Capital Income Taxes with Heterogeneous Discount Rates*

Peter Diamond† Johannes Spinnewijn‡

February 17, 2011

Abstract

With heterogeneity in both skills and discount factors, the Atkinson-Stiglitz theorem that savings should not be taxed does not hold. In a model with heterogeneity of preferences at each earnings level, introducing a savings tax on high earners or a savings subsidy on low earners increases welfare, regardless of the correlation between ability and discount factor. Extending Saez (2002), a uniform savings tax increases welfare if that correlation is sufficiently high. Key for the results is that types who value future consumption less are more tempted by a lower paid job. Some optimal tax results and empirical evidence are presented.

Keywords: Optimal Taxation, Capital Income, Discount Rates
JEL Classification Number: H21

*We are grateful to Richard Blundell, Jesse Edgerton, Louis Kaplow, Emmanuel Saez, Ivan Werning, seminar participants at MIT, Berkeley and the NBER, and two anonymous reviewers for valuable comments, and the National Science Foundation for financial support (Award 0648741). The research reported herein was supported by the Center for Retirement Research at Boston College pursuant to a grant from the U.S. Social Security Administration funded as part of the Retirement Research Consortium. The findings and conclusions are solely those of the authors and should not be construed as representing the opinions or policy of the Social Security Administration or any agency of the Federal Government; or the Center for Retirement Research at Boston College.

†Department of Economics, Massachusetts Institute of Technologys, E52-344, 50 Memorial Drive, Cambridge MA 02142, USA, and NBER (e-mail: pdiamond@mit.edu).

‡Department of Economics, London School of Economics and Political Science (LSE), Houghton Street, London WC2A 2AE, United Kingdom, and CEPR (e-mail: j.spinnewijn@lse.ac.uk).

1 Introduction

In the Mirrlees (1971) model, in the absence of bunching, all the workers at an earnings level are identical. The same property holds in the two-period extension (with one period of work) of Anthony B. Atkinson and Joseph E. Stiglitz (1976) to consider the taxation of capital income. This paper also analyzes two-period models with one period of work, but with the population varying in two dimensions - skill and preference for savings - resulting in heterogeneity in the population at each earnings level. The goal is to explore the implications of this heterogeneity for the welfare implications of introducing taxation of capital income, both uniformly and with rates possibly varying with earnings level.

The Atkinson-Stiglitz theorem states that when the available tax tools include nonlinear earnings taxes, optimal taxation is inconsistent with taxing savings when two key assumptions are satisfied: (1) that all consumers have preferences that are separable between consumption and labor and (2) that all consumers have the same sub-utility function of consumption. The models analyzed in this paper allow for differences in savings preferences as well as differences in ability. In the baseline model, these differences in savings preferences are introduced through differing discount factors in the (additive) utility function. The subutility functions of consumption thus vary in the population. The Atkinson-Stiglitz theorem does not apply.

Primary attention is focused on a model with four worker types - with two discount factors and two skill levels. With multi-dimensional heterogeneity, the implementation of the standard mechanism design solution, potentially separating all types, requires a highly complex tax system. In more realistic settings the complexity of the available tax tools is limited relative to the diversity in the population. Then, not all the workers at an earnings level will be identical. To explore the implications of heterogeneity at individual earning levels, the model assumes the existence of two jobs, rather than the standard model where each worker can select the number of hours to be worked.[1] Workers with the same earnings are subject to the same earnings tax rate. Earnings variation is plausibly much more highly correlated with skill differences than with preference differences and the redistribution across skill types is plausibly more important than across preference types. We therefore assume that at the optimum both high-skill types work at the high-skill job and that redistribution from high earners to low earners is the important redistribution. Given these assumptions social welfare increases with the introduction of a tax on the savings of high earners and with the introduction of a subsidy on the savings of low earners.

The relative frequencies of the four types in the population play no role in the derivation

[1] A limited number of jobs was assumed in Diamond (2006).

of this result, conditional on the assumed structure of the optimum. The case for taxing the savings of high earners appears to be more robust than the case for subsidizing the savings of low earners in some extensions. While the focus of the paper is the introduction of small taxes, we also consider optimal taxes under stronger assumptions. Savings tax policies, like a saver's credit for low-income households, as enacted in the US in 2001, or an absolute cap on tax-favored retirement savings, are in line with the finding that the savings tax should be progressive in earnings.[2]

The key underlying assumption is that those preferring to save more are more willing to work than those preferring to save less, conditional on skill (the disutility of work). This means that an incentive compatibility (IC) constraint just binding on a high skill worker who saves a little is not binding on a high skill worker who saves a lot. Earnings-dependent taxes and subsidies on savings allow an increase in redistribution by targeting types in a given job with saving preferences different from those of types who are just tempted to switch jobs. In particular, introducing taxation of savings of high earners (and transferring the revenue back equally to all high earners) eases the binding IC constraint, since it transfers resources from the high saver to the low saver for whom the IC constraint is binding. Introducing a subsidy on savings for low earners (financed by equal taxation on all low earners) also eases the binding IC constraint by making switching to the lower job less attractive to the high earner with low savings.

The assumption that those preferring to save more are more willing to work is implicit in a standard model with heterogeneity in discounting, representing preferences over first-period consumption x, second-period consumption c and output z by $u(x)+\delta_i u(c)-v(z/n_i)$. Types with higher δ_i prefer to save more and have a higher marginal valuation of net-of-tax earnings. Thus they are more willing to work harder for a given pay increase. As shown in Section 4.1, an alternative specification $(1/\delta_i)\, u(x) + u(c) - v(z/n_i)$ would imply the opposite pattern; types with higher δ_i would prefer to save more, but to work less. In these models, the sign of the relationship between savings and willingness to work is sufficient to determine the welfare effect of introducing earnings-dependent savings taxes. Since the results are exactly opposite when using the alternative specification for which the relationship is negative, we focus the analysis on our preferred specification only. Discount rate differences are just one example of the determination of the relationship between preferences for savings and willingness to work - the relationship depends more generally on heterogeneity in expected inheritances, medical expenditures, wealth, etc...

We examine some empirical support for our assumption that the relationship is positive,

[2] The analysis assumes rational savings by all workers. Concern about too little individual savings is also relevant for retirement savings policies.

using data from the Survey of Consumer Finances (SCF). We find that conditional on education and age, people with higher savings preferences tend to earn more. To proxy for savings, we use reported savings propensities and the time horizon people report having in mind when making spending and savings decisions.

The result on earnings-dependent savings taxes is independent of whether those with higher ability are more likely to have higher savings rates than those with lower ability, provided that the optimum has all the high skilled workers and only high skilled workers on the more productive job. Introduction of a uniform savings tax, however, only increases welfare if the correlation between ability and savings propensities is positive and sufficiently high. Empirical evidence suggests that on average those with higher skills do save at higher rates (Karen E. Dynan, Jonathan Skinner, and Stephen P. Zeldes 2004,[3] James Banks and Diamond, 2010[4]). We also use the proxies mentioned above to revisit the positive correlation between skills and savings propensities.

The result on the uniform savings tax builds on the analysis in Emmanuel Saez (2002). He derives conditions on endogenous variables to sign the effect on social welfare of introducing a uniform commodity tax or a subsidy, when consumers have heterogeneous sub-utility functions of consumption. With an optimal non-linear earnings tax, a small tax on savings increases welfare if either the net marginal social value is negatively correlated with savings, conditional on earnings, or on average those who choose to earn less would save less than those who choose to earn more, if restricted to the same earnings.

This paper is part of the literature on the optimal choice of the tax base and the joint taxation of labor and capital incomes in particular. Banks and Diamond (2010) review the literature on the inclusion of capital income in the tax base. Roger H. Gordon (2004) and Gordon and Wojciech Kopczuk (2008) argue that capital income reveals information about earnings ability and thus should be included in the tax base. Sören Blomquist and Vidar Christiansen (2008) analyze how people with different skills and different preferences for leisure who cannot be separated with an income tax, may be separated with a commodity tax. The four-types model with hours chosen by workers has been studied by Sanna Tenhunen and Matti Tuomala (2010), which calculates a set of examples, but explores the analytics only in two- and three-type models. They consider both welfarist and paternalist objective

[3] While Dynan, Skinner and Zeldes find that high earners save more, they state that a standard model with only discount rate differences can not explain both higher savings when working and higher savings when retired (in a three-period model with two retirement periods). Our focus is just on the savings propensities of workers. We agree that there are multiple factors affecting savings heterogeneity, but think that different discount factors with a positive correlation with skill is one of them.

