Preference Heterogeneity and Optimal Commodity Taxation

Mikhail Golosov, Aleh Tsyvinski, and Matthew Weinzierl*

October 14, 2009

Abstract

We analytically and quantitatively examine a prominent justification for capital taxation: goods preferred by the high-skilled ought to be taxed. We study an environment where commodity taxes are allowed to be nonlinear functions of income and consumption and find that optimal commodity taxes on these goods may be regressive. We first derive an expression for optimal commodity taxation, allowing us to study the forces for and against regressivity in that more general setting. We then parameterize the model to evidence on the relationship between skills and preferences and examine the quantitative case for regressive taxes on two specific, important categories of expenditure: future consumption (saving) and housing. The relationship between skill and time preference delivers quantitatively small, regressive capital taxes and does not justify substantial capital taxation, whether regressive or linear. In contrast to the case of capital taxation, there is a stronger case for regressive treatment of spending on owner-occupied housing. Our quantitative simulations yield an optimal policy that resembles the current mortgage interest deduction in the United States.

Introduction

One justification for positive capital taxation is that the goods preferred by high-ability individuals ought to be taxed because the consumption of these goods provides a signal of individuals' otherwise unobservable ability.[1,2] If individuals' abilities are positively related to preferences for saving, this argument implies that capital income should be taxed. Two prominent expositions of this justification are Saez (2002) and Banks and Diamond (2009). Saez shows that a small linear tax on a commodity preferred by individuals with higher skills 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 taxation and concludes "...the discount rate δ is probably negatively correlated with skills. This suggests that interest income ought

*Golosov: Yale and NES, Tsyvinski: Yale and NES, Weinzierl: Harvard Business School. Golosov and Tsyvinski thank NSF for support. We thank V.V. Chari, Larry Jones, and Greg Mankiw for comments.

[1] A different justification for the positive capital wedge is the New Dynamic Public Finance literature (see e.g., Golosov, Kocherlakota, and Tsyvinski 2003; Golosov, Tsyvinski and Werning 2006; Kocherlakota 2009).

[2] This result originated in Mirrlees (1976). Nearly all comprehensive treatments of modern tax policy contain a section on this result as a deviation from the standard Atkinson-Stiglitz (1976) recommendation of uniform commodity taxation. 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." Enthusiasm for this result may be because, as Mirrlees put it "This prescription is most agreeable to common sense."

to be taxed even in the presence of a non-linear optimal earnings tax." Banks and Diamond (2009) is the chapter on direct taxation in the Mirrlees Review. Commissioned by the Institute for Fiscal Studies, the 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 tradeoff by being a source of indirect evidence about who has higher earnings abilities and thus contributes to more efficient redistributive taxation."

In this paper, we analytically and quantitatively study this justification for taxing goods preferred by those with high ability, in particular future consumption (i.e., saving) and housing, when commodity taxes are allowed to be nonlinear functions of both income and consumption.[3]

We first derive analytical expressions that indicate 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 positive marginal tax on the good preferred by the high type. In other words, taxes are regressive in this case. We then derive the condition describing optimal commodity taxes in an economy with a continuum of types. The commodity tax on the agent with the highest skill is again equal to zero and is positive for other types. As is common in Mirrleesian models (e.g., Saez 2001) we then analytically study the forces for and against regressivity. The intuition for why regressive commodity taxation may be optimal starts with the realization that the goal of optimal tax policy (in the Mirrleesian framework) is to redistribute from high-ability workers without discouraging their work effort. With this as the goal, the optimal use of commodity taxation is to increase the attractiveness of earning a high income. Commodity taxes that are regressive (i.e., that fall with income) on those goods most 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 the low-skilled.[4]

The second result of the paper is to examine the quantitative case for regressive taxes in two specific important categories of expenditure, saving and housing. Our main conclusion here is twofold. The optimal nonlinear (and also linear) capital tax is small. Thus, our results do not justify substantial capital taxation despite the positive relationship between ability and patience that we find in the data. At the same time, the optimal regressive taxation of housing is quantitatively significant and, for a baseline calibration, is consistent with the magnitude and shape of the empirical estimates of the implicit tax distortion provided by the mortgage interest deduction (see Poterba and Sinai 2009).

To characterize the optimal taxation of saving, we use existing evidence from Lawrance (1991) as well as data from the National Longitudinal Survey of Youth (NLSY) to show a positive correlation between ability[5] and relative preference for future consumption. Using these data to estimate a mean value for time preference

[3] Though most research on this issue has focused on the linear tax problem, Mirrlees (1976) is clear that his results apply to *nonlinear* marginal commodity tax rates. A few later authors also noted the potential for optimal nonlinear rates: e.g., Kaplow (2008). Banks and Diamond (2009) look for but find no work on the nonlinear problem. They write: "In the context of this issue, how large the tax on capital income should be and how the marginal capital income tax rates should vary with earnings levels has not been explored in the literature that has been examined."

[4] The standard argument against nonlinear commodity taxation is arbitrage or retrading (see Hammond 1987, Golosov and Tsyvinski 2006). That may be an appropriate restriction for many goods, but important categories of personal expenditure can feasibly be taxed nonlinearly or as a function of income.

[5] In our analysis of Lawrance's data, we use the mean reported wage by income group as the proxy for ability. When we use NLSY data, 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.

by ability quantile, we find that optimal capital tax rates are regressive but quantitatively small relative to existing rates. For the baseline quantitative example the maximal capital tax rate in the nonlinear case is less than 0.14%, and the optimal linear capital tax rate is 0.08%. Moreover, welfare gains from these optimal capital taxes are negligible.[6] These results provide little support for the claim that preference heterogeneity may justify capital taxation, whether in a linear or regressive form.

The quantitative case for regressive housing taxes or subsidies is stronger. Again using data from the NLSY, we show that individuals with higher ability (AFQT score) own houses of greater market value relative to their income history, conditional on (a polynomial in) accrued lifetime income, gender, and age. Using an estimate of the mean preference for housing consumption by ability quintile, we calculate the optimal policy treatment of housing. In our baseline case, owner-occupied housing consumption should be regressively taxed (or subsidized), with the maximum distortion reaching 15 percent for the lowest-ability workers and 0 percent for the highest ability workers. The welfare gain from this optimal policy is orders of magnitude greater than that from optimal capital taxes and equals 3.7 billion dollars in terms of consumption variation for the United States, nearly 4% of the tax revenue lost to the mortgage interest deduction each year. The optimal distortions we simulate are in line with the pattern and levels of the effective mortgage interest deduction in the United States as estimated in Poterba and Sinai (2009). These results may provide one explanation for the persistence of the mortgage interest deduction despite decades of calls from tax experts for its removal.

Finally, this paper studies the importance of preference normalization in our optimal taxation model. We normalize preferences over commodities in two ways. These normalizations are similar to two assumptions made by Saez (2002) in his analysis of optimal commodity taxes with preference heterogeneity. First, we normalize preferences to eliminate any incentive for the planner to redistribute across agents based simply on their preferences over goods. Specifically, the marginal social value to a Utilitarian planner of allocating resources to an undistorted individual is independent of that individual's preferences over consumption goods. This normalization makes it more likely that the optimal capital tax is positive than if we assumed more patient individuals generated higher social welfare for the planner. We also normalize preferences in a second way. We model preferences over commodities, including future and current consumption, as having no direct effect on the labor supply decisions of individuals. Because the challenge of optimal tax policy is to encourage the high-skilled to work despite redistributive taxation, this normalization has a direct influence on the resulting form of optimal policy. This second normalization contrasts with the approach in recent work on optimal taxation of capital with heterogeneous discount rates by Diamond and Spinnewijn (2009), who model preferences such that more patient individuals are more willing to work.

The paper proceeds as follows. Section 1 provides an illustrative example of our theoretical results in an economy with two skill types and heterogeneity in preferences over two goods. Section 2 specifies a general model of optimal taxation with heterogeneity in ability and preferences and derives conditions on the optimal policy. In Sections 3 and 4, we parameterize the model with data on heterogeneous preferences for consumption over time and owner-occupied housing and calculate the optimal taxes for these data. For housing, we calculate and decompose the welfare benefits of the optimal policy, and we compare the results of our baseline simulation to the existing mortgage interest deduction in U.S. tax policy. In Section 5, we consider a generic calibration of the model to test the robustness of our results to a wide range of parameterizations and identify patterns in the decomposition of the welfare gain. Section 6 discusses the

[6] If we took into account variation around mean preference values within ability levels, the optimal taxes and welfare gains are likely to be even smaller.

importance of preference normalization in these models. An Appendix contains technical details referred to in the text.

1 A simple example

In this section we provide a simple example that captures the main intuition behind the more general model.[7] We show that, in this setting, the optimal commodity tax is regressive for goods preferred by the skilled. In particular, the tax is positive on the low-skilled individual's purchase of goods that are preferred by the high-skilled, while the high-skilled individual faces no distortion.

Consider an economy populated by a continuum of measure 1 of two types of individuals $i = \{1, 2\}$, where the size of each group is equal to 1/2. These individuals differ in wage w^i, where $w^2 > w^1$. The wage is private information to the agent. Suppose there are two commodities, c_1 and c_2. The utility function for an individual with wage w^i is given by:

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

The planner's problem is to specify consumption and income allocations for each individual to maximize a Utilitarian social welfare function.

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

$$\max_{\left\{c_1^i, c_2^i, y^i\right\}_{i=1,2}} \sum_i u\left(c_1^i, c_2^i, \frac{y^i}{w^i}\right) \tag{1}$$

subject to

$$u\left(c_1^2, c_2^2, \frac{y^2}{w^2}\right) \geq u\left(c_1^1, c_2^1, \frac{y^1}{w^2}\right), \tag{2}$$

$$\sum_i y^i - c_1^i - c_2^i \leq 0. \tag{3}$$

Constraint (2) is an incentive compatibility constraint stating that an individual of type $i = 2$ prefers the consumption and income bundle intended for it by the planner $\left\{c_1^2, c_2^2, y^2\right\}$ to a bundle $\left\{c_1^1, c_2^1, y^1\right\}$ allocated to an individual of type $i = 1$.[8] Constraint (3) is feasibility, where we assume that the marginal rate of transformation of commodities is equal to 1.

