w^i)}{1+\alpha(w^i)} \ln c_1^i + \frac{1}{1+\alpha(w^i)} \ln c_2^i - v\left(\frac{y^i}{w^i}\right) \right] \right] \qquad (33)$$

subject to the feasibility constraint

$$(1+r)(y-x) + \sum_{i=l,h} \pi^i \left(y^i - c_1^i - c_2^i\right) \geq 0, \qquad (34)$$

and incentive compatibility in period $t=2$:

$$\frac{\alpha(w^h)}{1+\alpha(w^h)} \ln c_1^h + \frac{1}{1+\alpha(w^h)} \ln c_2^h - v\left(\frac{y^h}{w^h}\right) \geq \frac{\alpha(w^h)}{1+\alpha(w^h)} \ln c_1^l + \frac{1}{1+\alpha(w^h)} \ln c_2^l - v\left(\frac{y^l}{w^h}\right), \qquad (35)$$

which says that the high-ability agents do not choose to mimic the low-ability agents in the second period [29]

Let μ denote the multiplier on the incentive compatibility constraint (35).

The first-order conditions of the planner's problem yield a condition describing optimal policy:

$$\beta(1+r)\frac{1}{u'(x)} = \sum_{i=l,h} \pi^i \left(\frac{1}{u'(c_1^i)} + \frac{1}{u'(c_2^i)} \right). \qquad (36)$$

Result (36) is the Inverse Euler Equation, and it is the same condition that describes optimal policy when preferences do not vary across types. Thus, the planner's optimal policy toward saving is unaffected by the relationship between stochastic ability and the preference for goods when the planner is able to use nonlinear, income-dependent commodity taxes.

How does the optimal policy treat consumption choices in the second period? Define θ^i as the distortion

---

[28] An equivalent expression combining the second-period commodities could be used, instead.

[29] As in the previous Sections, we assume that only the high ability agent can pretend to be the low ability agent and not vice versa.

to second-period consumption choices for an individual of second-period type i:

$$\theta^i = \frac{\frac{\alpha(w^i)}{1+\alpha(w^i)} u'\left(c_1^i\right)}{\frac{1}{1+\alpha(w^i)} u'\left(c_2^i\right)},$$

In the Appendix, we show that the optimal policy is the same as that from our analysis of optimal capital taxation in previous sections. In particular, the high type faces no distortion to its commodity choices ($\theta^h = 1$), while the low type is distorted away from the good that the high-type relatively prefers ($\theta^l > 1$).

6   Conclusion

Among others, Mirrlees (1976) and Saez (2002) have argued that goods preferred by the high-ability ought to be taxed as part of an optimal tax policy that seeks to redistribute toward the individuals with (unobservable) low ability. Recently, the logic for taxing goods preferred by those with high ability has been used to argue for positive capital income taxation, for example by Banks and Diamond (2008).

We study the case for nonlinear taxes on goods justified by a relationship between ability and preferences over them. We derive an analytical result characterizing the optimal distortion to consumption choices and decompose it into both conventional and novel factors. Then, we calibrate a model economy to micro data that allows us to estimate savings preferences and their relationship to measured ability. When we simulate optimal policy given these estimates, we find that the magnitude of optimal capital income taxes is modest–only 2% on average and 4.5% on high earner–for our baseline case with a unit intertemporal elasticity of substitution. The welfare gains from these taxes are small. These results are robust to variation in the social welfare function's concavity and the elasticity of labor supply. Substantially larger optimal capital income tax rates are implied if the intertemporal elasticity of substitution is lower, though even in that case the welfare gain from imposing optimal capital income taxes remains small.