[4] Beyond looking at empirical studies of savings and earnings, Banks and Diamond discuss the wide finding in the cognitive psychology literature, typically using experimental designs, that higher ability individuals are more patient.

functions. We focus on the four-types model since the result in a two-types model, while striking, does not seem relevant for policy inferences.[5] While the focus of this paper is on capital taxation, the intuition generalizes to the taxation of other commodities for which the preferences are heterogeneous, since this heterogeneity may impact the labor choice as well (Louis Kaplow 2008a).

The paper is organized as follows. Section 2 sets up the model with four types and two jobs. Section 3 characterizes the first best, given the restriction on jobs. Section 4 introduces incentive compatibility constraints and characterizes the second best. Our main result determines the welfare consequence of introducing earnings-varying savings tax rates. We discuss the robustness for some extensions of the model. We then analyze optimal savings tax rates and consider a uniform savings tax as well, rather than one varying with the level of earnings. Section 5 discusses empirical support for the assumptions and Section 6 has concluding remarks.

2 Model

We consider a model with two periods. Agents consume in both periods, but work only in the first period. Preferences are assumed to be separable over time and between consumption and work. Denoting first period consumption by x, second period consumption by c, and earnings by z, preferences satisfy

$$U(x, c, z) = u(x) + \delta u(c) - v(z/n),$$

with $u' > 0, u'' < 0$ and $v' > 0, v'' > 0$. An agent's ability n determines the disutility of producing output z. An agent's preference for future consumption depends on the discount factor δ.

We consider heterogeneity in both ability n and preference for future consumption δ. Although robust insights for optimal taxation have been derived in models with two types, considering heterogeneity in two parameters in a model with two types implies perfect correlation between the two parameters. The inference based on a simple two-types economy, although simple, may therefore be misleading. In order to allow for imperfect correlation, we consider a four-types model. We denote the four types by $l1, l2, h1, h2$ with frequencies

[5] Narayana Kocherlakota (2005) analyzes a model with asymmetric information about stochastically evolving skills, which is not present in this model. The mechanism design optimum has the inverse Euler condition as a discouragement to savings. He shows that the inverse Euler condition can be implemented with a linear wealth tax that is regressive in earnings when capital income is received. Iván Werning (2009) shows that implementation can be done with a nonlinear wealth tax that does not vary with earnings when income is received.

f_{ij} and welfare-weights η_{ij}. The first two types have low ability n_l, and differ in discount factors δ_1 and δ_2, with $\delta_2 > \delta_1$. The type with discount factor δ_2 has a stronger preference for second-period consumption than the type with discount factor δ_1. The second two types have high ability n_h, with $n_h > n_l$, and also differ in discount factors δ_1 and δ_2.

	high discount factor δ_2	low discount factor δ_1
high ability n_h	**h2**	**h1**
low ability n_l	**l2**	**l1**

There are only two jobs in the economy: a high skill job h and a low skill job l. The output z_i from job i is independent of the worker's type, while the disutility of holding a job varies with ability. We add the restriction that everyone holding a job receives the same pay (no taxes based on identity, only on earnings). Hence, the (after-tax) earnings y_i on a job i is independent of the worker's type as well. We assume that the low-ability types can only hold the low job, while the high-ability types can hold either job. However, the redistribution to the low-skilled types is sufficiently important and the type mix sufficiently balanced that all high-skilled workers hold high-skilled jobs at the various optima analyzed. This requires a restriction on the weights in the social welfare functions and the population distribution, which we do not explore.

The desired intertemporal consumption of a given earnings level depends only on a type's discount factor since preferences are separable in consumption and work. There is no dependence on the effort to achieve the earnings. The indirect utility-of-consumption, $w_j[y, R]$, for a type with discount factor δ_j as a function of earnings y and return to saving R equals

$$\begin{aligned} w_j[y,R] \equiv \quad & \max u[x] + \delta_j u[c] \\ \text{subject to:} \quad & x + R^{-1}c = y. \end{aligned}$$

For notational convenience, denote by x_{ij} and c_{ij} the first and second period consumptions levels for a type with discount factor δ_j working on job i. For later use, we also note that

$$\begin{aligned} \frac{\partial w_j}{\partial y} &= u'[x] \\ \frac{\partial w_j}{\partial R} &= R^{-2}cu'[x] = R^{-1}(y-x)u'[x]. \end{aligned}$$

3 Restricted First Best Analysis

To clarify the workings of the model, we begin with optimization with skill known, and so without any IC constraint. We assume that at the optimum each worker is assigned to the matching job: the high-ability types to the high-skilled job h, the low-ability types to the low-skilled job l. The analysis differs from the usual treatment given the restriction that both the output produced and the (after-tax) earnings gained on a job are the same for everyone holding the job. We first characterize the optimal earnings level at each job when savings can not be taxed. We then analyze the welfare consequence of a small earnings-dependent tax on the savings.[6] The (restricted) first best analysis shows that a distortionary tax on savings may increase welfare by redistributing between workers on each job. The second best analysis, however, shows the potential role for a savings tax to increase redistribution between the high earners and low earners, which is absent without an IC constraint.

3.1 No Taxation of Savings

In the (restricted) first best, the social welfare function is maximized with respect to the job-specific earnings and output levels, subject to a resource constraint. With the welfare weight of type ij denoted by η_{ij}, the first best solves:

$$\begin{aligned} &\text{Maximize}_{y,z} \quad \sum f_{ij}\eta_{ij}\left(w_j\left[y_i,R\right]-v\left[z_i/n_i\right]\right) \\ &\text{subject to:} \qquad E+\sum f_{ij}\left(y_i-z_i\right)\leq 0, \end{aligned} \tag{1}$$

where E is an exogenous resource requirement. Forming a Lagrangian with λ the Lagrange multiplier for the resource constraint, we have

$$\mathcal{L}=\sum_{i,j} f_{ij}\eta_{ij}\left(w_j\left[y_i,R\right]-v\left[z_i/n_i\right]\right)-\lambda\sum_{i,j} f_{ij}\left(y_i-z_i\right).$$

We define the net marginal social value of earnings for an individual of type ij as

$$g_{ij}\equiv\eta_{ij}\frac{\partial w_j\left[y_i,R\right]}{\partial y}-\lambda=\eta_{ij}u'\left[x_{ij}\right]-\lambda.$$

[6] In Section 4.4 we also consider analysis of the optimal linear earnings-dependent tax on savings. We do not consider optimization with general (non-linear) taxation of savings as that would call for further analysis of how to model the population in order to have a tax structure with plausible complexity relative to population heterogeneity.

The first order conditions with respect to earnings are

$$\sum_j f_{ij}\eta_{ij}u'[x_{ij}] - \lambda\sum_j f_{ij} = 0,$$

for $i = h, l$.[7] This implies that the population-weighted net marginal social values add to zero at each job,

$$\sum_j f_{ij}g_{ij} = 0 \text{ for } i = h, l.$$

Without restriction on earnings, the net marginal social value would be equal to zero for each type. With the equal pay restriction, redistribution between the workers on a job is desirable if the net marginal social values are different from zero. The welfare weights determine the direction of desired redistribution between workers on each job. In the absence of savings taxation, the intertemporal consumption allocation is undistorted

$$u'[x_{ij}] = \delta_j R u'[c_{ij}] \text{ for all } i, j.$$

3.2 Small Earnings-Dependent Taxes on Savings

Given the observability of earnings and savings, small linear taxes on savings (collected in the first period) could be set differently for high and low earners. This could be implemented by the rules on retirement savings accounts, like the IRA and 401(k) in the US. The (local) desire to redistribute can be met by a small linear tax or subsidy on savings by workers on a given job with the revenues returned equally to them by raising net-of-tax earnings on the job.

The introduction of a small savings tax changes the revenue constraint of the government and the consumption utilities. We differentiate the Lagrangian,

$$\mathcal{L} = \sum_{i,j} f_{ij}\eta_{ij}\left(w_j\left[y_i, (1-\tau_i)R\right] - v\left[z_i/n_i\right]\right) - \lambda\sum_{i,j} f_{ij}\left(y_i - z_i - \tau_i\left(y_i - x_{ij}\right)\right),$$

with respect to a savings tax rate τ_i on those with earnings level y_i. Evaluated at a zero tax

[7] The first order conditions with respect to output are

$$\sum_j f_{ij}\eta_{ij}v'[z_i/n_i]/n_i - \lambda\sum_j f_{ij} = 0 \text{ for } i = h, l.$$

Given the job restriction, the equality between the marginal disutility of work and the marginal value of consumption is only satisfied 'on average' among types at each job.

level, we find

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_i} &= \lambda \left(\sum_j f_{ij} \left(y_i - x_{ij} \right) \right) - \sum_j f_{ij} \eta_{ij} u' \left[x_{ij} \right] \left(y_i - x_{ij} \right) \\
&= \sum_j f_{ij} g_{ij} x_{ij} - \sum_j f_{ij} g_{ij} y_i.
\end{aligned}
$$

Using the FOC with respect to y_i, $\sum_j f_{ij} g_{ij} = 0$ for $i = h, l$, the derivative can be written as:

$$
\frac{\partial \mathcal{L}}{\partial \tau_i} = \sum_j f_{ij} g_{ij} x_{ij}.
$$

This implies that a tax on the savings by those on a given job increases welfare if the savings of the one type towards which redistribution is desirable saves less than the other type.