Let u_n be the partial derivative of $u(c_1, c_2, l)$ with respect to the n^{th} argument. Note that these partial derivatives may depend on the wage rate. For example, using the utility function from the general model, (18), $u_1\left(c_1^i, c_2^i, l^i\right) = \frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)} \frac{1}{c_1^i}$. Let μ be the multiplier on constraint (2). Using the first order conditions for consumption in the above problem, we obtain the following expressions for and individual of type $i = 2$:

$$\frac{u_1\left(c_1^2, c_2^2, \frac{y^2}{w^2}\right)}{u_2\left(c_1^2, c_2^2, \frac{y^2}{w^2}\right)} = 1. \tag{4}$$

[7] Similar examples are found in Diamond (2007) and Diamond and Spinnewijn (2008). However, as discussed in Section 6, we normalize preferences in important ways that these other examples do not. This normalization has direct effects on the optimal policies we derive.

[8] Writing this constraint we assumed that only an individual of type $i = 2$ can misrepresent his type. This is easy to ensure if the ratio w^2/w^1 is high enough.

and for the individual of type $i = 1$:

$$\frac{u_1\left(c_1^1, c_2^1, \frac{y^1}{w^1}\right)}{u_2\left(c_1^1, c_2^1, \frac{y^1}{w^1}\right)} = \frac{1 - \frac{u_2\left(c_1^1, c_2^1, \frac{y^1}{w^2}\right)}{u_2\left(c_1^1, c_2^1, \frac{y^1}{w^1}\right)}\mu}{1 - \frac{u_1\left(c_1^1, c_2^1, \frac{y^1}{w^2}\right)}{u_1\left(c_1^1, c_2^1, \frac{y^1}{w^1}\right)}\mu}. \tag{5}$$

Equation (4) shows that the consumption choices of the high-skill individual are undistorted. The marginal rate of substitution $\frac{u_1(\cdot)}{u_2(\cdot)}$ is equal to the marginal rate of transformation. Equation (5) shows that if the multiplier μ on the incentive compatibility constraint is not equal to zero, then the consumption choices of the low-skill individual are distorted. In particular, if an individual's ratio $\frac{u_1}{u_2}$ is less than 1, the policy has caused him to consume more of good 1 relative to good 2 than he would have chosen in autarky.

Now, suppose we impose a condition requiring that if all individuals are given the same consumption and income allocation, the marginal utility of good 2 relative to good 1 is higher for the high-ability individual j (type 2) than for the low-ability individual i (type 2).

Assumption 1 *If* $w^j > w^i$:

$$\frac{u_2\left(c_1, c_2, \frac{y}{w^j}\right)}{u_1\left(c_1, c_2, \frac{y}{w^j}\right)} > \frac{u_2\left(c_1, c_2, \frac{y}{w^i}\right)}{u_1\left(c_1, c_2, \frac{y}{w^i}\right)} \tag{6}$$

for any (c_1, c_2, y).

We now can summarize the argument in a proposition characterizing the distortions in the optimal allocation.

Proposition 2 *Suppose that* $\left\{c_1^i, c_2^i, y^i\right\}_{i=1,2}$ *is an optimal allocation solving (1). Then the optimal choice of consumption for the high-skill individual is not distorted. Suppose that Assumption 1 holds. Then the optimal choice of consumption of good 1 versus consumption of good 2 for the low-skill agent is distorted downwards:*

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

This Proposition states that if good 2 is particularly enjoyed by high-skilled workers, the planner should impose a positive distortion (a positive tax) on the consumption of good 2 by the low-skilled workers (but not on consumption of that good by high-skilled workers). The intuition for this result is as follows. The planner wants to discourage a high-skill individual from deviating and claiming that he is a low type. A high-skill agent will find deviating less attractive if doing so will cause him to face a positive tax on the good that he values highly. The cost of such a positive tax is a distortion in the consumption choices by the low-skill 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 skill level, not income. In particular, it does not apply to goods with an income elasticity of demand greater than 1 but for which preferences are unrelated to skill. For those goods, the inequality in (6) 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 regressive taxes requires the high-skilled to prefer good 2 even when at the same income level as the low-skilled.

2 Model

In this section, we set up a model with a continuum of ability types, as in the classic Mirrlees (1971) framework. We derive a formula for optimal relative commodity taxes that are allowed to be nonlinear in consumption and to depend on income. To capture preference heterogeneity, we assume that preferences across consumption goods are a function of ability. This simplifies the planner's problem by retaining a single dimension of heterogeneity: two or more dimensions introduce multiple screening problems for which a tractable analytical approach has not been developed.[9]

There is a continuum of measure one of individual agents. We index agents 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 $\{w_{\min}, w_{\max}\}$. The ability is private information to the agent. The utility function of an individual depends on $\alpha\left(w^i\right)$, so that the preference parameter for an individual depends directly on his or her wage.

Each individual maximizes the utility function:

$$U\left(w^i\right) = u\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right). \tag{7}$$

Note that utility is a function of the consumption of good 1, c_1, and the consumption of good 2, c_2, as well as of labor effort l, and the preference parameter $\alpha\left(w^i\right)$. Superscripts i on consumption and labor denote the values of these variables for the individual of wage w^i.

A social planner maximizes a utilitarian social welfare function. The planner offers incentive compatible triplets of $\left\{c_1^i, c_2^i, y^i\right\}_i$.

Problem 3

$$\max_{\left\{c_1^i, c_2^i, y^i\right\}_i} \int_{w_{\min}}^{w_{\max}} \left(c_1^i, c_2^i, \frac{y^i}{w^i}\right) f\left(w^i\right) dw^i u \tag{8}$$

subject to

$$\int_{w_{\min}}^{w_{\max}} \left(y^i - c_1^i - c_2^i\right) \leq 0. \tag{9}$$

and

$$u\left(c_1^i, c_2^i, \frac{y^i}{w^i}\right) \geq u\left(c_1^j, c_2^j, \frac{y^j}{w^i}\right), \tag{10}$$

for all i, j.

Constraint (10) is the incentive compatibility constraint stating that an individual of type i prefers the consumption and income allocation intended for it by the planner to an allocation intended for an individual of type j.

Solving the planner's problem in equations (8) through (10) can yield insights into the wedges that optimal policy drives into private optimization.

It is standard to rewrite the planner's problem with explicit tax functions. In this alternative formalization of the problem, each individual maximizes the utility function (7) subject to the individual's after-tax budget constraint,

$$l^i w^i - T\left(w^i l^i\right) - \left(c_1^i + t^1\left(w^i l^i, c_1^i\right)\right) - \left(c_2^i + t^2\left(w^i l^i, c_2^i\right)\right) \geq 0. \tag{11}$$

[9] 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.

The budget constraint requires careful examination. The nonlinear income tax $T\left(w^i l^i\right)$ is a continuous, differentiable function of income $y^i = w^i l^i$. The two other tax functions, $t^1\left(w^i l^i, c_1^i\right)$ and $t^2\left(w^i l^i, c_2^i\right)$, 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. The budget constraint (11) has the multiplier μ.

To characterize optimal taxes with this formalization of the planner's problem, we follow the formal techniques of the Mirrleesian literature. In particular, we consider the following social planner's problem:

Problem 4 *Planner's Problem*

$$\max_{\left\{c_1^i, c_2^i, l^i\right\}_i} \int_{w_{\min}}^{w_{\max}} U\left(w^i\right) f\left(w^i\right) dw^i \tag{12}$$

subject to feasibility

$$\int_{w_{\min}}^{w_{\max}} \left(w^i l^i - c_1^i - c_2^i\right) f\left(w^i\right) dw^i \tag{13}$$

and incentive compatibility, which is that each individual maximizes (7) *subject to* (11) *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 a budget constraint that assumes no government spending for simplicity. The government must also take into account that each individual will choose labor supply to maximize his or her utility subject to the specified tax system.

2.1 The optimal commodity choice wedge

We now derive a formula for the optimal commodity wedge, i.e., the wedge distorting commodity choices.[10] We formulate the Hamiltonian from the planner's problem above. The Hamiltonian includes the following differential constraint:

$$\frac{\partial U^i}{\partial w^i} = u_{w^i}\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right) + \mu\left(l^i\left(1 - T'\left(w^i l^i\right) - t_{y^i}^1\left(w^i l^i, c_1^i\right) - t_{y^i}^2\left(w^i l^i, c_2^i\right)\right)\right), \tag{14}$$

derived using the envelope condition on the individual's utility maximization problem. To remove the tax functions from this expression, we use the individual's first order condition with respect to labor l^i :

$$u_{l^i}\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right) = -\mu w^i\left(1 - T'\left(w^i l^i\right) - t_{y^i}^1\left(w^i l^i, c_1^i\right) - t_{y^i}^2\left(w^i l^i, c_2^i\right)\right). \tag{15}$$

Substituting (15) into (14) yields:

$$\frac{\partial U^i}{\partial w^i} = u_{w^i}\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right) - \frac{l^i u_{l^i}\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right)}{w^i}.$$

The Hamiltonian is then:

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

where subscripts denote partial derivatives and $(\cdot)$ denotes the set of arguments of the utility function, $\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right)$. The first term of the Hamiltonian is the utility of the individual with wage w^i. The

[10] For a textbook treatment, see Salanie (2003), chapter 5.2.

second is government's budget constraint multiplied by a shadow price 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 costate variable ϕ.