7 Appendix

7.1 The Derivation of the General Tax Ratio, expression (19)

The Hamiltonian from the main text, (16), includes the following differential constraint:

$$\frac{\partial U^i}{\partial w^i} = u_{w^i}\left(c_1^i, c_2^i, l^i, w^i\right) + \mu\left(l^i\left(1 - T'\left(w^i l^i\right) - t_{y^i}^1\left(w^i l^i, c_1^i\right) - p_2 t_{y^i}^2\left(w^i l^i, c_2^i\right)\right)\right), \tag{37}$$

derived using the envelope condition on the individual's utility maximization problem. Using the individual's first order condition with respect to labor l^i, we can rewrite the Hamiltonian as:

$$H\left(w^i\right) = \left(U\left(w^i\right) + \lambda\left(w^i l^i - c_1^i - p_2 c_2^i\right)\right) f\left(w^i\right) + \phi\left(u_{w^i}(\cdot) - \frac{l^i u_{l^i}(\cdot)}{w^i}\right),$$

where subscripts denote partial derivatives and $(\cdot)$ denotes the set of arguments of the utility function, $\left(c_1^i, c_2^i, l^i, w^i\right)$.

To solve for the optimal policy, choose l and c_1^i as the control variables. The first order conditions yield

$$\frac{dc_2^i}{dc_1^i} = \frac{-\left(\lambda f\left(w^i\right) - \phi\left(u_{w^i c_1^i}(\cdot) - \frac{l^i u_{l^i c_1^i}(\cdot)}{w^i}\right)\right)}{\left(\lambda p_2 f\left(w^i\right) - \phi\left(u_{w^i c_2^i}(\cdot) - \frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\right)\right)},$$

and

$$\frac{dc_2^i}{dl^i} = \frac{\lambda w^i f\left(w^i\right) + \phi\left(u_{w^i l^i}(\cdot) - \frac{u_{l^i}(\cdot)}{w^i} - \frac{l^i u_{l^i l^i}(\cdot)}{w^i}\right)}{\lambda p_2 f\left(w^i\right) - \phi\left(u_{w^i c_2^i}(\cdot) - \frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\right)}.$$

Individuals will allocate their after-tax income so that the following relationships hold:

$$\frac{dc_2^i}{dc_1^i} = -\frac{u_{c_1^i}}{u_{c_2^i}} = -\frac{1 + t_{c_1}^1\left(w^i l^i, c_1^i\right)}{p_2\left(1 + t_{c_2}^2\left(w^i l^i, c_2^i\right)\right)} \tag{38}$$

$$\frac{dc_2^i}{dl^i} = -\frac{u_{l^i}}{u_{c_2^i}} = \frac{w^i\left(1 - T'\left(w^i l^i\right) - \frac{\partial t^1\left(w^i l^i, c_1^i\right)}{\partial y^i} - p_2\frac{\partial t^2\left(w^i l^i, c_2^i\right)}{\partial y^i}\right)}{p_2\left(1 + t_{c_2}^2\left(w^i l^i, c_2^i\right)\right)}$$

so we can write:

$$\frac{1}{p_2}\frac{1 + t_{c_1}^1\left(w^i l^i, c_1^i\right)}{1 + t_{c_2}^2\left(w^i l^i, c_2^i\right)} = \frac{1}{p_2}\frac{\lambda f\left(w^i\right) - \phi\left(w^i\right)\left(u_{w^i c_1^i}(\cdot) - \frac{l^i u_{l^i c_1^i}(\cdot)}{w^i}\right)}{\lambda f\left(w^i\right) - \phi\left(w^i\right)\frac{1}{p_2}\left(u_{w^i c_2^i}(\cdot) - \frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\right)}. \tag{39}$$

The expression (39) includes multipliers from the planner's problem. Next, we derive expressions for them in terms of the model variables. This yields the optimal tax result (19). To do so, we write the Hamiltonian in terms of only the control and state variables. The individual's budget constraint implicitly defines c_2 as a function of the variables $\left\{l^i, w^i, c_1^i\right\}$ as well as taxes, which themselves depend on these three

variables. Therefore we can write:

$$c_2^i = \Phi\left(c_1^i, l^i, w^i\right),$$

and

$$\frac{\partial U^i}{\partial w^i} = \left(u_{w^i}\left(c_1^i, l^i, w^i\right) - \frac{l^i u_{l^i}\left(c_1^i, l^i, w^i\right)}{w^i}\right).$$