The welfare weights imply the desired direction of redistribution within productivity types and so the signs of g_{i2} and g_{i1} at each job i. With equal incomes and different discount factors, we have $x_{i2} < x_{i1}$ and $c_{i2} > c_{i1}$. Thus, if first period utilities get the same weights for both types, $\eta_{i1} = \eta_{i2}$, $g_{i1} < 0 < g_{i2}$, implying a desire to redistribute to the high saver. In contrast, if second period utilities get the same weights for both types, $\eta_{i1}\delta_1 = \eta_{i2}\delta_2$, the signs are reversed, implying a desire to redistribute to the low saver. If there is no desire to redistribute for high (low) skill types we have $\eta_{h2} u' \left[x_{h2} \right] = \eta_{h1} u' \left[x_{h1} \right]$ $\left(\eta_{l2} u' \left[x_{l2} \right] = \eta_{l1} u' \left[x_{l1} \right] \right)$. In general, with uniform weights for given discount factors, $\eta_{hi} = \eta_{li}$, we do not satisfy both conditions. In what follows, we assume that the welfare weights are such that at the optimum, lump sum redistribution between workers with the same earnings but different discount rates would not be as important as redistribution to those with lower jobs (or even of zero importance in some results).[8]

4 Second Best Analysis

We now assume that skill is not observable and so consider the second best with the presence of IC constraints involving taking a job with lower productivity (the reverse having been ruled out by assumption). Workers with the same discount factor (but different skills) have exactly the same preferences over consumption for any given level of earnings. Therefore, if a high skill worker were to take the lower productivity job, consumption would match that of the

[8] This is similar to the approach in Saez (1999) which examined optimal taxes to minimize aggregate deadweight burden, assuming no value to lump sum redistributions at the optimum. In contrast, Matthew C. Weinzierl (2009) looks for formulations such that there is no interest in redistribution at a laissez faire equilibrium. We prefer normative evaluations of actual or optimized equilibria to normative evaluations of hypothetical alternatives.

low skill worker with the same discount factor.[9] Hence, the prime issue is determining which high skill workers face a binding IC constraint of not imitating the low skill worker with the same discount factor.

We start with the further restriction, as above, that savings not be taxed. We also add the critical assumption that earnings distribution issues are sufficiently important that at the second-best optimum (with IC constraints) the net marginal social value of first period consumption $g_{ij} \equiv \eta_{ij} u' [x_{ij}] - \lambda$ is negative for both of the worker types holding the high-skill job and positive for both of the types holding the low-skilled job. Without a binding IC constraint, this condition could not hold at the optimum as noted above.

Assumption 1 *At the second-best optimum, the net marginal social values of first period consumption satisfy*

$$g_{hj} < 0, g_{lj} > 0, \text{ for } j = 1, 2.$$

4.1 No Taxation of Savings

We assume that the Pareto-weights and population fractions are such that all high-skilled workers work at the high-skilled job and the desired level of redistribution to lower earners is sufficient that at least one IC constraint is binding.

$$\text{Maximize}_{y,z} \quad \sum f_{ij} \eta_{ij} \left(w_j [y_i, R] - v [z_i / n_i] \right)$$

$$(2) \qquad \begin{aligned} \text{subject to:} \quad & E + \sum f_{ij} (y_i - z_i) \leq 0 \\ & w_h [y_h, R] - v [z_h / n_h] \geq w_h [y_l, R] - v [z_l / n_h] \\ & w_l [y_h, R] - v [z_h / n_h] \geq w_l [y_l, R] - v [z_l / n_h] . \end{aligned}$$

Forming a Lagrangian with μ_j the Lagrange multiplier for the corresponding IC constraint, and assuming that at the optimum each worker is assigned to the matching job, we have

$$\begin{aligned} \mathcal{L} = \sum_{i,j} f_{ij} \eta_{ij} \left(w_j [y_i, R] - v [z_i / n_i] \right) - \lambda \sum_{i,j} f_{ij} (y_i - z_i) \\ + \sum_j \mu_j \left(w_j [y_h, R] - v [z_h / n_h] - w_j [y_l, R] + v [z_l / n_h] \right) . \end{aligned}$$

[9] With no taxation of savings and equal pay for equal work, the IC constraints of not imitating the other discount rate type are trivially satisfied as each worker optimizes savings given his or her labor choice.

Since the first-period consumption of type hj if switching to the low job equals the first-period consumption of type lj, the FOC with respect to earnings are

$$\begin{aligned}\sum_j f_{hj}\eta_{hj}u'[x_{hj}] + \sum_j \mu_j u'[x_{hj}] - \lambda \sum_j f_{hj} &= 0, \\ \sum_j f_{lj}\eta_{lj}u'[x_{lj}] - \sum_j \mu_j u'[x_{lj}] - \lambda \sum_j f_{lj} &= 0.\end{aligned}$$

Given the definition of the net social utility $g_{ij} \equiv \eta_{ij}u'[x_{ij}] - \lambda$, this implies

$$\begin{aligned}\sum_j f_{hj}g_{hj} &= -\sum_j \mu_j u'[x_{hj}] < 0, \\ \sum_j f_{lj}g_{lj} &= \sum_j \mu_j u'[x_{lj}] > 0.\end{aligned}$$

The population-weighted values add to a positive expression

$$\sum_{i,j} f_{ij}g_{ij} = \sum_j \mu_j \left(u'[x_{lj}] - u'[x_{hj}]\right) > 0.$$

That is, transfers which would be worth doing without an IC constraint are restricted, raising the social marginal utilities of consumption, on average, above the value of resources in the hands of the government. Since the IC constraints are on the high skilled types, on average more redistribution from the high earners to the low earners is desirable.

IC constraints Given the equal pay constraint, it follows that only one of the IC constraints is binding, and it is the one on the low discount factor type. To see this consider the difference in consumption utility from different incomes,

$$\Delta[y_h, y_l, \delta_j, R] \equiv w_j[y_h, R] - w_j[y_l, R].$$

This difference in consumption utility is increasing in the discount factor,

$$\frac{\partial \Delta[y_h, y_l, \delta_j, R]}{\partial \delta} = u[c_{hj}] - u[c_{lj}] > 0.$$

The difference in labor disutility does not depend on the discount factor. Thus if the IC constraint is binding on the low discount factor type, it is not binding on the high discount factor type. The low discount factor type values earnings in the first period less and is therefore more tempted to switch to the less productive job.[10]

4.2 Small Earnings-Dependent Taxes on Savings

As above, the sign of the welfare impact of introducing a small linear savings tax or subsidy depends on the welfare weights. Given observability of earnings, the small linear taxes on savings could be different for high and low earners. The welfare impacts of introducing taxes on savings (collected in the first period) are obtained by differentiating the Lagrangian (with savings taxation included and the tax rates τ_i set at zero):

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tau_h} &= \lambda \left(\sum_j f_{hj} \left(y_h - x_{hj} \right) \right) - \sum_j f_{hj} \eta_{hj} u' \left[x_{hj} \right] \left(y_h - x_{hj} \right) - \mu_1 u' \left[x_{h1} \right] \left(y_h - x_{h1} \right), \\ \frac{\partial \mathcal{L}}{\partial \tau_l} &= \lambda \left(\sum_j f_{lj} \left(y_l - x_{lj} \right) \right) - \sum_j f_{lj} \eta_{lj} u' \left[x_{lj} \right] \left(y_l - x_{lj} \right) + \mu_1 u' \left[x_{l1} \right] \left(y_l - x_{l1} \right).\end{aligned}$$

That is, the impact on the Lagrangian is made up of three pieces: the impact on the revenue constraint, the impact on utilities, and the impact on the binding IC constraint.