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 is:

$$\lambda\left(-1-\frac{dc_2^i}{dc_1^i}\right)\frac{\pi^i}{dw^i}+\phi\left(u_{w^i c_1^i}(\cdot)+u_{w^i c_2^i}(\cdot)\frac{dc_2^i}{dc_1^i}-\frac{l^i u_{l^i c_1^i}(\cdot)}{w^i}-\frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\frac{dc_2^i}{dc_1^i}\right)=0,$$

or, rearranging

$$\frac{dc_2^i}{dc_1^i}=\frac{-\left(\lambda\frac{\pi^i}{dw^i}-\phi\left(u_{w^i c_1^i}(\cdot)-\frac{l^i u_{l^i c_1^i}(\cdot)}{w^i}\right)\right)}{\left(\lambda\frac{\pi^i}{dw^i}-\phi\left(u_{w^i c_2^i}(\cdot)-\frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\right)\right)}.$$

Individuals maximizing (7) subject to (11) will allocate their after-tax income so that the following relationships hold:

$$\frac{dc_2^i}{dc_1^i}=-\frac{u_{c_1^i}}{u_{c_2^i}}=-\frac{1+t1_{c_1}\left(w^i l^i, c_1^i\right)}{1+t2_{c_2}\left(w^i l^i, c_2^i\right)} \tag{16}$$

so we can write:

$$\frac{1+t1_{c_1}\left(w^i l^i, c_1^i\right)}{1+t2_{c_2}\left(w^i l^i, c_2^i\right)}=\frac{\lambda\frac{\pi^i}{dw^i}-\phi\left(w^i\right)\left(u_{w^i c_1^i}(\cdot)-\frac{l^i u_{l^i c_1^i}(\cdot)}{w^i}\right)}{\lambda\frac{\pi^i}{dw^i}-\phi\left(w^i\right)\left(u_{w^i c_2^i}(\cdot)-\frac{l^i u_{l^i c_2^i}(\cdot)}{w^i}\right)}. \tag{17}$$

To fully characterize the optimal distortion to commodity purchases given by (17), we solve for λ and $\phi\left(w^i\right)$ in a specific example.

2.1.1 A specific example

We assume the individual utility function is

$$U^i=u\left(c_1^i, c_2^i, l^i, \alpha\left(w^i\right)\right)=\frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)}\ln c_1^i+\frac{1}{1+\alpha\left(w^i\right)}\ln c_2^i-\frac{1}{\sigma}\left(l^i\right)^{\sigma}. \tag{18}$$

It is important to note that this utility function normalizes preferences over consumption goods in the two ways mentioned in the Introduction. The first normalization, following the techniques of Weinzierl (2009), ensures that the marginal social value to a Utilitarian planner of allocating resources to an undistorted individual is independent of that individual's preference parameter $\alpha\left(w^i\right)$. This prevents preference heterogeneity, which is inherently ordinal, from artificially driving redistribution by making the cardinal utility of consumption higher for an individual depending on his or her preferences. The second normalization separates heterogeneity in commodity preferences from the consumption-leisure choice of individuals. Specifically, it ensures that two individuals of the same ability w^i will choose the same labor effort when undistorted.[11]

The next proposition derives an expression for the optimal commodity taxes.

[11] Logarithmic utility of consumption makes is possible to achieve these two normalizations simultaneously. For a more general case, the Appendix to this paper contains the details of both normalizations.

Proposition 5 *Given the individual utility function* (18), *the solution to the Planner's Problem satisfies:*

$$\frac{1+t_{c_1}^1\left(w^i l^i, c_1^i\right)}{1+t_{c_2}^2\left(w^i l^i, c_2^i\right)} = \frac{f\left(w^i\right) + u_{w^i c_1^i}\left(1 - F_t\left(w^i\right)\right)\left(\begin{array}{c} \frac{1}{1-F(w^i)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j \\ -\frac{1}{1-F(w_{\min})}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j \end{array}\right)}{f\left(w^i\right) + u_{w^i c_2^i}\left(1 - F\left(w^i\right)\right)\left(\begin{array}{c} \frac{1}{1-F(w^i)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j \\ -\frac{1}{1-F(w_{\min})}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j \end{array}\right)} \tag{19}$$

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

$$\lambda = \frac{1}{\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c_2^j}} f\left(w^j\right) dw^j}$$

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

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

As with the conditions for optimal marginal income tax rates from, e.g., Saez (2001), concave utility of consumption prevents result (19) from being fully closed-form, instead relying on optimal utility and consumption levels. Nevertheless, we can establish some important lessons from it.

First, on the top type, $(1 - F(w_{\max}))$ is zero, and the result reduces to

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

so the commodity distortion is zero on the highest ability worker.

Second, the distortion is also zero on the lowest ability worker, as the terms in large parentheses in the numerator and denominator are zero.

In addition, examination of terms in (19) gives detail about the determinants of the optimal distortion.

The parenthetical term common to the numerator and denominator is the difference in the average cost of raising utility for the population with wages above w^i and for the entire population. It is positive, since if it were negative the planner could raise social welfare by incentive-compatible and feasible transfers of c_2 from the overall population to the high-skilled. As such, this difference measures the loss in welfare that results from having to satisfy the incentives of the high-skilled rather than being able to spread resources across all workers. When this loss is large, the optimal distortion to consumption at wage w^i is larger because that distortion discourages higher-skilled workers from working less.

The relationship between $u_{w^i c_1^i}$ and $u_{w^i c_2^i}$ determines whether policy discourages consumption of good 1 or good 2 for intermediate ability levels. With utility function (18) this relationship is determined by the sign of $\alpha'\left(w^i\right)$. If $\alpha'\left(w^i\right) < 0$, then high-ability workers relatively prefer good 2, and $u_{w^i c_1^i} < 0$ while $u_{w^i c_2^i} > 0$. Then, the ratio on the right-hand side of (19) is less than one, and the optimal distortion discourages marginal consumption of good 2. That is, the good preferred by the more able workers ought to be marginally taxed.

The term $f\left(w^i\right)$ provides a measure of the share of the population distorted by a given commodity tax. When this share is high, the optimal consumption distortion is smaller, as the planner wants to avoid distortions on large sub-populations. Mathematically, $f\left(w^i\right)$ enters both the numerator and the denominator, pushing the tax ratio toward unity.

The term $\left(1-F\left(w^i\right)\right)$ is the share of individuals with higher wages who are encouraged to exert more effort due to the distortion at w^i. The larger this term, the more valuable is the distortion to the planner, all else the same. Mathematically, $\left(1-F\left(w^i\right)\right)$ multiplies the terms in the numerator and denominator that push the tax ratio away from unity. We know that $\left(1-F\left(w^i\right)\right)$ falls as the wage rises, so this lowers the optimal distortion as we move up the ability distribution.

Finally, suppose there exists an ability level $\tilde{w}$ such that the distribution of all abilities above that level follows a Pareto form, as in Saez (2001). Then for all such $w^i > \tilde{w}$, $\frac{w^i f\left(w^i\right)}{\left(1-F\left(w^i\right)\right)}$ is constant. Rearrange the expression (19) to obtain

$$\frac{1+t^1_{c_1}\left(w^i l^i, c^i_1\right)}{1+t^2_{c_2}\left(w^i l^i, c^i_2\right)} = \frac{\frac{w^i f\left(w^i\right)}{1-F\left(w^i\right)} + u_{w^i c^i_1} w^i \left(\frac{1}{1-F\left(w^i\right)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c^j_2}} f\left(w^j\right) dw^j - \frac{1}{1-F\left(w_{\min}\right)}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c^j_2}} f\left(w^j\right) dw^j\right)}{\frac{w^i f\left(w^i\right)}{1-F\left(w^i\right)} + u_{w^i c^i_2} w^i \left(\frac{1}{1-F\left(w^i\right)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{1}{u_{c^j_2}} f\left(w^j\right) dw^j - \frac{1}{1-F\left(w_{\min}\right)}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{1}{u_{c^j_2}} f\left(w^j\right) dw^j\right)}.$$

From above, we know that the parenthetical terms are positive; they are also increasing in w^i following the same argument. Therefore, assuming $u_{w^i c^i_1} < 0$ and $u_{w^i c^i_2} > 0$, whether the optimal tax on good 2 is regressive or progressive in the upper tail of the income distribution depends on how quickly $u_{w^i c^i_1}$ and $u_{w^i c^i_2}$ converge to zero. If they do not converge quickly enough, the tax on good 2 is progressive in the tail.

Though these interpretations aid in understanding result (19), we may want to reformulate that result in terms of observable quantities in the spirit of Saez (2001). The Appendix derives the following version of result (19):

$$\frac{1+t^1_{c_1}}{1+t^2_{c_2}} = \frac{w^i f\left(w^i\right) + \varepsilon_{c^i_1 w^i}\frac{1+t^1_{c_1}}{1+t^2_{c_2}}\left(1-F\left(w^i\right)\right)\left(\frac{1}{1-F\left(w^i\right)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{\hat{y}^j}{\hat{y}^i} f\left(w^j\right) dw^j - \frac{1}{1-F\left(w_{\min}\right)}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{\hat{y}^j}{\hat{y}^i} f\left(w^j\right) dw^j\right)}{w^i f\left(w^i\right) + \varepsilon_{c^i_2 w^i}\left(1-F\left(w^i\right)\right)\left(\frac{1}{1-F\left(w^i\right)}\int_{w^j=w^i}^{w^j=w_{\max}} \frac{\hat{y}^j}{\hat{y}^i} f\left(w^j\right) dw^j - \frac{1}{1-F\left(w_{\min}\right)}\int_{w^j=w_{\min}}^{w^j=w_{\max}} \frac{\hat{y}^j}{\hat{y}^i} f\left(w^j\right) dw^j\right)} \tag{20}$$

where $\varepsilon_{c_m w}$ denotes the Frisch elasticity (holding marginal utility constant) of consumption of good m with respect to the wage, $\hat{y}^i$ is the disposable income individual i would choose to earn in an economy with income taxes only (i.e., before the introduction of optimal commodity taxes, the planner can observe the distribution of $\hat{y}^i$). This alternative representation of the main result on optimal commodity taxes can be more readily applied with observable data.