Next, we use expression (8), the individual's utility function, from the main text:

$$U\left(w^i\right) = u\left(c_1^i, c_2^i, l^i, w^i\right)$$

to write the following implicit expression for c_2^i:

$$c_2^i = \psi\left(U^i, c_1^i, l^i, w^i\right).$$

With these substitutions, write the Hamiltonian as

$$H\left(w^i\right) = \left(U\left(w^i\right) + \lambda\left(w^i l^i - c_1^i - p_2\psi\left(U^i, c_1^i, l^i, w^i\right)\right)\right) f\left(w^i\right) + \phi\left(u_{w^i}\left(c_1^i, l^i, w^i\right) - \frac{l^i u_{l^i}\left(c_1^i, l^i, w^i\right)}{w^i}\right)$$

Pontryagin's Maximum Principle implies

$$\phi'\left(w^i\right) = -\left[f\left(w^i\right) - \lambda p_2 f\left(w^i\right)\psi_{U^i}\left(U^i, c_1^i, l^i, w^i\right)\right].$$

The transversality conditions yield:

$$\lambda = \frac{1}{\int_{w^j=w_{\min}}^{w^j=w_{\max}} p_2\psi_{U^j}\left(U^j, c_1^j, l^j, w^j\right) f\left(w^j\right) dw^j},$$

$$\phi\left(w^i\right) = \left(1 - F\left(w^i\right)\right)\left(1 - \frac{\frac{1}{(1-F(w^i))}\int_{w^j=w^i}^{w^j=w_{\max}}\psi_{U^j}\left(U^j, c_1^j, l^j, w^j\right) f\left(w^j\right) dw^j}{\frac{1}{(1-F(0))}\int_{w^j=w_{\min}}^{w^j=w_{\max}}\psi_{U^j}\left(U^j, c_1^j, l^j, w^j\right) f\left(w^j\right) dw^j}\right).$$

Use

$$\psi_{U^j}\left(U^j, c_1^j, l^j, w^j\right) = \frac{\partial c_2^j}{\partial U^j} = \frac{1}{u_{c_2^j}},$$

to derive the expressions (23) and (24) from the main text.

29

7.1.1 Expression for optimal marginal income taxes

For income taxes, we are interested in the extra tax an individual pays when he earns a dollar of income. This will include commodity taxes. We start by combining results from the previous section to obtain:

$$\frac{dc_2^i}{dy^i} = \frac{\left(1 - T'\left(w^i l^i\right) - \frac{\partial t^1\left(w^i l^i, c_1^i\right)}{\partial y^i} - p_2 \frac{\partial t^2\left(w^i l^i, c_2^i\right)}{\partial y^i}\right)}{p_2\left(1 + t_{c_2}^2\left(w^i l^i, c_2^i\right)\right)} = \frac{\lambda f\left(w^i\right) + \phi\left(u_{w^i y^i}\left(\cdot\right) - \frac{u_{y^i}\left(\cdot\right)}{w^i} - l^i u_{y^i y^i}\left(\cdot\right)\right)}{\lambda p_2 f\left(w^i\right) - \phi\left(u_{w^i c_2^i}\left(\cdot\right) - l^i u_{y^i c_2^i}\left(\cdot\right)\right)}.$$