The FOC for earnings are

$$\begin{aligned}\sum_j f_{hj} g_{hj} + \mu_1 u' \left[x_{h1} \right] &= 0, \\ \sum_j f_{lj} g_{lj} - \mu_1 u' \left[x_{l1} \right] &= 0.\end{aligned}$$

Multiplying these by the earnings level at the job, y_i, and substituting, we have

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tau_h} &= \sum_j f_{hj} g_{hj} x_{hj} + \mu_1 u' \left[x_{h1} \right] x_{h1}, \\ \frac{\partial \mathcal{L}}{\partial \tau_l} &= \sum_j f_{lj} g_{lj} x_{lj} - \mu_1 u' \left[x_{l1} \right] x_{l1}.\end{aligned}$$

[10] The conclusion would be reversed if the low discount factor type values earnings more. With the alternative preference specification, $\frac{1}{\delta} u(x) + u(c) - v(z/n)$, the difference in consumption utility is decreasing in the discount factor δ

$$\frac{\partial \Delta \left[y_h, y_l, \delta_j, R \right]}{\partial \delta} = -\frac{1}{\delta_j^2} \left(u \left[x_{hj} \right] - u \left[x_{lj} \right] \right) < 0.$$

Hence, if the IC constraint is binding on the high discount factor type, it is not binding on the low discount factor type.

Substituting for $\mu_1 u'[x_{i1}]$ from the FOC for earnings, we find

$$\frac{\partial \mathcal{L}}{\partial \tau_h} = f_{h2} g_{h2} (x_{h2} - x_{h1}) > 0 \tag{3}$$

and

$$\frac{\partial \mathcal{L}}{\partial \tau_l} = f_{l2} g_{l2} (x_{l2} - x_{l1}) < 0. \tag{4}$$

The signs follow from the assumption on the net social marginal utilities and the differences in savings behavior by types $i2$ and $i1$ for $i = h, l$. The Proposition immediately follows.

Proposition 1 *At the second best optimum, assuming that all high skill workers hold high skill jobs and $g_{hj} < 0, g_{lj} > 0$, for $j = 1, 2$, then introduction of a small linear tax on savings that falls on high earners is welfare improving; and introduction of a small linear subsidy on savings that falls on low earners is welfare improving.*

At the second best optimum, without a tax on savings, only the high earner/low saver's IC constraint is binding. One can thus increase the redistribution from high earners/high savers by taxing savings, but increasing net-of-tax earnings just enough that the high earners/low savers remain indifferent to job change and thus the binding IC constraint is unchanged. One can also increase the redistribution towards the low earners/high savers by subsidizing their savings, but decreasing net-of-tax earnings such that the low earners/low savers remain indifferent so that it does not become more attractive for the high earners/low savers to take the low job.

The results in Proposition 1 are driven by the implicit assumption that workers with higher savings preferences are also more willing to work for additional pay and thus less tempted to switch to the low-earning job. The results are reversed if the opposite is true. We discuss our assumption providing some empirical evidence in Section 5. Heterogeneity in discount factors is not the only source of heterogeneity affecting the relationship between willingness to save and work. For instance, workers who anticipate higher future expenditures, perhaps medical expenditures, have a higher preference for savings, and also value earnings more. The opposite is true with heterogeneity in current wealth or anticipated inheritances. The sign of the relationship between preferences for savings and earnings, conditional on skill, is sufficient to determine the welfare effect of introducing earnings-dependent savings taxes.

The correlation between skill and discount plays no role in signing the expressions in (3) and (4). As a consequence, this correlation does not affect the results in Proposition

1, provided that the optimum has heterogeneity among the high skill workers on the more productive job. In a two-type model with heterogeneous earnings at the optimum, there is homogeneity at each earnings level. With the usual assumptions, one IC constraint will bind and there is no gain from taxing the savings of the type not being imitated. Whether the gain is from taxing or subsidizing the savings of the other type depends on the correlation between savings and skill (Diamond 2003).[11] The model with four types can have an optimum with only the high skill-high savers at the high job, resulting in a similar conclusion as in the two-types model.[12] The source of this inference does not seem robust to realistic diversity in the economy.

4.3 Robustness

The relative discount factor of the *pivotal* type, who is just tempted to switch to the lower earnings job is at the heart of the results. The desirability of a savings tax on high earners depends on the difference between this type's discount factor and the discount factors of the other high earners. The desirability of a savings subsidy for low earners depends on the difference between this pivotal type's discount factor and the discount factors of the low earners. Changing assumptions in a way that changes the identity of the pivotal worker can alter the findings. In general, for a pivotal type $\hat{\imath}\hat{\jmath}$ with $(n_{\hat{\imath}}, \delta_{\hat{\jmath}})$, the expressions (3) and (4) generalize to

$$\frac{\partial \mathcal{L}}{\partial \tau_h} = \sum_{ij:y_i=y_h} f_{ij} g_{ij} \left(x_j \left[y_h, R \right] - x_{\hat{\jmath}} \left[y_h, R \right] \right) \tag{5}$$

and

$$\frac{\partial \mathcal{L}}{\partial \tau_l} = \sum_{ij:y_i=y_l} f_{ij} g_{ij} \left(x_j \left[y_l, R \right] - x_{\hat{\jmath}} \left[y_l, R \right] \right). \tag{6}$$

We illustrate this by introducing more heterogeneity in ability and discount factors separately. We also show how the analysis extends to three skill levels in workers and jobs, preserving the assumptions of two discount rates and uniform skill on each job. Clearly, the role of job filling - which type works on which job - is central in determining the structure of optimal taxes. Limits on skill variation in who gets hired will tend to support the central

[11] Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski and Weinzierl (2009) consider a model with a continuum of types and perfect correlation between ability and savings preference.

[12] For a numerical example with logarithmic preferences, this holds if the correlation between ability and discount is sufficiently positive. The cut-off correlation below which the optimum has both high skill types on the high skill job, is decreasing in the disutility of additional work $(v(z_h/n) - v(z_l/n))$ and in the welfare weights η_{lj} of the low skill workers.

case of results. Willingness to hire to a job as well as willingness to work at a job both matter for the determination of the earnings levels of different workers.

4.3.1 More Heterogeneity in Ability

We relax the assumption that the two types with high skill have exactly the same skill. With differences in both dimensions among those holding the high skill job, either type might be pivotal, which can reverse the sign of the small tax on savings of high earners that raises welfare. Willingness to work depends on both dimensions, although willingness to hire depends only on skill.

As long as the high skill type with high discount factor has higher skill than the high skill type with low discount factor, Proposition 1 continues to hold. However, if the type with low discount factor is sufficiently more skilled, the type with high discount factor may be more tempted to switch to the low skill job. For given skill of type $h2$, n_{h2}, this reversal of which IC constraint is binding holds when the ability level n_{h1} of type $h1$ is higher than $\hat{n}_{h1}$ $(> n_{h2})$, where the cut-off level $\hat{n}_{h1}$ is such that the IC constraint is just binding on both types,

$$
\begin{aligned}
&\{w_l\,[y_h, R] - v\,(z_h/\hat{n}_{h1})\} - \{w_l\,[y_l, R] - v\,(z_l/\hat{n}_{h1})\} = \\
&\qquad\qquad \{w_h\,[y_h, R] - v\,(z_h/n_{h2})\} - \{w_h\,[y_l, R] - v\,(z_l/n_{h2})\}\,.
\end{aligned}
$$

With $v\,[z/n]$ convex, the difference in labor disutilities between jobs, $\{v\,(z_h/n) - v\,(z_l/n)\}$, is decreasing in n. Hence, for values of n_{h1} higher than $\hat{n}_{h1}$ the IC constraint is more stringent for the high discount saver. In this case, a savings subsidy on the high earners and a savings tax on the low earners are welfare improving. This is the opposite of Proposition 1.

4.3.2 More Heterogeneity in Discount Factors

We relax the assumption that the discount factors of the two types with low skill match those of the two types with high skill. With job-specific earnings and no taxation of savings, a high skill worker considering switching to the low job chooses optimal savings without needing to match any particular worker holding the low job. Thus, with the same skill among high earners, the gain from switching to the low job is always higher for the high skill worker who has lower preference for savings, regardless of the discount rates among the low skill workers. We continue to have a welfare gain from introducing taxation of savings among high earners as in Proposition 1.

Subsidization of savings of low earners will continue to generate a welfare gain as long as the discount factor of the high-skill low-saver is small enough relative to the distribution of discount factors among holders of the low skill job. Denoting by $\widetilde{x}_{h1}$ the first-period consumption of the high-skilled low saver if taking the low skill job, the FOC for earnings on that job is:

$$\sum_j f_{lj}g_{lj} - \mu_1 u'\left[\widetilde{x}_{h1}\right] = 0.$$

The impact of a savings tax on low earners is

$$\frac{\partial \mathcal{L}}{\partial \tau_l} = \sum_j f_{lj}g_{lj}x_{lj} - \mu_1 u'\left[\widetilde{x}_{h1}\right]\widetilde{x}_{h1}.$$

Comparing the consumption in the IC constraint with a weighted average of consumptions among low earners, this derivative is negative (and the gain from the subsidization of the savings of low earners in Proposition 1 continues to hold) if and only if $\widetilde{x}_{h1} > \overline{x}_l$, where

$$\overline{x}_l = \frac{\sum_j f_{lj}g_{lj}x_{lj}}{\sum_j f_{lj}g_{lj}}.$$

With the net marginal social values assumption, $g_{lj} > 0$, for $j = 1, 2$, $\overline{x}_l$ is a proper weighted average of the x_{lj}. Since the discount rate for the marginal high skill type may well be too high to meet this condition, we consider the tax of the savings of higher earners to be a more robust policy conclusion than the subsidization of the savings of low earners.[13]

4.3.3 Three Ability Levels and Three Jobs

We introduce an intermediate skill level and intermediate productivity job in the model. We extend the assumption that welfare weights and population fractions are such that at the optimum all the high skilled are on the most productive job to also have all the intermediate skilled on the intermediate job. We again consider the case in which agents may be tempted to switch to jobs designed for less skilled people. Only two downward constraints are relevant though.