If we restrict attention to commodity taxes that are a linear function of the consumption of the good, a modification of result (19) confirms the results of the previous literature (e.g., Saez 2002, Salanie 2003) that goods preferred by the highly able ought to be taxed.

3 Example 1: Capital Taxes

The results of Sections 1 and 2 suggest that optimal commodity taxes may be regressive on goods preferred by the high-skilled, but the analytical expression (19) made it clear that the shape of optimal commodity taxes will depend on many details of the economy. In this section and the next, we study the shape of optimal commodity taxation numerically, focusing on two prominent cases: capital income and housing.

We simulate the optimal tax treatment of capital income using empirical evidence on the relationship between ability and time preference, or intertemporal discounting. We consider two sources of data: Lawrance (1991) and our own analysis of the National Longitudinal Survey of Youth.

Our baseline case for these simulations will use the utility function previously given in expression (18) :

$$u\left(c^i, l^i, w^i\right) = \frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)} \ln c_1^i + \frac{1}{1+\alpha\left(w^i\right)} \ln c_2^i - \frac{1}{\sigma}\left(l^i\right)^{\sigma}$$

with $\sigma = 3$, for a constant-consumption elasticity of labor supply with respect to the wage of $\frac{1}{2}$. In the context of capital taxation, we interpret c_1 and c_2 as consumption in two different time periods. We will study the ratio:

$$\frac{t^2_{c_2^i} - t^1_{c_1^i}}{1 + t^2_{c_2^i}} = \frac{u_{c_2^i} - u_{c_1^i}}{u_{c_2^i}} \tag{21}$$

which measures the relative distortion provided by the tax system toward good 1 and away from good 2 at each income level. Under the capital tax interpretation, it measures the distortion to the individual's intertemporal Euler condition, which would set $u_{c_2^i} = u_{c_1^i}$. If this ratio is positive, the tax policy is discouraging future consumption relative to current consumption. More informally, it is taxing saving, so we will refer to it as the implied capital tax.

3.1 Based on Lawrance (1991)

Lawrance (1991) calculates annual time preference rates, by income percentile, in a standard intertemporal optimization model using data on food consumption. We denote her estimate of the discount rate for percentile i as ρ^i. For each income percentile, we find the wage at the same percentile in the 2007 U.S. wage distribution, and we assume that wages proxy for ability. For ease of comparison with the theoretical model, we also provide the implied value for $\alpha\left(w^i\right) = \exp\left(\rho^i\right)$, the preference parameter in the utility function (18). A representative set of Lawrance's results, and the implied $\alpha\left(w^i\right)$ and w^i values, are as follows:

Income percentile	5^{th}	10	25	50	75	90	95
ρ^i (discount rate)	.064	.049	.042	.039	.036	.035	.034
$\alpha\left(w^i\right)$	1.066	1.050	1.043	1.040	1.037	1.036	1.035
Wage (ability)	$5	7	10	15	22	31	38

The data in this table imply that the annual rate at which the highest-ability individuals discount future flows of consumption is approximately half that of the lowest-ability individuals. The pattern of ρ^i translates into a declining but concave path for $\alpha\left(w^i\right)$. Wages rise rapidly with income percentiles, such that the highest-ability group in the table have a wage more than seven times as high as the lowest-ability group.

Figure 1 plots the discount rates ρ^i from Lawrance (1991) and the simulated optimal capital tax rates against wages. It also shows the optimal linear capital tax rate, which we obtain by forcing the planner to set the (21) equal across wage types.

The results confirm that optimal capital taxes are regressive. However, Figure 1 shows that the optimal distortion is very small throughout the ability distribution. The highest marginal tax is less than 0.8%. The optimal linear capital tax is also very small (less than 0.2%). The welfare gains from adding linear or optimal nonlinear capital taxes are close to negligible. Starting with no capital taxation, adding optimal nonlinear capital taxation yields a consumption-equivalent gain of 0.0002 percent of total output, about half of which is due to the gain of going from the best linear capital tax to the optimal nonlinear capital tax.

3.2 Based on NLSY data

In this calibration, we use data from the National Longitudinal Survey of Youth (NLSY), a nationally representative sample of individuals born between 1957 and 1964 and first interviewed in 1979. This sample has been interviewed annually or biannually since. The key advantage of the NLSY for our purposes is that it contains data on individuals' net worth and income over time as well as a standard measure of ability. These data allow us to compute a measure of permanent income as well as to avoid using the wage as a proxy for ability (as we did with the data from Lawrance 1991). 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. 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 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 perhaps 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:

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

subject to

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

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 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. The first-order conditions of the individual's problem yield the following expression for δ:

$$1 + \delta + \delta^2 = \frac{y_1}{c_1}\frac{1+R}{R}.$$

or

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

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, 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. Formally, $y_1 = \sum_{t=1979}^{2004} r^{(2004-t)} y_t$, where $r = 1.05$ is the annual gross rate of return.[12] 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 $c_1 = y_1 - NW$.

In the following table, we show the mean and standard deviations of δ by AFQT quintile:

AFQT89 quintile	*Bottom*	2	3	4	*Top*
δ	0.34	0.37	0.39	0.42	0.47
st.dev. of δ	0.16	0.18	0.18	0.21	0.25

The variation in δ within AFQT quintiles is large relative to the variation across wage levels. Of course, the data are likely to be very noisy, and our inference of δ is based on a highly simplified model. Nevertheless, our findings are consistent with Lawrance's data above and with the findings of Benjamin, Brown, and Shapiro (2006), who find a "positive relationship between AFQT score and the propensity to have positive net assets" in the NLSY.

Consistent with these patterns, Table 1 shows the results of a regression of δ on age, age squared, gender, a 4th-degree polynomial in y_1 (to control for income effects), and AFQT score for the same sample.[13] Though the explanatory power of this set of independent variables is low, the coefficient on AFQT score is positive and significant at the 1 percent level. The magnitude of the coefficient, 0.00066, implies that a twenty-point increase in AFQT raises δ by .013, roughly in keeping with the pattern by quintile shown above.

To perform the optimal policy simulations, we convert δ into an annualized discount rate ρ^i for each AFQT quintile using the following identities: $\delta^{\frac{1}{23}} = \beta^i$, $\alpha\left(w^i\right) = \frac{1}{\beta^i}$, and $\alpha\left(w^i\right) = \exp\left(\rho^i\right)$. The NLSY also has data on wage and salary earnings and hours worked. We use these to impute a wage for each individual. We use data from 1992, the middle of the observed data range, to estimate wages by AFQT89 quintile. The results are:

AFQT89 quintile	*Bottom*	2	3	4	*Top*
β^i	0.954	0.958	0.960	0.963	0.967
ρ^i (discount rate)	0.047	0.043	0.040	0.038	0.033
$\alpha\left(w^i\right)$	1.049	1.044	1.041	1.039	1.034
Mean wage (ability)	\$8.93	11.11	12.23	13.44	15.49

Figure 2 plots ρ^i and the optimal capital tax from expression (21) for each AFQT89 quintile against the wage. The same lessons hold as in the simulations using the results of Lawrance (1991): optimal nonlinear capital taxes are regressive but are extremely small, the highest tax is less than 0.14%. The best linear capital tax is positive but even smaller (0.08%). The welfare gain from optimal capital taxation is nearly zero in this simulation, which is not surprising given the relatively small variation in preferences apparent from the NLSY data.

[12] 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.

[13] 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 0.0043, but it remains significant at the 1% level.

In sum, evidence from two sources we considered yields too weak a relationship between ability and time preferences to justify, in our model, substantial capital taxation, whether linear or regressive.

4 Example 2: Housing

While our analysis above did not find a justification for quantitatively substantial regressive (or even linear) capital taxation, nonlinear taxation may still be important for other categories of consumption. In this section, we consider one example: owner-occupied housing. Building on the results of the previous sections, if individuals of greater ability have a greater preference for the consumption of housing services, regressive subsidization of housing may be optimal.

One reason to study the potential for this relationship is the existence of a regressive housing subsidy in the United States, specifically the mortgage interest deduction. This policy has repeatedly come under fire from academic economists as not only distortionary but unfair, as larger benefits accrue to those in higher tax brackets and who can afford more expensive homes. Despite academic opposition to the mortgage interest deduction, and repeated proposals for the deduction's repeal, political pressures have kept it alive. This section provides one possible explanation for that survival: a regressive housing subsidy may be optimal. If high-ability workers particularly value housing services, they may be willing to accept higher earned income taxes in exchange for preferential treatment of housing. We show that an implied optimal subsidy to housing may be significant and comparable to that found in empirical studies such as Poterba and Sinai (2009).

One reason why high-ability workers may pay more for housing, holding income constant, is that these workers highly value a quality public school system.[14] In the United States, local property taxes fund most of the public school systems, so the prices of homes rises dramatically with the quality of the schools to which they give access (see, e.g., Black 1999). In the end, more redistribution is made possible by the regressive housing subsidy. Of course, whether this is a plausible explanation depends quantitatively on how strong such a correlation of preferences and ability is, if it exists at all.

Therefore, we turn to estimating the preference for housing by ability level. Suppose individuals maximize the following utility function

$$\max_{c_1,c_2,c_3} \left[\ln(c) + \eta \ln(h) - v(y)\right]$$

where c is consumption, h is spending on housing, and y is income, subject to the budget constraint

$$y - c - h = 0.$$

The first order conditions yield:

$$\frac{1}{\eta} h = c$$

which, in the budget constraint, implies

$$\frac{h}{y} = \frac{\eta}{1+\eta}$$

for the value of expenditure on housing as a share of total income.