Denote the labor wedge relative to good 2 as τ_{l,c_2}. Then, using the results (23) and (24), we can write:

$$\left(1 - \tau_{l,c_2}\right) = \frac{f\left(w^i\right) - p_2\left(u_{w^i y^i}\left(\cdot\right) - \frac{u_{y^i}\left(\cdot\right)}{w^i} - l^i u_{y^i y^i}\left(\cdot\right)\right)\left(\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j - \frac{\left(1-F\left(w^i\right)\right)}{\left(1-F\left(w^0\right)\right)} \int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j\right)}{p_2\left(f\left(w^i\right) + \left(u_{w^i c_2^i}\left(\cdot\right) - l^i u_{y^i c_2^i}\left(\cdot\right)\right)\left(\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j - \frac{\left(1-F\left(w^i\right)\right)}{\left(1-F\left(w^0\right)\right)} \int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j\right)\right)}.$$

Using expressions (20), (21), and (22), this simplifies to:

$$\left(1 - \tau_{l,c_2}\right) = \frac{-\left(u_{w^i y^i}\left(\cdot\right) - \frac{u_{y^i}\left(\cdot\right)}{w^i} - l^i u_{y^i y^i}\left(\cdot\right)\right) + \frac{C\left(w^i\right)}{B\left(w^i\right)}}{A_2\left(w^i\right) + p_2 \frac{C\left(w^i\right)}{B\left(w^i\right)}}. \tag{40}$$

Note that if $A_2\left(w^i\right) > 0$, as we assumed throughout the analysis, $\left(1 - \tau_{l,c_2}\right)$ is smaller than if $A_2\left(w^i\right) = 0$. Applying (19) to (40) yields the parallel result for the labor wedge relative to good 1 (τ_{l,c_2}):

$$\left(1 - \tau_{l,c_1}\right) = \frac{-\left(u_{w^i y^i}\left(\cdot\right) - \frac{u_{y^i}\left(\cdot\right)}{w^i} - l^i u_{y^i y^i}\left(\cdot\right)\right) + \frac{C\left(w^i\right)}{B\left(w^i\right)}}{A_1\left(w^i\right) + \frac{C\left(w^i\right)}{B\left(w^i\right)}}. \tag{41}$$

Here, if $A_1\left(w^i\right) < 0$, the labor wedge is greater than if there is no relationship between preferences and ability.

7.2 Estimating time preference from NLSY data

Our measure of preferences will be the discount factor implied by using NLSY data on income and net worth in a simple model of individual optimization . Suppose individuals live for three periods. In the first two periods, roughly corresponding to ages 20 through 42 and 43 through 65, they work, consume, and borrow or save. In the third period, they are retired and live for 23 years (for simplicity, as this makes all three periods of similar length). The individual solves the following utility maximization problem:

$$\max_{c_1,c_2,c_3} \left[\frac{\left(c_1\right)^{1-\gamma}-1}{1-\gamma} + \delta \frac{\left(c_2\right)^{1-\gamma}-1}{1-\gamma} + \delta^2 \frac{\left(c_3\right)^{1-\gamma}-1}{1-\gamma} - v\left(y_1, y_2\right)\right]$$

subject to

$$\left(\left(y_1 - c_1\right) R^2 + \left(y_2 - c_2\right)\right) R - c_3 = 0.$$

where c_t and y_t are consumption and income in period t, δ is the discount factor across 23-year periods (i.e., if the one-year-ahead discount factor is β, then $\delta = \beta^{23}$), $R = (1.05)^{23}$ is the average return to saving over a 23-year period, and $v(\cdot)$ is an unspecified function for the disutility of earning income.