First, as above, for two agents with the same skill, but different discount factors, the IC constraint is slack for the type with the higher discount factor if it is binding for the type with the lower discount factor. The reason is that, with $\Delta\left[y_1, y_2, \delta_j, R\right] \equiv w_j\left[y_1, R\right] - w_j\left[y_2, R\right]$, both $\partial\Delta\left[y_m, y_l, \delta_j, R\right]/\partial\delta > 0$ and $\partial\Delta\left[y_h, y_i, \delta_j, R\right]/\partial\delta > 0$ for $i = m, l$.

[13] With heterogeneity in discount factors, people who discount the future less may choose to invest more in education. If only education determines a worker's skill level, high-skilled workers have higher discount factors than low-skilled workers.

Second, with $v[z/n]$ convex, we have a similar condition for the difference in the disutility of labor between jobs. That is, with $\Delta'[z_h, z_l, n] \equiv v[z_h/n] - v[z_l/n]$,

$$\frac{\partial \Delta'[z_h, z_l, n]}{\partial n} = \left(-v'[z_h/n]\, z_h + v'[z_l/n]\, z_l\right)/n^2 < 0.$$

Thus, for two agents with the same discount factor, the IC constraint of switching to the low-skilled job is slack for the type with the highest ability if it is satisfied for the type with the intermediate ability switching to the low-skilled job and for the type with highest ability switching to the intermediate job. That is, the local IC constraints imply the global IC constraint.

In a similar way as for the four-types model, we can set up the Lagrangian for the constrained maximization problem. The two relevant IC constraints are

$$\begin{aligned} w_l[y_h, R] - v[z_h/n_h] &\geq w_l[y_m, R] - v[z_m/n_h], \\ w_l[y_m, R] - v[z_m/n_m] &\geq w_l[y_l, R] - v[z_l/n_m]. \end{aligned}$$

The impact of the introduction of earnings-dependent savings taxes on the Lagrangian equals respectively

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \tau_h} &= f_{h2} g_{h2} (x_{h2} - x_{h1}) > 0, \\ \frac{\partial \mathcal{L}}{\partial \tau_m} &= f_{mh} g_{mh} (x_{m2} - x_{m1}) \gtrless 0, \\ \frac{\partial \mathcal{L}}{\partial \tau_l} &= f_{l2} g_{l2} (x_{l2} - x_{l1}) < 0. \end{aligned}$$

This implies that Proposition 1 continues to hold for the high earners and the low earners.

For the intermediate earners, the introduction of a savings tax affects the temptation for type $h1$ to switch to the intermediate job and for type $m1$ to switch away for the intermediate job. Both types would, however, choose the same consumption allocation when taking the intermediate job. This simple structure holds since the discount rate is the same for the pivotal workers for movements into and out of the intermediate job. Thus the difference in savings between the types with high discount factors and low discount factors is what matters for the welfare effect of the savings tax. The following Proposition applies for the intermediate earners.

Proposition 2 *In a model with three ability levels and three jobs, the introduction of a small linear tax (subsidy) on savings that falls on the intermediate earners is welfare improving if redistribution from (to) the intermediate earners to (from) general revenues is welfare* *improving.*

From Propositions 1 and 2, there is a single sign change as a function of earnings in the response of welfare to taxing savings. This result generalizes for more than three jobs as well, if the welfare weights are non-increasing in skill. The savings of workers with earnings below a given level are subsidized, the savings of workers with earning above that level are taxed. The result depends on the assumption that types with the same skill are all at the same job, which becomes increasingly strained with many jobs.

4.4 Optimal Linear Earnings-Dependent Taxes on Savings

We have considered the introduction of small savings taxes on high and low earners. Part of the interest in this analysis comes from the possible link to the signs of the optimal taxes. Derivation of the FOC for the optimal linear savings taxes is straightforward; we show that it matches the signs of the small improvements given the additional condition that workers save more if the after-tax return to savings are higher.

One difference in analysis is that changes in both the earnings and savings taxes have a first order effect on tax revenues through the behavioral change in savings. In first period units, the tax revenue from a linear savings tax τ_i levied on the savings of workers with discount factor δ_j and earnings y_i equals $\tau_i s_j [y_i, R(1-\tau_i)]$. For notational convenience, denote optimal savings $s_j [y_i, R(1-\tau_i)]$ by s_{ij}. A second difference is that the relative size of the utility loss of a marginal increase in the savings tax compared to the utility gain of a marginal increase in earnings depends on the level of the savings tax. That is,

$$\frac{\partial w_j [y_i, R(1-\tau_i)]}{\partial \tau_i} = -s_{ij} u' [c_{ij}] \delta R = \frac{-s_{ij}}{1-\tau_i} u' [x_{ij}] = \frac{-s_{ij}}{1-\tau_i} \frac{\partial w_j [y_i, R(1-\tau_i)]}{\partial y_i}.$$

Forming a Lagrangian, and assuming that at the optimum each worker is assigned to the matching job, we now have

$$\begin{aligned}\mathcal{L} = \sum_{i,j} f_{ij} \eta_{ij} (w_j [y_i, (1-\tau_i) R] - v [z_i/n_i]) - \lambda \sum_{i,j} f_{ij} \{(y_i - z_i) - \tau_i s_{ij} [y_i, (1-\tau_i) R]\} \\ + \mu_1 (w_l [y_h, (1-\tau_h) R] - v [z_h/n_h] - w_l [y_l, (1-\tau_l) R] + v [z_l/n_h]).\end{aligned}$$

The FOC for earnings are

$$\begin{aligned}
\sum_j f_{hj}\eta_{hj}u'[x_{hj}] + \mu_1 u'[x_{h1}] - \sum_j \lambda f_{hj}\left(1-\tau_h\frac{\partial s_{hj}}{\partial y_h}\right) &= 0, \\
\sum_j f_{lj}\eta_{lj}u'[x_{lj}] - \mu_1 u'[x_{l1}] - \sum_j \lambda f_{lj}\left(1-\tau_l\frac{\partial s_{lj}}{\partial y_l}\right) &= 0.
\end{aligned}$$

The FOC for savings tax rates are

$$\begin{aligned}
\sum_j f_{hj}\eta_{hj}u'[x_{hj}]\frac{s_{hj}}{1-\tau_h} + \mu_1 u'[x_{h1}]\frac{s_{h1}}{1-\tau_h} - \sum_j \lambda f_{hj}\left\{s_{hj}+\tau_h\frac{\partial s_{hj}}{\partial \tau_h}\right\} &= 0, \\
\sum_j f_{lj}\eta_{lj}u'[x_{lj}]\frac{s_{lj}}{1-\tau_l} - \mu_1 u'[x_{l1}]\frac{s_{l1}}{1-\tau_l} - \sum_j \lambda f_{lj}\left\{s_{lj}+\tau_l\frac{\partial s_{lj}}{\partial \tau_l}\right\} &= 0.
\end{aligned}$$

Denote by $R_h \equiv R(1-\tau_h)$ and $R_l \equiv R(1-\tau_l)$ the after-tax returns to savings for respectively the high and low skill types. Combining the first order conditions as before, we find that the optimal linear savings tax is such that

$$f_{h2}g_{h2}(x_{h2}-x_{h1}) = \tau_h\sum_j \lambda f_{hj}\left\{s_{hj}-\frac{\partial s_{hj}}{\partial y_h}s_{h1}+\frac{\partial s_{hj}}{\partial R_h}R_h\right\} \tag{7}$$

$$f_{h2}g_{h2}(x_{h2}-x_{h1}) = \tau_h\sum_j \lambda f_{hj}\left\{s_{hj}+R_h\frac{\partial s_{hj}}{\partial R_h}|_c\right\}+\frac{\partial s_{h2}}{\partial y_h}[s_{h2}-s_{h1}] \tag{8}$$

and

$$f_{l2}g_{l2}(x_{l2}-x_{l1}) = \tau_l\sum_j \lambda f_{lj}\left\{s_{lj}-\frac{\partial s_{lj}}{\partial y_l}s_{l1}+\frac{\partial s_{lj}}{\partial R_l}R_l\right\}. \tag{9}$$