To examine whether η varies with ability, we return to the NLSY data from above. We use the reported 2004 market value of the respondent's primary residence as the measure of h. For income y, we use the same cumulative income measure as in the capital tax simulations from the previous section. There are 7,280

[14] On the positive relationship between ability and the returns to schooling, which could generate the preference pattern suggested here, see Belzil and Hansen (2002).

observations for $\frac{h}{y}$. The median value of h is \$85,000, the mean is \$152,058, and the maximum is \$1.816 million.[15] The following table shows, by AFQT quintile for these observations, the ratio $\frac{h}{y}$, its standard deviation, the value of the taste parameter $\alpha\left(w^i\right) = \left(\frac{1}{\eta^i} - 1\right)$, and the average wage for individuals.

AFQT89 quintile	*Bottom*	2	3	4	*Top*
$\frac{h}{y}$	0.080	0.116	0.131	0.157	0.170
st.dev. of $\frac{h}{y}$	0.150	0.169	0.161	0.221	0.188
$\alpha\left(w^i\right)$	10.5	6.6	5.6	4.4	3.9
Mean wage (ability)	\$8.93	11.11	12.23	13.44	15.49

These are the same AFQT quintiles, with the same corresponding mean wages, as were used in the NLSY capital tax simulation of Section 3. Note that the variation in $\frac{h}{y}$ is large within quintile. We will ignore this variation and use the point estimates of $\frac{h}{y}$ to simulate optimal tax policy toward housing consumption. The data imply the highest-skilled individuals place approximately twice the weight on housing relative to other goods compared to the lowest-skilled individuals.

Regression results provide support for the relationship between ability and preferences seen in this table. Table 2 shows the results of a regression of $\frac{h}{y}$ on age, age squared, gender, a 4th-degree polynomial in y_1 (to control for income effects), and AFQT score.[16] The coefficient on AFQT score is positive and highly significant, with a t-statistic of 8.38. In magnitude, the coefficient of 0.0007 implies that a twenty-point increase in AFQT, i.e., approximately a one-quintile increase, would generate an increase in $\frac{h}{y}$ of approximately 0.014. This magnitude is consistent with the pattern of $\frac{h}{y}$ across AFQT quintiles shown above.

Using the distribution of ability and preferences for housing given above, we simulate optimal taxation of housing. We will summarize policy with the term

$$\frac{u_h - u_c}{u_h} = \frac{\frac{\eta}{h} - \frac{1}{c}}{\frac{\eta}{h}}, \tag{22}$$

which measures the relative distortion toward non-housing consumption and away from housing consumption at each income level. As in the capital tax simulations, we assume that the normalized utility function in the social welfare function takes the form (18) with $\sigma = 3$.

Figure 3 plots $\frac{h}{y}$, expression (22), and the optimal linear housing tax against wages. The positive relationship between the mean values of $\frac{h}{y}$ and AFQT scores generates a sizeable and regressive optimal tax policy toward housing consumption. Figure 3 shows that the optimal tax on housing consumption, relative to other consumption, starts at 15 percent at the bottom of the income distribution and falls, with one blip up in the middle of the distribution, as income rises.

The welfare gain from this regressive housing distortion is orders of magnitude larger than the gain estimated from the optimal capital taxes, though it remains moderate in absolute terms. The estimated gain is 0.03 percent of total income, or just over \$3.7 billion in the current U.S. economy. About half of this gain could be captured by using the optimal linear tax on housing shown in Figure 3.

We can decompose the welfare gain from the optimal nonlinear housing tax into four components, as shown in Table 3. The first component is "efficiency," or the increase in overall production in the economy

[15] If we restrict the sample to the 4,716 observations for the ratio $\frac{h}{y}$ is positive, the median value is \$160,000 and the mean value is \$234,531.

[16] 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 0.0005, but it remains significant at the 1% level.

due to lower distortions to labor effort. The second is the reallocation of consumption across individuals in general, and toward less able individuals in particular, that is made possible by the housing subsidy. The third is the reallocation of consumption across goods for each individual. Finally, the fourth component is the reallocation of required income across individuals, in particular toward those with high ability and, therefore, low marginal disutility of earned income.

Table 3 shows that the largest contributor to the welfare gain is the redistribution of consumption across types, which yielded a gain equal to 169% of the overall increase. Partially offsetting this was, as would be expected, a large decrease in welfare due to reallocation of consumption from housing to non-housing consumption due to the distortion. This loss was equal to -116% of the overall increase in welfare. Smaller positive components of the change were greater overall efficiency due to lower distortions (13% of the gain), and a redistribution of required income to those of higher ability (35% of the gain).

Finally, using results from Poterba and Sinai (2009), we can show that the existing mortgage interest deduction in the United States creates distortions of a similar magnitude to that implied by the simulation above. Poterba and Sinai provide average tax savings from the mortgage interest deduction and average market values of the homes of households in five income brackets. For instance, for those with over $250,000 in annual income, the annual tax saving from the deduction is $5,459, and the mean home value is $1.072 million. For each income bracket, we calculate the implied subsidy to housing by assuming a real discount rate and a term length for the mortgage. We assume each household has a 5 percent real discount rate and has financed its house with a standard 30-year mortgage. Then, we calculate the present value of tax deductions as a percent of the market value of the home for each income bracket. We transform this subsidy rate into a relative tax on housing consumption by subtracting the subsidy in each income bracket from the subsidy for the top bracket. We label this transformed series the "existing relative housing tax due to mortgage interest deduction," and we show it in Figure 4. As this figure shows, the existing mortgage interest deduction in the United States is roughly consistent, in terms of magnitude, with the optimal results we have derived.

5 Other numerical simulations and robustness

The data used above for simulating optimal capital and housing tax policy were limited to a narrow range of wages and preferences. To supplement these simulations and to consider robustness, we numerically simulate an example of the model above with a hypothetical distribution of wages and preferences. For ease of interpretation, we consider good 1 to be current consumption and good 2 to be future consumption, so we use the language of capital taxation to describe the results.

5.1 Robustness: preference heterogeneity

We use a wage distribution that runs from $4 to $50 with 24 discrete values. We parameterize the distribution as lognormal, based on the 2007 wage distribution for full-time workers in the United States as reported in the Current Population Survey. We set

$$\alpha\left(w^i\right) = \left(\frac{w^I}{w^i}\right)^{\frac{1}{10}} \tag{23}$$

so that $\alpha\left(w^i\right)$ ranges from approximately 1.29 to 1.00.[17] For our baseline case, we use the utility function previously given in expression (18) with $\sigma = 3$. This utility function and the pattern of $\alpha\left(w^i\right)$ means that higher-ability individuals have a relative preference for good 2, i.e., future consumption, or saving. For ease of interpretation, we can also convert the preferences to a standard discount rate $\rho^i = \ln \alpha\left(w^i\right)$. Given our assumption on preferences, ρ^i varies from 0.25 to 0.00.

Figure 5 plots ρ^i and the capital tax expression (21) against optimal income. This term measures the relative distortion provided by the tax system toward good 1 and away from good 2 at each income level. In the capital tax interpretation, it measures the distortion to the individual's intertemporal Euler condition, which would set $u_{c_2^i} = u_{c_1^i}$. If this ratio is positive, the tax policy is discouraging future consumption relative to current consumption. More informally, it is taxing saving, so we will refer to it as the implied capital tax.

Figure 5 shows that the optimal implied capital tax is monotonically decreasing in the wage. This suggests that the restriction of previous work to linear commodity taxation is not innocuous. At the same time, our results confirm the previous work's conclusion that the best linear tax on a good preferred by the high-skilled is positive. Figure 5 plots this best linear capital tax: that is, the distortion to intertemporal consumption that results if we restrict the planner to policies that set expression (21) equal for all types i.

As we have merely assumed a pattern of preferences for this example, the magnitude of the optimal taxes and the welfare gains they yield are not meaningful. However, we can use the results in two ways: first, to gauge their sensitivity to alternative parameterizations; second, to decompose the welfare gains into separate parts. The next two subsections take up these tasks. For reference, in this baseline simulation, both the optimal nonlinear and best linear capital tax rates are small. The former never rises above 3.3 percent, while the latter is less than 2 percent. The welfare gain from capital taxation is also very small.[18] Starting from a policy with no capital taxation, the welfare gain is only 0.005 percent of total consumption (and output). The gain from the best linear capital tax to the optimal nonlinear capital tax makes up a small fraction, about 8 percent, of this gain.

5.2 Robustness: risk aversion and elasticity of labor

The results above are for a particular parameterization of the policy problem. Here, we test whether its lessons are robust to variations in the parameterization. In particular, we vary the curvature of utility as a function of consumption and the elasticity of labor supply.

One complication in considering alternative parameterizations is that the simple utility function assumed in (18) is no longer appropriate. Recall that we normalized all individuals' utility functions according to two criteria. First, the marginal social value to a Utilitarian planner of allocating resources to an individual in autarky is independent of its preference parameter $\alpha\left(w^i\right)$. Second, to avoid a direct connection between time preference and consumption-leisure preferences, we wrote the individual utility function so that two individuals with the same ability choose the same labor supply at their autarkic individual optima. With logarithmic utility of consumption, these restrictions are captured by the normalization in expression (18), but with a more general utility function, a more complex normalization is required. In the Appendix, such a normalization is derived. Here, we use the general expression for utility:

$$U = \frac{1}{\varphi^i}\left(\left(\frac{\alpha\left(w^i\right)}{1+\alpha\left(w^i\right)}\right)^{\gamma}\frac{\left(c_1^i\right)^{1-\gamma}-1}{1-\gamma} + \left(\frac{1}{1+\alpha\left(w^i\right)}\right)^{\gamma}\frac{\left(c_2^i\right)^{1-\gamma}-1}{1-\gamma} - \frac{1}{\sigma}\left(l^i\right)^{\sigma}\right) \tag{24}$$

[17] A second dimension along which to check robustness is the assumed pattern for $\alpha\left(w^i\right)$. We are currently examining this.
[18] To measure the welfare gain, we compute the factor by which all consumption levels would have to increase (uniformly) under a restricted policy in order to reach the level of welfare achieved with optimal nonlinear capital taxation.

where φ^i is a normalization factor that depends only on the preference parameters $\alpha\left(w^i\right)$ and the parameters γ and σ.