We make the assumption that an individual's total value of income prior to age 43 is identical to the income it will earn from age 43 until retirement. In the notation of the model, we assume $y_1 = y_2$ for all individuals. Solving the individual's problem yields:

$$\left((\delta)^{\frac{1}{\gamma}} \right)^2 (R)^{\frac{2-2\gamma}{\gamma}} + (\delta)^{\frac{1}{\gamma}} (R)^{\frac{1-\gamma}{\gamma}} + \left(1 - \frac{y_1 R + y_2}{R c_1} \right) = 0.$$

Assuming $y_1 = y_2$,

$$\delta = \left(\frac{1}{2R^{\frac{1-\gamma}{\gamma}}} \left(\left(-3 + 4 \frac{y_1}{c_1} \frac{1+R}{R} \right)^{\frac{1}{2}} - 1 \right) \right)^{\gamma}.$$

If $\gamma = 1$, so if the utility from consumption is logarithmic, this simplifies to:

$$\delta = \frac{1}{2} \left(\left(-3 + 4 \frac{y_1}{c_1} \frac{1+R}{R} \right)^{\frac{1}{2}} - 1 \right)$$

As expected, the higher is income relative to consumption, the greater the estimated δ for an individual. We drop 37 individuals whose estimated δ is negative or exceeds two in the $\gamma = 1$ specification, leaving 7,008 observations.

To estimate δ, we need values for y_1 and c_1 for each individual. For y_1, we use the NLSY's observations on income over time for each individual to calculate the "future value" of income earned prior to and including 2004. We do not observe income in all years for each individual. To obtain an income figure comparable to ending net worth for each individual, we calculate the future value of the observed incomes for each individual. Then, we scale that future value by the maximum number of years observable over the number of years observed for each individual. We also do not observe initial net worth. However, if we control for net worth in 1985, just six years after the survey began, the coefficient on AFQT is hardly changed.

Formally, $y_1 = \sum_{t=1979}^{2004} R^{\frac{1}{23}(2004-t)} y_t$. Using the full time series of income rather than simply the most recent observation of income is important for two reasons. First, it gives a better measure of the individual's likely lifetime or permanent income. Second, to calculate c_1, we assume that any income not accumulated as net worth by 2004 was consumed. Formally, we denote the NLSY variable "family net worth" NW and calculate $c_1 = y_1 - NW$.

Our data do not include components of individuals' expected future income, such as Social Security payments or other social transfers. To the extent that these omissions bias down the estimate of net worth, we will understate saving rates. Therefore, if these transfers are progressive, we will be overestimating the slope of discount factors versus ability. In a similar way, expected future gifts and inheritances are not taken into account in the data. To the extent that these are increasing in recipient income, we are underestimating the slope of discount factors versus ability.

Finally, a note on converting the estimates of δ into the preferences in expression (25). The following

31

equality relates the estimated $\delta\left(w^i\right)$ of individual i between the two periods t and $t+1$ to its annualized level, $\hat{\beta}\left(w^i\right)$:

$$\left(\delta\left(w^i\right)\right)^{\frac{1}{23}} = \hat{\beta}\left(w^i\right).$$

Next, $\hat{\beta}\left(w^i\right)$ is related to the model's representation of preferences, denoted $\hat{\alpha}\left(w^i\right)$, by the following expression:

$$\hat{\beta}\left(w^i\right) = \left(p_2\right)^{1-\gamma}\left(\frac{1}{\hat{\alpha}\left(w^i\right)}\right)^{\gamma},$$

Simplifying, note that the price ratio is the inverse of the annual return to saving, so $p_2 = \frac{1}{R^{\frac{1}{23}}}$ and:

$$\hat{\beta}\left(w^i\right) = \left(\frac{1}{R^{\frac{1}{23}}}\right)^{1-\gamma}\frac{1}{\left(\hat{\alpha}\left(w^i\right)\right)^{\gamma}};$$

$$\hat{\alpha}\left(w^i\right) = \left(\frac{1}{R^{\frac{1}{23}}}\right)^{\frac{1-\gamma}{\gamma}}\left(\frac{1}{\hat{\beta}\left(w^i\right)}\right)^{\frac{1}{\gamma}}. \tag{42}$$

We calculate R as described in the footnote to expression (30), using the estimated $\hat{\beta}\left(w^i\right)$. Specifically, we calculate $R = 1 + r = 1 - \ln\left(\sum_i \hat{\beta}\left(w^i\right)\pi^i\right)$, where π^i is the population proportion of type i. This reflects that the net rate of return r is set equal to the average discount rate $\rho = -\ln\left(\sum_i \hat{\beta}\left(w^i\right)\pi^i\right)$ in the data.