The left-hand sides in equations (7) and (9) correspond to the welfare changes of introducing earnings-dependent taxes on the high earners and low earners respectively. Thus, if the sum of the terms in brackets on the right-hand side is positive, the optimal linear tax is positive if the introduction of a small tax is welfare-improving and vice versa. Since preferences are additive, $\partial s_{ij}/\partial y_i < 1$, and so $s_{ij} - (\partial s_{ij}/\partial y_i)\, s_{i1} > 0$ for $i = h, l$. Hence, a sufficient condition for the right-hand side term to be positive is that savings are increasing in the after-tax return, $\partial s_{ij}/\partial R_i \geq 0$.[14,15]

Proposition 3 *At the second best optimum, assuming that savings are increasing in the after-tax returns, all high skill workers hold high skill jobs, and $g_{hj} < 0, g_{lj} > 0$ for $j = 1, 2$, the optimal linear savings tax is positive for the high earners and negative for the low earners.*

When considering more earnings levels, the result of a single sign change as a function of earnings also holds for the optimal linear earnings-dependent savings taxes when workers have CRRA preferences, $u[x] = x^{1-\gamma}/(1-\gamma)$, and $\gamma < 1$. With logarithmic preferences, $u[x] = \log[x]$, the optimal savings tax rate is strictly increasing in the earnings of workers if they are uniformly distributed across jobs, $f_{ij} = f_j$ for $\forall i, j$. Since for logarithmic preferences $s_{ij} = (\partial s_{ij}/\partial y_i)\, y_i$ and $\partial s_{ij}/\partial R_i = 0$, the optimal tax on the savings of earners at job i satisfies

$$f_{i2}g_{i2}\left(x_{i2} - x_{i1}\right) = \tau_i \sum_j \lambda f_{ij} \frac{\partial s_{ij}}{\partial y_i} x_{i1}.$$

With $f_{ij} = f_j$ for $\forall i, j$ and first-period consumption $x_{ij} = (1+\delta_j)^{-1} y_i$, we find the following expression for the optimal savings tax,

$$\tau_i = \frac{f_h\left(\delta_1 - \delta_2\right)}{\sum_j \lambda f_j \frac{\delta_j}{1+\delta_j}} \left(\frac{\eta_{i2}}{y_i} - \frac{\lambda}{1+\delta_2}\right).$$

Since $\delta_2 > \delta_1$, with the welfare weights non-increasing in skill, the optimal linear savings tax is increasing in earnings,

$$\frac{\partial \tau_i}{\partial y_i} > 0.$$

4.5 Uniform Taxes on Savings

The previous analysis considers taxes on savings with rates that vary with the level of earnings. This leaves the question of a tax rate on savings that is the same for both earnings levels, as with the Nordic dual income tax. Uniform taxes on savings may be used for instance to lower administration costs and reduce arbitrage opportunities.

[14] The relationship between improvements from small taxes and optimal taxes would be reversed if the terms in brackets were negative. However, reversing the sign of the impact of the rate of return on savings would not be a sufficient condition for the sign to be negative.

[15] Tenhunen and Tuomala (2010) analyze the mechanism design optimal allocations with varying correlations between discount and skill. In contrast with the optimal tax model analyzed here, the mechanism design optimal allocation allows distortion of the savings of each type separately. Their calculations for the four-type model with CES preferences suggest a similar pattern of savings taxes for (some) high skill types and negative savings taxes for (some) low skill types, as long as the correlation is not too high.

In order to evaluate the introduction of a small uniform tax on savings, we add the responses to the two separate tax changes,

$$\frac{\partial \mathcal{L}}{\partial \tau} = \frac{\partial \mathcal{L}}{\partial \tau_h} + \frac{\partial \mathcal{L}}{\partial \tau_l} = f_{h2} g_{h2} \left(x_{h2} - x_{h1}\right) + f_{l2} g_{l2} \left(x_{l2} - x_{l1}\right).$$

In contrast with the earnings-varying tax on savings, the correlation between skill and discount factor plays a role here.

The previous results were driven by a desire to redistribute from higher earners to low earners, $g_{hj} < 0, g_{lj} > 0$, for $j = 1, 2$. For simplicity, here we make the further assumption that there is no desire to redistribute within a job, $g_{i2} = g_{i1}$, for $i = h, l$. Using the FOC for earnings y_i, we find

$$\begin{aligned} f_{h2} g_{h2} &= \frac{f_{h2}}{\sum_j f_{hj}} \sum_j f_{hj} g_{hj} = -\frac{f_{h2}}{\sum_j f_{hj}} \mu_1 u' \left[x_{h1}\right], \\ f_{l2} g_{l2} &= \frac{f_{l2}}{\sum_j f_{lj}} \sum_j f_{lj} g_{lj} = \frac{f_{l2}}{\sum_j f_{lj}} \mu_1 u' \left[x_{l1}\right]. \end{aligned}$$

Thus, the welfare impact of introducing a small uniform tax on savings equals

$$\frac{\partial \mathcal{L}}{\partial \tau} = \mu_1 \left(\frac{f_{h2}}{\sum_j f_{hj}} u' \left[x_{h1}\right] \left(x_{h1} - x_{h2}\right) - \frac{f_{l2}}{\sum_j f_{lj}} u' \left[x_{l1}\right] \left(x_{l1} - x_{l2}\right) \right).$$

It is convenient to write this as

$$\frac{\partial \mathcal{L}}{\partial \tau} = \mu_1 \frac{f_{l2}}{\sum_j f_{lj}} u' \left[x_{l1}\right] \left(x_{l1} - x_{l2}\right) \left(\frac{f_{h2} / \sum_j f_{hj}}{f_{l2} / \sum_j f_{lj}} \Omega - 1 \right),$$

with

$$\Omega \equiv \frac{u' \left[x_{h1}\right] \left(x_{h1} - x_{h2}\right)}{u' \left[x_{l1}\right] \left(x_{l1} - x_{l2}\right)} > 0.$$

Since $x_{l1} > x_{l2}$,

$$sign \left(\frac{\partial \mathcal{L}}{\partial \tau} \right) = sign \left(\frac{f_{h2} / \sum_j f_{hj}}{f_{l2} / \sum_j f_{lj}} \Omega - 1 \right).$$

The sign of this expression depends on both the relative proportions of savings types, $(f_{h2} / \sum_j f_{hj}) / (f_{l2} / \sum_j f_{lj})$, and the relative importance of differences in savings, Ω. Note that, $\Omega \geq 1$ is a sufficient condition for a positive correlation of skill with proportions, $f_{h2} / \sum_j f_{hj} > f_{l2} / \sum_j f_{lj}$, to imply that introducing a savings tax increases social welfare.

Assuming homothetic consumption preferences, so that $x_{h1}/x_{l1} = x_{h2}/x_{l2}$, the expression for Ω becomes $\left(u' \left[x_{h1}\right] x_{h1}\right) / u' \left[x_{l1}\right] x_{l1}$. This expression is equal to one for the log utility function. For CRRA preferences, $u[x] = x^{1-\gamma}/(1-\gamma)$, we find

$$\Omega = \left(\frac{x_{h1}}{x_{l1}}\right)^{1-\gamma}.$$

Thus, if the relative risk aversion γ is smaller than 1, then $\Omega \geq 1$ and a positive correlation between ability and discount factor (i.e. $f_{h2}/\sum_j f_{hj} > f_{l2}/\sum_j f_{lj}$) implies that $\partial\mathcal{L}/\partial\tau$ is positive. If γ is larger than 1, the sign of $\partial\mathcal{L}/\partial\tau$ depends on the size of the correlation and the magnitude of the earnings difference between jobs. Conversely, when the correlation is negative, $\partial\mathcal{L}/\partial\tau$ is negative if γ is greater than 1.[16] This implies the following proposition.