We solve the same planner's problem as above, but for a range of values for γ and σ. Table 4 shows the implied optimal capital taxes, the implied best linear capital tax rate, and the welfare gains from nonlinear capital taxation for each combination of γ and σ. For the parameterizations with the most plausible, moderate values for γ and σ, the optimal nonlinear capital taxes are small and regressive over the entire wage distribution. In all cases, optimal capital taxes are zero for the highest-wage, highest-patience type. In all but one case, they are monotonically decreasing with the wage; the exception is when $\gamma = 2$, $\sigma = 4.5$, for which the very small optimal capital tax rates take an inverse-U shape, peaking at 2.6 percent. Larger capital tax rates can be obtained only by raising the labor supply elasticity (which makes the income tax more distortionary) or the concavity of utility from consumption (which makes redistribution more valuable to the planner). Even with a larger elasticity of labor supply, however, only the lowest few types face rates well above the typical results, and increasing the concavity of utility from consumption to $\gamma = 4$ yields rates that start below 10% for the lowest types and fall as wage increase. Related to these characteristics of the optimal nonlinear capital taxes, the best linear capital tax rates shown in Table 4 are positive but generally low. Again, we can elicit larger optimal linear rates by increasing the elasticity of labor supply or the concavity of utility from consumption.

The welfare gains from capital taxation in this setting increase with the curvature of utility from consumption and the elasticity of labor supply. More curvature (higher risk aversion) increases the value of redistribution, while more elasticity increases the distortionary costs of traditional taxation. More sophisticated taxation, such as nonlinear capital taxation, is therefore more valuable in these cases. Nevertheless, the welfare gains remain small in these variations, with the highest gain reaching only 0.026 percent of total output in the case with the highest elasticity of labor supply ($\sigma = 1.5$). The incremental reform from the best linear capital tax to the optimal nonlinear tax consistently makes up a small fraction of the gain obtained from reforming a system with no capital taxation at all.

5.3 Welfare gain decomposition

As shown in Table 4, the planner raises social welfare when it uses nonlinear capital taxes. Though these gains are small in size, it is revealing to decompose these welfare gains along the dimensions explored in the optimal housing policy simulation from the previous section.

Table 5 shows that the largest contributor to welfare improvement in all cases is the redistribution of consumption across individuals, while the reallocation of goods for each individual lowers welfare in all cases. Intuitively, the planner's ability to cater to the preferences of high-ability workers for saving makes it possible to redistribute more to those with low ability, while at the same time the taxes that the planner uses to encourage the high-skilled to work reduce the welfare of lower-skilled workers by distorting their choices among commodities.

Table 6 shows another way to measure the increased redistribution made possible by regressive commodity taxes. It shows the average tax rate on the lowest-wage workers under the optimal tax regime, the best linear tax regime, and in the regime of no capital taxation. This average tax rate falls with the sophistication of policy, indicating more redistribution toward lower-wage workers. As might be expected from the previously discussed results, however, the magnitude of the difference in redistribution is not large.

6 Role of preference normalization

In this section, we explore the role of preference normalization in the study of optimal commodity taxation. We normalize preferences in two ways: to neutralize the role of preferences over goods in how much the planner values individuals; and to neutralize the effect of preferences over goods on the labor supply choices of individuals.

First, we normalize so as to eliminate any incentive for the planner to redistribute across agents based simply on their preferences over goods. Consider the following two representations of the same preferences over consumption goods:

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

and

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

Expression (25) normalizes preferences as in our main analysis, whereas (26) does not. Specifically, starting an individuals' undistorted optimal allocations, a planner using (25) has no desire to redistribute across preference types (conditional on the wage) because the marginal social value of resources is equalized across preference types. In contrast, a planner using (26) obtains a larger increase in social welfare from allocating a marginal unit of resources to the individual with lower $\alpha\left(w^i\right)$. Therefore, optimal tax policy will favor individuals with lower $\alpha\left(w^i\right)$. For example, in the context of capital taxation when ability is positively correlated with patience (i.e., $\alpha'\left(w^i\right) < 0$), using (26) rather than (25) will cause the planner to favor those with lower $\alpha\left(w^i\right)$. Since these individuals value saving, the planner will be discouraged from taxing capital. Importantly, if we were to multiply expression (26) through by $\alpha\left(w^i\right)$, the impact of preferences would reverse even though we would be using an observationally equivalent representation of them. In that case, lower $\alpha\left(w^i\right)$ types would yield smaller increases in social welfare to the planner, so the planner would be predisposed toward capital taxation.

Second, we use a representation of preferences that implies no relationship between preferences across goods and the willingness to work. To do so we use the utility function (a generalization of (25) for γ not necessarily equal to one):

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

To see that chosen income is independent of preferences, note that individual maximization yields:

$$\varphi^i \lambda = \left(w^i\right)^{\frac{-2\gamma}{\sigma-1+\gamma}},$$

and thus

$$y^i = \left(w^i\right)^{\frac{\sigma(\sigma-1+\gamma)-2\gamma}{\sigma(\sigma-1+\gamma)-(\sigma-1+\gamma)}}.$$

Without this normalization, we would be forced to assume that preferences over goods were systematically related to preferences between leisure and consumption.

These normalizations are similar to two assumptions Saez (2002) states in his analysis of this topic. His Assumption 1 is that the planner's marginal social welfare weights on individuals are independent of their tastes for goods, conditional on their incomes. Our first normalization pursues the same neutrality of marginal social welfare weights, though we use the laissez-faire allocations rather than the optimal allocations as the starting point for the normalization. This normalization captures the idea that the government does not want to redistribute resources across individuals simply because they will spend them on different consumption baskets. Our second normalization parallels Saez's Assumption 2, which states that, conditional their income, individuals' labor supply responses to tax changes are unaffected by their preferences. Though our normalization focuses on isolating from preferences the chosen level of labor supply, rather than its response to tax changes, the idea of the two approaches is similar. Intuitively, this normalization means that individuals choose how much to work without regard to how they plan to spend their disposable income. Saez notes that both of his Assumptions seem like reasonable ones in the context of capital taxation. We take a similar perspective, believing that our normalizations provide a natural, and neutral, starting point for modeling preference heterogeneity and its effects on optimal commodity taxation.

Diamond and Spinnewijn (2009) use a representation of preferences similar to (26) to study optimal capital taxation. As the discussion above implies, the use of this utility function introduces two factors in determining the optimal tax policy that are absent from our setup. First, if the planner's objective function simply sums utilities of the form in (26), a positive capital tax is less likely to be part of the optimum policy, as $\alpha'\left(w^i\right) < 0$ means that high-wage individuals are more valuable to their planner. For example, recall that our baseline simulation of optimal capital taxes in Section 3 yielded a constrained optimal linear capital tax of 0.08%. Using (26) instead yields a constrained optimal linear capital tax of only 0.04%. Diamond and Spinnewijn consider a planner's objective function that weighs different individuals differently. For their main analysis they assume that the planner puts a small enough weight on the high-skilled so as to want to redistribute away from them. Second, their use of (26) makes capital taxation less distortionary, as the individuals who value saving also respond less elastically to tax changes. As Diamond and Spinnewijn note, if they were to assume an observationally equivalent representation of preferences in which more patient individuals would choose to work less, their results on optimal capital taxation are reversed. Only limited empirical evidence is available to determine which is the appropriate assumption.

7 Conclusion

Among others, Mirrlees (1976) and Saez (2002) have argued that goods preferred by the high-skilled ought to be taxed as part of an optimal tax policy that seeks to redistribute from the (unobservably) high-skilled. This argument has been used, in particular, to justify positive capital income taxation. We show that, contrary to these previous results, optimal commodity taxation when preferences vary with ability may be regressive in income on those goods preferred by those who are more able. We obtain this result by allowing taxes on goods to be nonlinear functions of income and the consumption of the good, which is plausible for many important categories of consumption such as education, health, housing, and future consumption.

We parameterize the model with data on preferences for current relative to future consumption and for housing. We find that the relationship between ability and time discounting is unlikely to justify substantial capital taxation, whether regressive or linear. However, we present a stronger case for regressive taxes or subsidies on housing, such as the mortgage interest deduction in the United States.

References

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

[2] Belzil, Christian and Hansen, Jorgen (2002). "Unobserved Ability and the Return to Schooling," *Econometrica*, 70(6).

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

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

[5] Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2).

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

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

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

[9] Golosov, Mikhail and Aleh Tsyvinski (2006). "Optimal Taxation with Endogenous Insurance Markets," *Quarterly Journal of Economics*, 122(2).

[10] Golosov, Mikhail and Aleh Tsyvinski (2007). "New Dynamic Public Finance: A User's Guide," *NBER Macroannual 2006*, MIT Press.

[11] Hammond, Peter J. (1987). "Markets as Constraints: Multilateral Incentive Compatibility in Continuum Economies," *Review of Economic Studies*, 54.

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

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

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

[15] 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.

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

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

[18] Poterba, James, and Todd Sinai (2009). "Tax Expenditures for Owner-Occupied Housing: Deductions for Property Taxes and Mortgage Interest and the Exclusion of Imputed Rental Income," Working Paper.

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

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

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

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

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

[24] Weinzierl, Matthew (2009). "Incorporating Preference Heterogeneity into Optimal Tax Models." Working Paper.