7.2.1 Utility function normalization, expressions (25) and (26)

Here, we detail the normalization of preferences in the expression (25). The goal is to scale the preferences across goods so that they do not mechanically affect labor effort. For an example of such an effect, consider the case of an individual solving an intertemporal optimization problem with utility

$$U = \sum_t^T\left(\left(\delta^i\right)^{t-1}\frac{\left(c_t^i\right)^{(1-\gamma)}-1}{(1-\gamma)} - \frac{1}{\sigma}\left(\frac{y_t^i}{w_t^i}\right)^{\sigma}\right)$$

and budget constraint

$$\sum_t^{T-1} p_t\left(y_t^i - c_t^i\right) - p_T c_T^i = 0,$$

with multiplier λ. Substituting the individual's first order conditions into the budget constraint yields the following value for λ :

$$\lambda = \left(\frac{\sum_t^{T-1} p_t\left(\frac{p_t}{\left(\delta^i\right)^{t-1}}\right)^{\frac{-1}{\gamma}} + p_T\left(\frac{p_T}{\left(\delta^i\right)^{T-1}}\right)^{\frac{-1}{\gamma}}}{\sum_t^{T-1} p_t\left(w_t^i\right)^{\frac{\sigma}{\sigma-1}}\left(p_t\right)^{\frac{1}{\sigma-1}}}\right)^{\frac{(\sigma-1)\gamma}{\sigma-1+\gamma}}.$$

32

This expression for λ implies:

$$y_t^i = \left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{1}{\sigma-1}} \left(\frac{\sum\limits_t^T (p_t)^{\frac{\gamma-1}{\gamma}} \left(\left(\delta^i\right)^{t-1}\right)^{\frac{1}{\gamma}}}{\sum\limits_t^{T-1} (p_t)^{\frac{\sigma}{\sigma-1}} \left(w_t^i\right)^{\frac{\sigma}{\sigma-1}}} \right)^{\frac{\gamma}{\sigma-1+\gamma}}.$$

In this case, the chosen income level y_t^i depends on preferences.

We wish to avoid that dependence, so we specify preferences in a way that will cause each individual's labor effort to be independent of time preferences. Consider a normalized version of the previous utility function:

$$U = \sum_{t=1}^T \left(\left(\frac{\alpha \left(w^i\right)^{T-t}}{\sum\limits_{s=1}^T \left(\alpha\left(w^i\right)\right)^{s-1}} \right)^\gamma (p_t)^{1-\gamma} \frac{\left(c_t^i\right)^{(1-\gamma)} - 1}{(1-\gamma)} - \frac{1}{\sigma} \left(\frac{y_t^i}{w_t^i}\right)^\sigma \right).$$

This is a generalization of (25) to T rather than 2 time periods. Substituting the individual's first order conditions into the budget constraint yields the following value for λ:

$$\lambda = \left(\frac{1 + \sum\limits_t^{T-1} \alpha \left(w^i\right)^{T-t}}{\sum\limits_t^{T-1} \left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{\sigma}{\sigma-1}} \sum\limits_{s=1}^T \left(\alpha\left(w^i\right)\right)^{s-1}} \right)^{\frac{(\sigma-1)\gamma}{\sigma-1+\gamma}}.$$

This expression for λ implies:

$$y_t^i = \frac{\left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{1}{\sigma-1}}}{\left(\sum\limits_t^{T-1} \left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma}{\sigma-1+\gamma}}} \left(\frac{1 + \sum\limits_t^{T-1} \alpha\left(w^i\right)^{T-t}}{\sum\limits_{s=1}^T \left(\alpha\left(w^i\right)\right)^{s-1}} \right)^{\frac{\gamma}{\sigma-1+\gamma}}.$$