Proposition 4 *If there is no desire to redistribute within a job, $g_{i2} = g_{i1}$, for $i = h, l$, with CRRA preferences, a uniform small tax on savings increases welfare if the relative risk aversion is smaller than one and the correlation between ability and discount factor is positive. A uniform small subsidy on savings increases welfare if the relative risk aversion is greater than one and the correlation between ability and discount factor is negative.*

Corollary 1 *If there is no desire to redistribute within a job, $g_{i2} = g_{i1}$, for $i = h, l$, with logarithmic preferences, a uniform small tax (subsidy) on savings increases welfare if and only if the correlation between ability and discount factor is positive (negative).*

As with the earnings-varying taxes, the sign result for introducing a uniform tax matches that for optimal linear taxation in some interesting cases. Denote by $R_\tau \equiv R(1-\tau)$ the after-tax returns to savings and by s_{ij} the savings of type ij as a function of after-tax earnings and the after-tax interest rate. Setting the derivative of the Lagrangian with respect to τ equal to zero, we find the following condition for the optimal linear tax,

$$f_{h2}g_{h2}(x_{h2} - x_{h1}) + f_{l2}g_{l2}(x_{l2} - x_{l1}) = \tau \sum_{i,j} \lambda f_{ij} \left\{ s_{ij} - \frac{\partial s_{ij}}{\partial y_i} s_{i1} + \frac{\partial s_{ij}}{\partial R_\tau} R_\tau \right\}.$$

If the sum of the terms in brackets is positive, we have that the optimal uniform tax is positive if the introduction of a small uniform tax is welfare improving. This is the case for logarithmic preferences and CRRA preferences with relative risk aversion $\gamma < 1$.

Proposition 5 *If there is no desire to redistribute within a job, $g_{i2} = g_{i1}$, for $i = h, l$, with logarithmic preferences or CRRA preferences with $\gamma < 1$, the optimal linear uniform tax on savings is positive if the correlation between ability and discount factor is positive.*

[16] For CARA preferences, $\frac{\partial\mathcal{L}}{\partial\tau}$ is negative when the correlation between ability and discount factor is negative. When the correlation is positive, $\frac{\partial\mathcal{L}}{\partial\tau}$ is positive if the absolute risk aversion is sufficiently small.

5 Preferences and IC Constraints

Above we used the utility functions $u[x] + \delta_j u[c] - v[z/n_j]$. This family of utility functions has the property that those with higher savings rates (larger values of δ_j) are more willing to increase work for a given amount of additional pay. But that is not the only way in which the savings and labor supply decisions can be connected in this simple setting. As noted above, with the utility functions $(u[x] + \delta_j u[c])/\delta_j - v[z/n_j] = u[x]/\delta_j + u[c] - v[z/n_j]$, the relationship is reversed - those with higher savings rates are less willing to increase work for additional pay. If we had assumed this class of functions, then we would have reversed the pattern of desirable savings taxes in Proposition 1 - having the IC constraint bind for the high saver would imply that it is not binding for the low saver, implying, in turn, that there should be a subsidy of savings for high earners and a tax on savings for low earners.

That it is standard practice to write utility in the form employed does not, by itself, shed light on its empirical reality. More generally, a one-dimensional family of separable utility functions, $U[\phi[x, c, j], z, j]$, can have any pattern between the variation in the subutility function of consumption and the variation in the interaction between consumption and labor. In the example of a distribution of inheritances or medical expenses, the IC constraint that is binding depends on the timing of the event - a future event fits with the proposition in the text; earlier events reverse it. This raises the question of identifying an empirical basis for distinguishing which case is more relevant.

It is not easy to find data applying directly to this issue. The models we have examined have two periods with one period of work. They can be thought of as modeling working life, and then retirement. The question we want to answer is whether, for a given level of skill, those with higher savings rates tend to have greater labor supply functions. While the model has only two types at each skill and so a perfect correlation between these two characteristics, presumably a more heterogeneous population and recognition of the stochastic nature of employment opportunities would move the empirical test to one of correlation.

Perhaps the most direct empirical measure relating to this picture would be the willingness to work beyond age 62 being positively correlated with wealth for a given lifetime average wage level. The model leaves out many features that affect wealth accumulation, such as random returns on investments, and a proper test would need to recognize the variation in earnings opportunities at age 62 relative to earlier earnings opportunities. Attempting to control for these factors would require an empirical study well beyond what would fit as casual evidence on the sign of the correlation between skill and savings propensity. Instead we look at properties within a single year as simply measured in the Survey of Consumer Finances (SCF).

Before turning to the data, we believe that there are also reasons based on casual empiricism for supporting the appropriateness of using our formulation and expecting a positive correlation more generally. Modeling savings with rationality and discounting combines underlying preferences and issues of self-control. As discussed in Banks and Diamond (2010), psychological analyses suggest these are mixed together. We see no reason to think that this does not apply to working as well as to consuming - whether that is working for later consumption or working to influence future work opportunities. That is, working (at a job with disutility) involves self-control for a future payoff. And saving involves self-control. So those with less difficulty in self-control may show greater willingness to both work and save, which would be captured in the standard utility function expression. How serious a self-control problem should be before moving away from accepting revealed preference as a normative criterion is not clear. In a richer model, human capital investment involves discounting in a similar way to savings decisions and so may generate the pattern in the standard model structure, although formal modeling would distinguish between human and financial capital accumulations.

5.1 Empirical Correlations

We use the SCF panels in 1998, 2001 and 2004, containing information on 13,266 households in total.[17] The tendency to save, in a way perhaps robust over time, is plausibly captured in whether subjects confirm the statement that they: "Save regularly by putting money aside each month."[18] First, we use this proxy to confirm the correlation of saving with potential for earnings. Of course, circumstances that can affect regular savings behavior vary with age, on average. So, we do the calculation using separate age cells. As shown in Table 1, the proportion of regular savers rises steadily with education, taken as a proxy for earnings potential (skill). The clear pattern supports the positive correlation between higher earnings potential and saving habits. This is consistent with the empirical evidence reviewed in Banks and Diamond (2010) and Dynan, Skinner and Zeldes (2004).

[17] Population weights are used to convert the SCF sample to a representative national sample.

[18] The results are similar (with sign reversal) with the statement "Don't save - usually spend about as much as income."

TABLE 1: Proportion of regular savers

	30-34	35-39	40-44	45-49	50-54	55-59
< **High School**	0.21	0.16	0.26	0.25	0.28	0.27
High School	0.34	0.40	0.37	0.39	0.41	0.48
Some College	0.37	0.45	0.50	0.42	0.45	0.51
College Degree	0.58	0.54	0.51	0.57	0.57	0.50
Graduate Degree	0.64	0.56	0.62	0.59	0.54	0.52

Turning to our primary concern, whether for a given skill, those who save more are more willing to work, we look at the correlation between the savings proxy and log earnings within the age-education cells in Table 2.

TABLE 2: Correlation log(earnings) and saving regularly

	30-34	35-39	40-44	45-49	50-54	55-59
< **High School**	0.30*	0.17*	0.22*	0.20*	0.31*	0.40*
High School	0.24*	0.17*	0.29*	0.22*	0.32*	0.34*
Some College	0.18*	0.21*	0.20*	0.23*	0.20*	0.17*
College Degree	0.22*	0.34*	0.20*	0.28*	0.19*	0.10*
Graduate Degree	0.30*	-0.01	0.17*	0.15*	0.15*	0.13*

*Significantly different from 0 at the 5 percent level.

Except for one, all of the within-cell correlations in Table 2 are positive and significant. The correlation for the full sample between the logarithm of earnings and saving regularly, conditional on cell dummies, equals .22. The positive correlations are supportive of the positive correlation which fits with the assumptions in Proposition 1, which applies with both the discount rate interpretation and the future income interpretation.

The SCF offers a second proxy that relates directly to the model with different discount rates. We use the question asking: "In planning (your/your family's) saving and spending, which of the following is most important to [you/you and your (husband/wife/partner)]: the next few months, the next year, the next few years, the next 5 to 10 years, or longer than 10 years?" To interpret the averages, we quantify this variable assigning $0.5, 1, 3, 7.5$ and 15 to the respective answers. Table 3 shows the average time horizon per cell and Table 4 shows the correlation between time horizon and earnings within cells. Time horizon increases with education and so wage rate, which is supportive of the assumed positive correlation between discount factor and skill used in analysis of the taxation of savings that is not earnings-varying.

TABLE 3: Average time horizon (converted into years)

	30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**
< High School	3.3	3.0	4.5	4.4	4.5	5.3
High School	4.2	4.9	5.5	5.6	5.2	5.5
Some College	4.7	5.9	6.3	5.7	6.0	6.4
College Degree	6.5	7.0	7.3	7.1	7.9	7.5
Graduate Degree	7.4	8.7	8.6	8.4	8.3	8.2

TABLE 4: Correlation log(earnings) and time horizon

	30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**
< High School	−0.00	0.12*	0.11*	0.19*	0.22*	0.29*
High School	0.13*	0.14*	0.19*	0.09*	0.13*	0.22*
Some College	0.12*	0.11*	0.14*	0.23*	0.16*	0.12*
College Degree	0.17*	0.37*	0.20*	0.29*	0.29*	0.22*
Graduate Degree	0.31*	0.18*	0.24*	0.28*	0.24*	0.24*

*Significantly different from 0 at the 5 percent level.