Table 1. Results of OLS regression of discount factor (delta) on ability and controls

Variable	Coefficient	Standard Error	T-statistic
age	-1.95E-02	1.60E-02	-1.22
agesq	5.98E-04	4.51E-04	1.32
sex	7.88E-03	4.39E-03	1.79
fvincome**	5.46E-08	1.64E-08	3.34
fvincsq	1.45E-14	8.22E-15	1.77
fvinccu**	-3.80E-21	1.27E-21	-2.99
fvincq**	1.43E-28	4.40E-29	3.25
afqt**	6.60E-04	9.36E-05	7.05
Note: * indicates significance at the 5% level or lower; ** at 1%			
Observations	7,008		
F-statistic	92.44		
R-squared	0.095		

Table 2. Results of OLS regression of h/y on ability and controls

Variable	Coefficient	Standard Error	T-statistic
age*	-3.70E-02	1.47E-02	-2.52
agesq**	1.06E-03	4.14E-04	2.56
sex	6.35E-03	4.03E-03	1.57
fvincome**	9.42E-08	1.42E-08	6.64
fvincsq**	-2.74E-14	6.85E-15	-4.01
fvinccu*	2.56E-21	1.02E-21	2.51
fvincq*	-7.04E-29	3.50E-29	-2.01
afqt**	7.18E-04	8.57E-05	8.38
Note: * indicates significance at the 15% level or lower; ** at 1%			
Observations	7,280		
F-statistic	46.07		
R-squared	0.047		

Table 3: Welfare gain decomposition of housing subsidy

Share of gain due to:				
Increase in efficiency	Redistribution of total consumption	Individual allocation of consumption	Redistribution of required income	Total
13%	169%	-116%	35%	100%

Table 4. Robustness Checks: results varying curvature of utility and elasticity of labor supply												
gamma	sigma		Capital tax rates at select wages						Optimal *flat* capital tax rate	Welfare gain over no capital tax, (as % of output)	Welfare gain over flat capital tax, (as % of output)	
			$4=Min	$12	$22	$30	$50=Max					
2.0	1.5		16.7%	7.8%	5.1%	3.8%	0%		6.12%	0.0259%	0.0048%	19%
2.0	3.0		4.5%	4.1%	3.6%	3.0%	0%		3.55%	0.0079%	0.0004%	5%
2.0	4.5		1.2%	2.5%	2.6%	2.4%	0%		2.34%	0.0034%	0.0001%	4%
1.0	3.0		3.3%	2.2%	1.8%	1.5%	0%		1.82%	0.0046%	0.0004%	8%
2.0	3.0		4.5%	4.1%	3.6%	3.0%	0%		3.55%	0.0079%	0.0004%	5%
4.0	3.0		9.3%	9.0%	7.7%	6.4%	0%		7.83%	0.0179%	0.0009%	5%
Memo: cdf at wage			0.02	0.37	0.76	0.90	1.00					
Memo: Discount rate			0.25	0.14	0.08	0.05	0.00					

Table 5: Welfare gain decomposition							
gamma	sigma		Share of gain due to:				
			Increase in efficiency	Redistribution of total consumption	Individual allocation of consumption	Redistribution of required income	Total
2.0	1.5		45%	155%	-177%	77%	100%
2.0	3.0		77%	118%	-121%	26%	100%
2.0	4.5		89%	103%	-109%	17%	100%
1.0	3.0		26%	180%	-116%	10%	100%
2.0	3.0		77%	118%	-121%	26%	100%
4.0	3.0		92%	98%	-134%	44%	100%

Table 6. Average tax rates on top earner

gamma	sigma	Welfare gain over no capital tax, (as % of output)	Average tax rate at min wage in optimal policy	Average tax rate at min wage in linear tax policy	Average tax rate at min wage with no capital tax
2.0	1.5	0.0259%	-81.1%	-81.0%	-80.8%
2.0	3.0	0.0079%	-66.8%	-66.7%	-66.7%
2.0	4.5	0.0034%	-65.2%	-65.2%	-65.1%
1.0	3.0	0.0046%	-75.0%	-75.0%	-75.0%
2.0	3.0	0.0079%	-66.8%	-66.7%	-66.7%
4.0	3.0	0.0179%	-57.3%	-57.2%	-57.1%

Figure 1: Implied capital taxes using Lawrance (1991)