To simplify, note that

$$\sum_{s=1}^T \left(\alpha\left(w^i\right)\right)^{s-1} = 1 + \sum_{t=1}^{T-1} \alpha\left(w^i\right)^{T-t}$$

so that the expression for y_t^i simplifies to:

$$y_t^i = \frac{\left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{1}{\sigma-1}}}{\left(\sum\limits_t^{T-1} \left(w_t^i\right)^{\frac{\sigma}{\sigma-1}} (p_t)^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\gamma}{\sigma-1+\gamma}}}.$$

With this normalization, the choice of effort does not depend on preferences.

Note that if $\gamma = 1$, the normalized utility function becomes

$$U = \sum_{t=1}^T \left(\frac{\alpha \left(w^i\right)^{T-t}}{\sum\limits_{s=1}^T \left(\alpha\left(w^i\right)\right)^{s-1}} \ln c_t^i - \frac{1}{\sigma} \left(\frac{y_t^i}{w_t^i}\right)^\sigma \right).$$

These normalized utility functions are used in the main paper.

7.3 Dynamic Model, result (36)

The planner's problem is stated in the main text. The first-order conditions are:

$$u'(x) = R\lambda,$$

$$\frac{1}{1 + \alpha(w^h)} u'(c_1^h)(\pi^h \beta + \mu) = \pi^h \lambda,$$

$$\frac{\alpha(w^h)}{1 + \alpha(w^h)} u'(c_2^h)(\pi^h \beta + \mu) = \pi^h \lambda,$$

$$\frac{1}{1 + \alpha(w^l)} u'(c_1^l)\left(\pi^l \beta - \frac{1 + \alpha(w^l)}{1 + \alpha(w^h)}\mu\right) = \pi^l \lambda,$$

$$\frac{\alpha(w^l)}{1 + \alpha(w^l)} u'(c_2^l)\left(\pi^l \beta - \frac{\alpha(w^h)}{\alpha(w^l)}\frac{1 + \alpha(w^l)}{1 + \alpha(w^h)}\mu\right) = \pi^l \lambda.$$

Together, they yield

$$\left(\frac{\pi^h}{1 + \alpha(w^h)} + \frac{\pi^l}{1 + \alpha(w^l)}\right)\beta = \left(\frac{\pi^h}{u'(c_1^h)} + \frac{\pi^l}{u'(c_1^l)}\right)\lambda,$$

$$\left(\frac{\alpha(w^h)}{1 + \alpha(w^h)}\pi^h + \frac{\alpha(w^l)}{1 + \alpha(w^l)}\pi^l\right)\beta = \left(\frac{\pi^h}{u'(c_2^h)} + \frac{\pi^l}{u'(c_2^l)}\right)\lambda,$$

and thus the Inverse Euler Equation:

$$\beta R \frac{1}{u'(x)} = \sum_{i=l,h} \pi^i \left(\frac{1}{u'(c_1^i)} + \frac{1}{u'(c_2^i)}\right).$$

These first order equations also imply no distortion on the consumption choices in period 2 of the high type but a distortion to the choices of the low type, just as in the simple two-type model in the main text.

References

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

[2] Ales, Laurence, and Pricila Maziero (2009). "Accounting for Private Information," Working Paper.

[3] Atkinson, Anthony and Joseph Stiglitz (1976). "The Design of Tax Structure: Direct vs. Indirect Taxation," *Journal of Public Economics*.

[4] Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2006). "Who is 'Behavioral'? Cognitive ability and anomalous preferences." Harvard University mimeo, May.

[5] Banks, James and Peter Diamond (2008). "The Base for Direct Taxation," *Mirrlees Review*, Institute for Fiscal Studies.