As with regular savings, those with higher education have longer horizons and within education cells, log earnings and time horizon are positively correlated.[19] While this is a robust prediction of the two-period model, in a three-period model with consistent discount rates within the working years, types with higher discount factors will work relatively more while they are young but may not when older. In the appendix we look at hours worked per week, which gives a similar, but less clean answer, reflecting the role of uniform hours on many jobs. We consider again whether one reports to be saving regularly and the reported time horizon.

Education choices reflect both ex ante "skill" and discount rate and then affect wage rates, which matter for later taxation. Presumably, the level of completed education is increasing in both ex ante skill and discount factor, on average. In addition to affecting ex post skill, education may affect one's discount rate thereafter. Thus education is a proxy for both skill

[19] Christopher F. Chabris et al. (2008) find that experimentally measured discount rates have strong predictive power for field behavior relative to other variables in their sample (e.g., sex, age, education). However, they find that the correlation between the discount rate and each field behavior is small; none exceeds 0.28 and many are near 0.

and discount rate and can not be used in a simple way to distinguish between them. A further difficulty in interpreting the correlations is that education is a discrete variable while skill is continuous and varying within education classes. In general, there are many factors that affect savings and work that are not in the basic model used for analysis. Thus, this evidence is merely suggestive.

6 Conclusion

Design of the taxation of capital income needs to reflect many factors. This paper focuses on heterogeneity in savings rates, an important dimension of heterogeneity for tax setting. The paper uses a model with jobs, rather than one with individual worker choices of hours, in order to have diversity in savings behavior at each earnings level in a tractable form. Neither labor market model describes the nature of opportunities for all workers, leaving room for learning from both types of models. And neither model has a role for hiring decisions, which are relevant for determining opportunities. In an hours model, workers make changes in response to small changes in marginal taxes. In a jobs model, there are many workers who are not at the margin of switching to a different job. Their lack of labor supply response to small changes in taxes is important for tax policy and seems plausible for many workers.

In keeping with the optimal tax literature the objective function in this paper has been defined in terms of individual lifetime utilities. Rather than considering how to weight the utilities of those with different discount factors (or preference differences more generally), we have reported some results in terms of different welfare weights.[20] And we have used a model where the social objective function respects all preferences, not allowing for concerns that some people save too little for their own good.[21] Moving from this analysis toward concrete policy recommendations calls for addressing these issues, as well as the more complex need to move from analyses based on lifetimes to ones that incorporate additional concerns that are relevant for taxes set annually, primarily on annual tax bases.

[20] A number of other papers have considered optimal taxes with heterogeneous preferences. See, for example, Agnar Sandmo (1993), Helmuth Cremer and Firouz Gahvari (1998, 2002), Katherine Cuff (2000), Robin Boadway et al. (2002), Blomquist and Christiansen (2004), Ritva Tarkiainen and Tuomala (2007), Kaplow (2008b).

[21] Addressing this concern would include consideration of mandatory retirement income programs and the effect of the design of savings incentives on consumer behavior, beyond the standard model of lifetime utility maximization.

Appendix: Hours of Work

The results for the correlation between hours worked and the two proxies for the savings preferences are in Tables A1 and A2 respectively. The question we use for hours of work asks: "How many hours (do you/does [he/she]) work on (your/her/his) main job in a normal week? (if not self-employed) How many hours (do you/does [he/she]) work in this business in a normal week? (self-employed)."[22]

TABLE A1: Correlation hours worked and saving regularly

	30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**
< High School	0.05	−0.00	0.17*	0.02	0.24*	0.04
High School	0.07*	0.15*	0.05	−0.01	0.10*	0.02
Some College	−0.03	0.04	0.18*	0.13*	0.06*	−0.18*
College Degree	0.11*	0.10*	0.05*	−0.03	0.10*	0.14*
Graduate Degree	0.08	−0.04	0.07*	0.04*	0.01	0.10*

* denotes statistical significance at the 5 percent level

TABLE A2: Correlation hours worked and time horizon

	30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**
< High School	0.23*	0.22*	−0.07	−0.15*	−0.01	0.08
High School	−0.02	0.05	0.07*	0.03	0.19*	0.08*
Some College	−0.03	0.07*	0.13*	0.15*	−0.01	0.14*
College Degree	−0.01	−0.01	0.04	0.11*	0.07*	0.05
Graduate Degree	0.14*	0.14*	0.15*	−0.02	0.05*	−0.10*

* denotes statistical significance at the 5 percent level

Most of the correlations are positive. Many of them are significant as well. The correlations for the full-sample using the two different proxies for saving preference, conditional on cell dummies, are both .06.

[22] The result for annual hours worked are very similar. This leaves out the role of second jobs.

References

[1] **Atkinson, Anthony B. and Joseph E. Stiglitz.** 1976. "The Design of Tax Structure: Direct Versus Indirect Taxation." *Journal of Public Economics*, 6(1-2): 55-75.

[2] **Banks, James and Peter Diamond.** 2010. "The Base for Direct Taxation." In Reforming the Tax System for the 21st Century: The Mirrlees Review, ed. Institute for Fiscal Studies. Oxford: Oxford University Press.

[3] **Blomquist, Sören and Vidar Christiansen.** 2008. "Taxation and Heterogeneous Preferences." *Finanz Archiv*, 64: 218-244.

[4] **Boadway, Robin, Maurice Marchand, Pierre Pestieau, and María del Mar Racionero.** 2002. "Optimal Redistribution with Heterogeneous Preferences for Leisure." *Journal of Public Economic Theory*, 4(4): 475-498.

[5] **Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathan P. Schuldt, and Dmitry Taubinsky.** 2008. "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty,* 37(2-3): 237-269.

[6] **Cremer, Helmuth, and Firouz Gahvari.** 1998. "On Optimal Taxation of Housing." *Journal of Urban Economics,* 43: 315-335.

[7] **Cremer, Helmuth, and Firouz Gahvari.** 2002. "Nonlinear Pricing, Redistribution and Optimal Tax Policy." *Journal of Public Economic Theory,* 4: 139-161.

[8] **Cuff, Katherine.** 2000. "Optimality of workfare with heterogeneous preferences." *Canadian Journal of Economics*, 33(1): 149-74.

[9] **Diamond, Peter.** 2003. *Taxation, Incomplete Markets, and Social Security.* Cambridge: MIT Press.

[10] **Diamond, Peter.** 2006. "Optimal Tax Treatment of Private Contributions for Public Goods with and without Warm Glow Preferences", *Journal of Public Economics*, 90(4-5): 897-919.

[11] **Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes.** 2004. "Do the Rich Save More?" *Journal of Political Economy*, 112(2): 397-444.

[12] **Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Matthew C. Weinzierl.** 2010. "Preference Heterogeneity and Optimal Capital Income Taxation." NBER Working Paper Series 16619.

[13] **Gordon, Roger H.** 2004 "Taxation of Interest Income." *International Tax and Public Finance*, 11(1): 5-15.

[14] **Gordon, Roger H. and Wojciech Kopczuk.** 2008. "The Choice of the Personal Income Tax Base." http://www.columbia.edu/~wk2110/bin/choiceBase.pdf

[15] **Kaplow, Louis.** 2008a. *The Theory of Taxation and Public Economics.* Princeton: Princeton Press.

[16] **Kaplow, Louis.** 2008b. "Optimal Policy with Heterogeneous Preferences." *B.E. Journal of Economic Analysis and Policy*, Vol. 8, Iss. 1 (Advances), Article 40: 1-28.

[17] **Kocherlakota, Narayana.** 2005. "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica*, 73(5): 1587-1621.

[18] **Mirrlees, James A.** 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38: 175-208.

[19] **Saez, Emmanuel.** 1999. "A Characterization of the Income Tax Schedule Minimizing Deadweight Burden." Unpublished.

[20] **Saez, Emmanuel.** 2002. "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics,* 83(2): 217-30.

[21] **Sandmo, Agnar.** 1993. "Optimal redistribution when tastes differ." *Finanz Archiv*, 50(2): 149-163.

[22] **Tarkiainen, Ritva and Matti Tuomala.** 2007. "On optimal income taxation with heterogeneous work preferences." *International Journal of Economic Theory*, 3(1): 35 - 46.

[23] **Tenhunen, Sanna and Matti Tuomala.** 2010. "On Optimal Lifetime Redistribution Policy." *Journal of Public Economic Theory,* 12(1): 171-198.

[24] **Weinzierl, Matthew C.** 2009. "Incorporating Preference Heterogeneity into Optimal Tax Models." http://www.people.hbs.edu/mweinzierl/paper/PreferenceHeterogeneity_OptimalTax.pdf

[25] **Werning, Iván**. 2009. "Nonlinear Capital Taxation." Unpublished