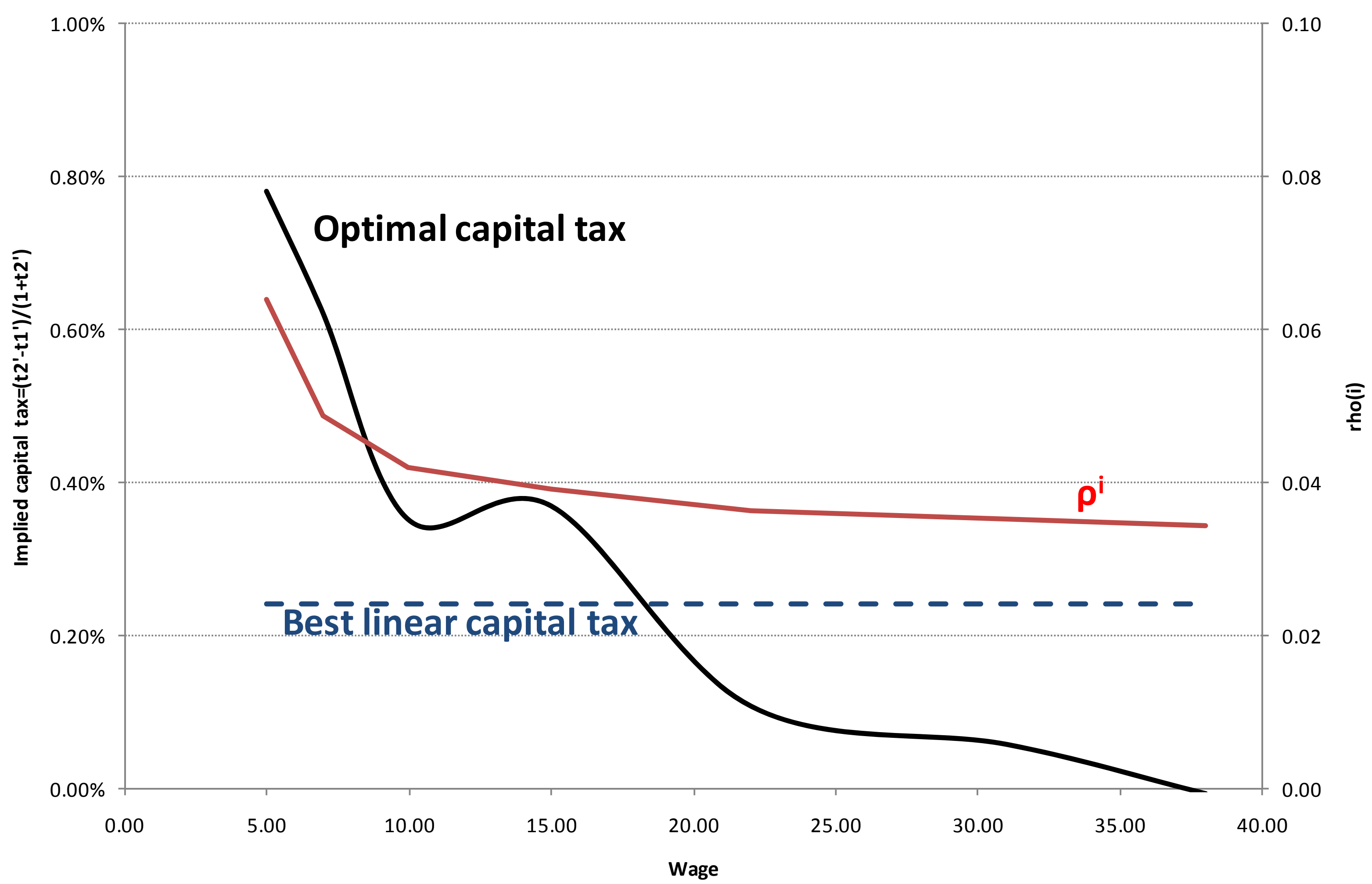


Figure 2: Implied capital taxes using NLSY

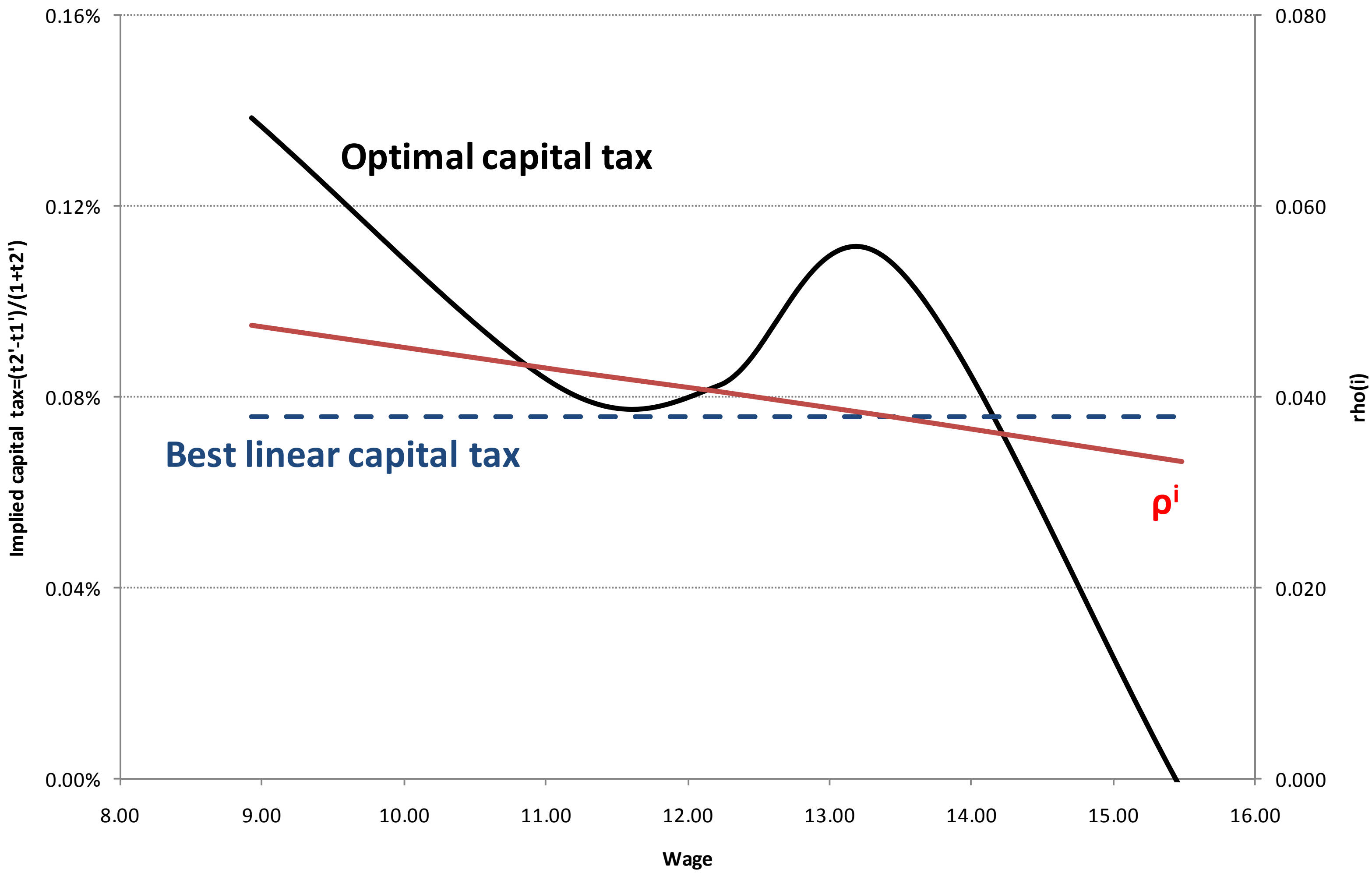


Figure 3: Implied housing consumption taxes

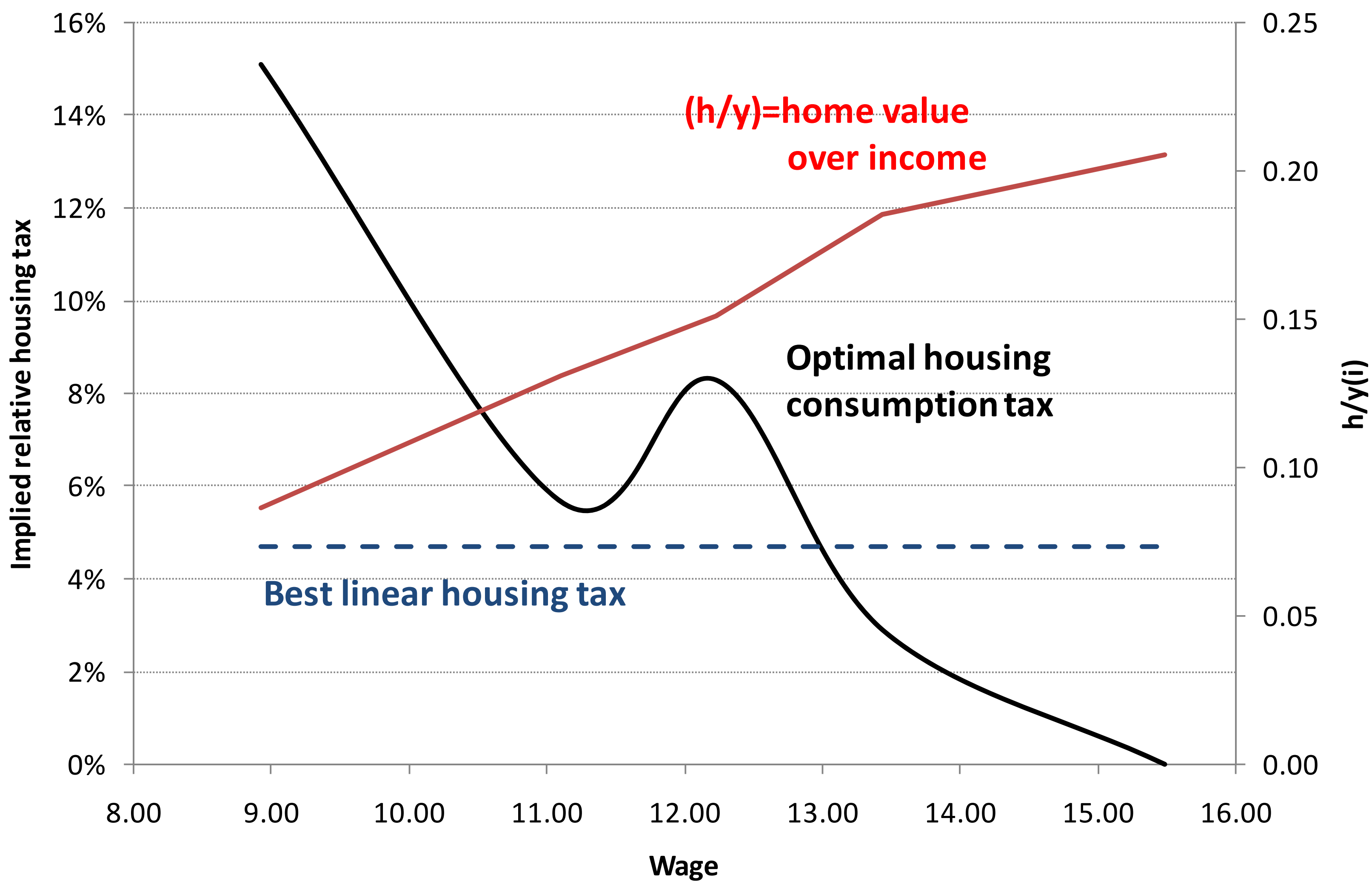
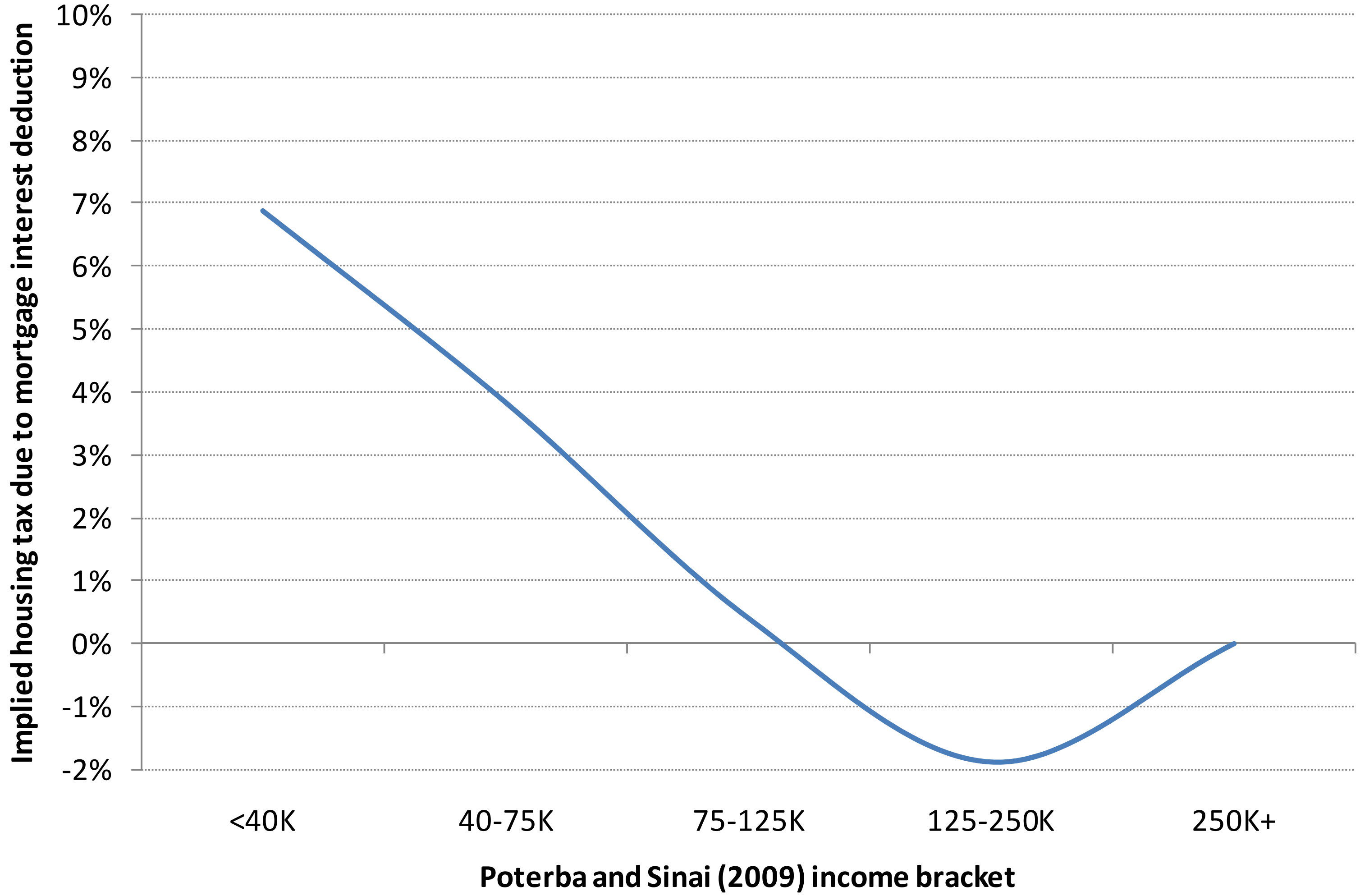


Figure 4: Existing Relative Housing Tax due to Mortage Interest Deduction*



* Calculations based on Poterba and Sinai (2008). Data shown is the present value of mortgage interest deduction savings enjoyed by the top income group, as a percent of home value, less the same present value as a percent of home value for each other income group.

Figure 5: Implied capital taxes in numerical example