[6] Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). "Consumption Inequality and Partial Insurance," *American Economic Review 98(5).*

[7] Cagetti, Marco (2003). "Wealth Accumulation over the Life Cycle and Precautionary Saving," *Journal of Business and Economic Statistics*, 21(3).

[8] Chetty, Raj (2010). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," NBER Working Paper 15616 version December 2009. Newer version available from author's website.

[9] Corlett, W.J. an D.C. Hague (1953). "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies* (21), pp. 21-30.

[10] Diamond, Peter. Discussion of "New Dynamic Public Finance: A User's Guide," in the *NBER Macroannual 2006*, pp. 365-379.

[11] Diamond, Peter and Johannes Spinnewijn (2009) "Capital Income Taxes with Heterogeneous Discount Rates," MIT working paper, May 19.

[12] Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes (2004). "Do the Rich Save More?," *Journal of Political Economy.*

[13] Dynan, Karen E. (1993)."The Rate of Time Preference and Shocks to Wealth: Evidence from Panel Data." Board of Governors of the Federal Reserve System, Economic Activity Section WP 134.

[14] Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski (2003). "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3).

[15] Golosov, Mikhail and Aleh Tsyvinski (2006). "Designing Optimal Disability Insurance: A Case For Asset Testing," *Journal of Political Economy.*

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

[17] Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2010a). "Optimal Taxation: Merging Micro and Macro Approaches," forthcoming *Journal of Money, Credit, and Banking.*

[18] Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2010b). "Optimal Dynamic Taxes," Working Paper.

[19] Gordon, Roger H., and Wojciech Kopczuk (2010). "The Choice of the Personal Income Tax Base," Working paper.

[20] Hubbard, R. Glenn, Jonthan Skinner, and Stephen P. Zeldes (1998). "Precautionary Saving and Social Insurance," *Journal of Political Economy, 103.*

[21] Kleven, Henrik, Claus Kreiner and Emmanuel Saez (2009). "The Optimal Income Taxation of Couples", *Econometrica* 77(2), 2009, 537-560.

[22] Kaplow, Louis (2008). *The Theory of Taxation and Public Economics.*

[23] Kocherlakota, Narayana (2009). *The New Dynamic Public Finance.* Princeton University Press.

[24] Kocherlakota, Narayana and Luigi Pistaferri (2009). "Asset Pricing Implications of Pareto Optimality with Private Information," *Journal of Political Economy 117(3).*

[25] Kocherlakota, Narayana and Luigi Pistaferri (2007). "Household Heterogeneity and Real Exchange Rates," *Economic Journal 117.*

[26] Judd, Kenneth and Che-Lin Su (2006). "Optimal Income Taxation with Multidimensional Taxpayer Types," Working Paper, earliest available draft April, 2006; more recent draft November 22, 2006.

[27] Meghir, C and Luigi Pistaferri (forthcoming). "Earnings, consumption, and life cycle choices," *Handbook of Labor Economics*, vol. 4.

[28] Mirrlees, James (1971). "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies.*

[29] Mirrlees, James (1976). "Optimal Tax Theory: A Synthesis," *Journal of Public Economics.*

[30] Mirrlees, James (1986). "The Theory of Optimal Taxation," *Handbook of Mathematical Economics*, eds. K.J. Arrow and M.D. Intrilligator: Elsevier.

[31] OECD (2008). "Table II.4: Overall statutory tax rates on dividend income," downloads from the OECD Tax Database: www.oecd.org/ctp/taxdatabase.

[32] Salanie, Bernard (2003). *The Theory of Taxation.*

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

[34] Saez, Emmanuel (2002). "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics.*

[35] Tarkiainen, Ritva and Matti Tuomala (2007). "On optimal income taxation with heterogeneous work preferences", *International Journal of Economic Theory,* 3(1).

[36] Tuomala, Matti (1991). *Optimal Income Tax and Redistribution.*

[37] Weinzierl, Matthew (forthcoming). "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies.*